Where to Spend the Next Million?

Applying Impact Evaluation to Trade Assistance

chidi A. Okoye

edited by **Olivier Cadot, Ana M. Fernandes, Julien Gourdon and Aaditya Mattoo**

CE
PR

THE WORLD BANK

WHERE TO SPEND THE NEXT MILLION?

Where to Spend the Next Million? Applying Impact Evaluation to Trade Assistance

Centre for Economic Policy Research

The Centre for Economic Policy Research is a network of over 700 Research Fellows and Affiliates, based primarily in European universities. The Centre coordinates the research activities of its Fellows and Affiliates and communicates the results to the public and private sectors. CEPR is an entrepreneur, developing research initiatives with the producers, consumers and sponsors of research. Established in 1983, CEPR is a European economics research organization with uniquely wide-ranging scope and activities.

The Centre is pluralist and non-partisan, bringing economic research to bear on the analysis of medium- and long-run policy questions. CEPR research may include views on policy, but the Executive Committee of the Centre does not give prior review to its publications, and the Centre takes no institutional policy positions. The opinions expressed in this report are those of the authors and not those of the Centre for Economic Policy Research.

CEPR is a registered charity (No. 287287) and a company limited by guarantee and registered in England (No. 1727026).

The World Bank

The World Bank Group is a major source of financial and technical assistance to developing countries around the world, providing low-interest loans, interest-free credits and grants for investments and projects in areas such as education, health, public administration, infrastructure, trade, financial and private sector development, agriculture, and environmental and natural resource management. Established in 1944 and headquartered in Washington, DC, the Group has over 100 offices worldwide. The World Bank's mission is to fight poverty with passion and professionalism for lasting results and to help people help themselves and their environment by providing resources, sharing knowledge, building capacity and forging partnerships in the public and private sectors.

Where to Spend the Next Million?

Applying Impact Evaluation to Trade Assistance

edited by

OLIVIER CADOT, ANA M. FERNANDES,
JULIEN GOURDON AND AADITYA MATTOO

Contents

List of Figures

List of Tables

Foreword

Economic development is a process of continuous industrial and technological upgrading. For any country, regardless of its level of development, a necessary condition for success is that it develops industries that are consistent with its comparative advantage, determined by its endowment structure. The graduation from low-skilled manufacturing activities in China and other large middle-income countries will open up unprecedented industrialisation and trade opportunities for African and other low-income countries. The lower-income countries that can design and implement a viable strategy to capture this new opportunity will start a dynamic process of structural change that can lead to poverty reduction and prosperity.

But what constitutes a viable strategy, and how do we design and implement it? The 'negative agenda', *ie* eliminating policy barriers to trade that distort incentives, is reasonably well understood and well accepted. Thus, there has been considerable progress on trade liberalisation, though the results have sometimes fallen short of expectation. The 'positive agenda', *ie* proactive policy to enhance trade competitiveness, is less well understood and a bigger implementation challenge. In principle, there is clearly a role for policy to address market failures that increase trade costs and inhibit exports. In practice, successful intervention has proved hard to design and execute.

In order to intervene effectively and efficiently, developing countries need evidence-based advice. They need to know which interventions work and which do not, in which sectors, in which sequence and which are most cost-effective. But the traditional trade policy literature provides little guidance on the design of proactive policies for cutting trade costs or promoting exports. To draw an analogy: we know from impact evaluation that for the purpose of keeping an East African child in school an extra year of de-worming is more cost-effective than other complex and costly transfer schemes. Similarly, we need to know where the biggest bang for our buck would be in trade facilitation and export promotion.

Sceptics argue that rigorous evaluation of trade-related interventions is inherently infeasible. I do not agree with such 'trade exceptionalism'. I believe rigorous impact evaluation is feasible if we are open to the use of a variety of methodologies: not just randomised evaluations, but any technique that can enable us to assess the extent to which observed outcomes can be attributed to a policy intervention, a project or a programme.

I am very glad that this volume seeks to facilitate, and in the process develop methodologies for, rigorous impact evaluation of trade-related interventions (whether financed by World Bank projects or by governments). As part of our Open Data, Open Knowledge and Open Solutions initiative, I am also proud to report that a parallel process is underway to create a large cross-country firm-level customs-transactions database which will be made publicly available and will allow evaluators everywhere to assess the impact of trade interventions.

Justin Yifu Lin
Senior Vice President and Chief Economist
World Bank

Acknowledgements

Impact Evaluation of Trade Assistance: Paving the Way

OLIVIER CADOT, ANA M. FERNANDES, JULIEN GOURDON AND
AADITYA MATTOO[1]

1 INTRODUCTION

Trade policy has changed fundamentally since the days of structural adjustment and economy-wide trade reforms. Partly in reaction to the uneven results of trade policy reforms, the focus has shifted to more targeted interventions aimed at reducing trade costs and addressing market failures that inhibit exports. Significant national resources and international assistance are now devoted to trade facilitation and export promotion, and the international development community has galvanised around a new 'aid-for-trade' (AfT) mantra as a means of helping low-income countries integrate into the global economy.

The environment in which trade-related assistance is provided has also changed. In times of fiscal austerity, taxpayers increasingly question the justification for large aid flows and, at the very least, demand results and accountability.[2] The development community has struggled to respond to these demands because there is surprisingly little evidence about what does and does not work in the area of trade and industrial policies.

An authoritative survey of trade and industrial policy recently acknowledged that there is hardly any microeconomic evidence to guide specific trade interventions (Harrison and Rodrigues-Clare 2010). Even the most basic ques-

[2]A recent poll featured by the *Financial Times* (12 July 2010) showed that the majority of respondents in OECD countries considered defence and development aid as priority areas for spending cuts.

tions go unanswered. For instance:

- if reducing trade costs is the objective, should limited resources be focused on transport or customs reform?
- in customs reform, should the emphasis be on computerising processes or on creating incentives for integrity?
- in transport, should it be on containerising ports or improving inland links?
- if the object is to enhance the ability of firms to export, should the focus be on assistance to particular firms or on improving the business environment?
- if particular firms, should the focus be on the few large firms who can use it but may not need it, or the many small firms who need it but may not be able to use it?
- if the focus is on improvements in the business environment, should it be economy-wide or good-governance enclaves such as export processing zones?

There are two reasons for the disappointing pace at which evidence on these questions has gathered. First, trade policy research has been slow to respond to changing needs. Tariffs continue to occupy centre stage in policy research, in spite of their declining importance as trade barriers, simply because they are easy to measure. Second, the AfT community has been slow to build a culture of rigorous evaluation. For instance, a review of 85 recent World Bank trade-related projects conducted by the authors revealed that only 5 of them included rigorous evaluation components. Worse, those few evaluations relied on crude before–after comparisons, which are known to be vulnerable to confounding influences.

Still, the tools for a serious evaluation of trade-related interventions are there. Originally developed in the medical sciences, impact-evaluation (IE) methods have spread to the social sciences and are routinely employed in the areas of health and education. In essence, an impact evaluation compares the outcomes of entities—individuals or firms—that received support from a programme or were directly affected by a policy with the counterfactual outcome of those same entities had the programme or policy not been in place. Because such counterfactual outcomes are not observable, they are approximated by the outcomes of a control group.

The recent creation by the World Bank of a separate impact-evaluation unit as part of the Development Impact Evaluation Initiative (DIME) has helped spread IE methods to new areas of development research and practice.[3] For instance, World Bank researchers have led the way in analysing the impact of business registration reform or bankruptcy reform (Klapper and Love 2010;

---

[3]Information on DIME can be obtained at http://go.worldbank.org/1F1W42VYV0.

Bruhn 2011; Gine and Love 2010). Researchers have also begun to use these methods to evaluate programmes and policies in the area of private sector development, where the treated 'entities' are firms (see McKenzie (2010) for a survey).

IE methods have also provided powerful tools in other fields to help guide policy choices and minimise the cost of interventions. For instance, Banerjee and Duflo (2008) showed how a comparison of IE results established that, in order to raise school attendance rates among Kenyan children, a programme to treat intestinal worms was 20 times more cost-effective than hiring teachers, suggesting a clear prioritisation of actions.[4] Similar evaluations could be used to guide trade interventions.

The usual excuse for *not* using IE methods in assessing the effectiveness of trade assistance is that the 'clinical' nature of the treatment needed for a proper definition of treatment and control groups is absent from trade policy. This was perhaps true of old-style trade policies like structural adjustment or tariff reforms, but it is not true of the new trade interventions like export promotion. *Trade exceptionalism*—the notion that trade-related interventions are inherently not amenable to IE—is, as this volume intends to show, groundless. Trade-related interventions can be evaluated formally, provided that we are not wedded to a particular methodology such as randomised controlled trials (RCTs). Although, as we will see, the range of application of RCTs is broader than might be assumed, other quasi-experimental methods are available and can shed light on what works and what does not.

RCTs are only one of the possible approaches for rigorous impact evaluation. For instance, some countries implement regulatory reforms in a staggered fashion, starting in a small set of locations before extending to all locations. The impact of such reforms can be rigorously evaluated by using locations where the reforms are introduced later as a control group for the locations where reforms are introduced earlier and using a difference-in-differences estimation methodology (Bruhn 2008). Similarly, *ex post* evaluation of programmes and policies is a possible approach, provided that information is available both on those firms that received support from a programme or were directly affected by a policy, and on the entire (or a large portion of the) universe of firms. In these circumstances, it is possible to use propensity score matching combined with difference-in-differences estimation (see, for example, Lopez-Acevedo and Tinajero 2010; Tan 2009).

---

[4]This ratio was established by comparing the evaluation of a de-worming programme by Miguel and Kremer (2004) with a separate evaluation of a programme to reduce teacher–student ratios by Banerjee *et al* (2005). Comparing impact estimates from separate impact evaluations is tricky since each has been established in a particular context with limited external validity (we will return to the issue later in this chapter). However, when the difference in cost effectiveness is as large as this one, the risk of getting the prioritisation order wrong is reduced.

These methods have already been applied in a number of recent studies, some of which are included in this volume, and have produced interesting and unexpected results. Consider the following three examples.

First, in an *ex post* evaluation of export promotion programmes in six Latin American countries using rich firm-level data sets, Volpe Martincus shows in Chapter 2 that these programmes were effective in facilitating export expansion primarily along the extensive margin (*ie* through an increase in the number of products exported or in the number of export markets served) rather than along the intensive margin (an increase in exports of existing products to existing markets). He also shows that programmes benefited small and relatively inexperienced firms more than larger and already established exporters, and that bundled services providing support to firms throughout the export-development process were more effective than isolated actions.

Gourdon *et al* use similar *ex post* evaluation methods in Chapter 3 to assess the impact of a World Bank-financed export promotion programme in Tunisia (FAMEX), which provided a mixture of counselling and matching grants to new exporters. Their findings suggest that export promotion has a significant effect on overall export growth: a 39% increase in the average annual growth rate of programme beneficiaries relative to the control group over a four-year period. The effect of the programme on the extensive margin of exports—in terms of products and destinations—is more subdued: about 5% higher growth for beneficiaries, which is significant only for destinations. They also find a significant increase in employment growth, *ie* 10% more for programme beneficiaries than for control firms. The effect on export growth is stronger for firms that were initially only marginal exporters (exports represented less than 20% of turnover). Interestingly, their sample also includes service firms, for which the effect of export promotion is significantly larger than for manufacturing firms.

In Chapter 6, Datt and Yang analyse a natural experiment in which the Philippines government suddenly reduced the minimum value threshold under which shipments were exempt from pre-shipment inspections (PSIs), closing a loophole that had encouraged importers to slice shipments in order to escape inspection. They show that the reform failed to curb under-invoicing, and thus to raise duty collection, as importers switched to an alternative loophole, namely, the use of an export-processing zone (EPZ). As this alternative loophole involved high fixed costs (setting up a presence in the EPZ), in the end the Philippine government was no better off, while importers were worse off. The authors also discuss the effects of a related policy reform in Colombia, where the government sought to remedy undervaluation of certain imports by mandating PSI on a subset of products. This, however, left open the loophole of misclassification of those products as similar products that did not require a PSI. Both cases illustrate the importance of careful, incentive-compatible reform design.
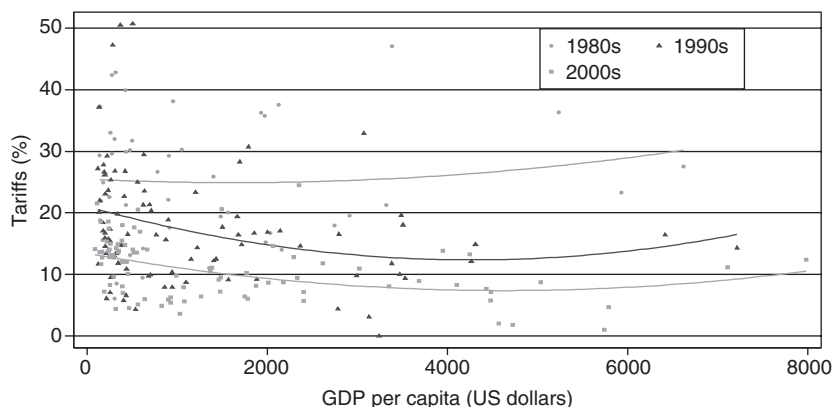
Figure 1.1: *Tariffs and GDP per capita.*
Source: World Integrated Trade Solution/World Development Indicators.

The rest of this chapter is organised as follows: in Section 2 we discus the changing nature of trade policy. In Section 3 we review the available evidence on the impact of trade assistance. In Section 4 we consider a detailed menu of trade-related interventions and discuss the challenges to their evaluation. In Section 5 we address the data issues crucial to impact evaluation. Finally, in Section 6 we look at the future challenges to doing IE in trade assistance.

2   THE CHANGING NATURE OF TRADE POLICY

2.1   *From Old-Style Reforms to Focused Interventions*

Most developing countries have moved beyond the first generation of trade reforms, which involved across-the-board cuts in tariffs and the elimination of import quotas. As shown by Figure 1.1, tariffs have fallen substantially over the last 20 years. The simple average applied tariff of OECD countries on all goods was 2.9% in 2008, and the developing country average is down to around 10% compared to 30% in 1990.

Recourse to quantitative restrictions has also substantially declined. One reason is the narrower interpretation of the *balance-of-payments exception* in the WTO and the stricter enforcement of the conditions under which it can be invoked. Countries like India have been forced to phase out numerous quotas that had been maintained for a long time, ostensibly to address balance-of-payments difficulties. Another reason is the tighter interpretation following the Uruguay Round Agreement of the *national treatment* provision in the WTO, which precludes the local-content requirements that many developing countries had favoured and other members had tolerated.
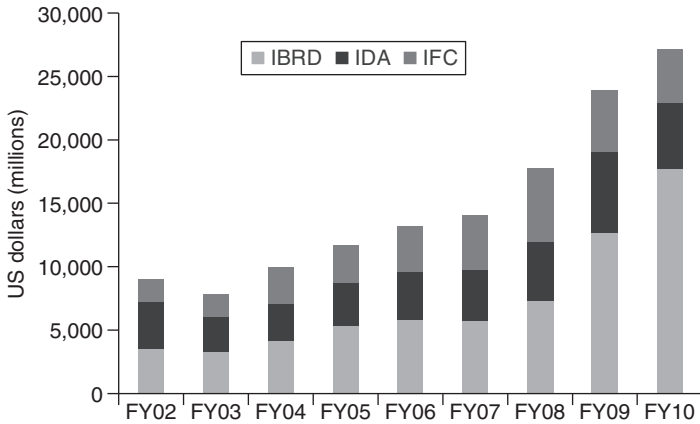
Figure 1.2: *World Bank aid-for-trade commitments 2002–10.*
Source: authors' calculations based on data from the World Bank Business Warehouse website.

With this decline in traditional barriers to market access, supply-side constraints are seen as the main obstacle faced by developing countries in taking advantage of new opportunities in international markets. Therefore, trade interventions are becoming more targeted, focusing on either the *trade-facilitation* agenda, involving, for example, customs reforms and infrastructure (*eg* port) improvements, or the *trade competitiveness* agenda, consisting of proactive industrial policies, involving productive capacity building, EPZs or export promotion. In designing such trade interventions, developing countries need policy advice, in particular, more evidence-based advice. They need to know which interventions work and which do not, in which sectors, in which sequence and which ones are most cost-effective.

2.2   The Shifting Focus of Multilateral Assistance

The World Bank has shifted its emphasis in trade assistance from broad trade liberalisation reforms in the 1980s and 1990s to more targeted interventions since the early 2000s, to reduce the costs of trade and to equip producers to export. The declaration of WTO ministers in Hong Kong in 2005 and the first Global Aid for Trade Review in Geneva in 2007 gave an impetus to the expansion of AfT to help developing countries build their supply-side capacity and trade-related infrastructure.

The World Bank responded by expanding its commitments on trade competitiveness, trade facilitation and infrastructure, and is now a leading contributor to AfT. As shown by Figure 1.2, recent commitments by the World Bank are substantial and growing: concessional trade-related lending (as per the OECD/WTO definition) to low-income countries grew from US$3.18 billion

Figure 1.3: *World Bank Group trade portfolio 2008.*
Source: authors' calculations based on data from the World Bank Business Warehouse website.

annually in 2002–5 to an average of US$4.84 billion in 2007–8, while non-concessional trade-related lending to middle-income countries increased from US$4.16 billion in 2002–5 to US$9.8 billion in 2007–8 (World Bank 2011).[5] Since 2001, the World Bank has approved 437 trade-related lending projects in 90 countries and 53 trade-related lending operations in ten regional groups, with Africa and Eastern Europe and Central Asia accounting for most of the operations (World Bank 2011).

Trade-facilitation-related infrastructure is the largest single component of World Bank trade-related investments in developing countries, while the rest consist mostly of improving competitiveness. Figure 1.3 shows the distribution of World Bank commitments on AfT as of fiscal year 2008, while Table 1.1 details the types of interventions falling under the 'trade competitiveness' and the 'trade-facilitation' agendas.

Given the increase in AfT, donors and recipients would like to see evidence that this new type of assistance will be more effective than past aid efforts.

---

[5]The numbers presented in the figure are based on the OECD/WTO definition of aid for trade. The sectors that fall under this definition are (1) for IBRD/IDA: agriculture, fishing and forestry; information and communication; energy and mining; transportation; industry and trade; (2) for IFC: agriculture and forestry; information; oil, gas and mining; chemical; utilities; transportation and warehousing; construction and real estate; food and beverages; non-metallic mineral product manufacturing; primary metals; pulp and paper; textiles, apparel and leather; plastics and rubber; industrial and consumer products; wholesale and retail trade; professional, scientific and technical services; accommodation and tourism services.

Table 1.1: *Focused trade interventions.*

Trade competitiveness (including trade finance)	Trade facilitation and logistics
· Export promotion/diversification	· Customs reform
· Support to producer/exporter organisations	· Ports/airports rehabilitation
· Quality testing and export certification	· Railway privatisation/rehabilitation
· Technology upgrading and support services	· Roads construction/rehabilitation
· Strengthening policy/regulatory framework	
· Export credit insurance	
· Export credit guarantee	
· Line of credit	
· Support for financial institutions	

These concerns are especially strong in the aftermath of the 2008–10 global financial crisis, when pressures to reduce fiscal deficits and debt are weakening political support for foreign assistance. In fact, a recent opinion poll in OECD countries revealed that a large majority of the public favoured cuts in defence and aid spending rather than in other categories of expenditure.[6]

3   EVALUATING THE IMPACT OF FOCUSED TRADE INTERVENTIONS

In this section, we review three kinds of existing evaluation effort. The first involves broad, inconclusive assessments of AfT and its impact. The second examines the effect of national trade interventions, such as export promotion activities, but still at a highly aggregate level, considering mostly aggregate exports as outcomes; this provides some support for certain types of focused interventions. The third set of efforts involves assessments by the World Bank of its own trade-related projects. While the last set are in principle as focused as the interventions themselves, they have for the most part not been based on the collection or analysis of any hard evidence on impact.

3.1   *Evaluating Aid for Trade*

The literature on the impact of AfT is fairly limited, in part because AfT projects are not always distinguishable from other aid projects. As in the rest of the aid-effectiveness literature, the results are ambiguous (Rajan and Subramanian 2008). Regarding the cross-country allocation of AfT, Gamberoni and Newfarmer (2009) find that, after controlling for absorption capacity (related, for example, to governance), more AfT is directed towards countries with a

---

[6]See *Financial Times*, 12 July 2010.

higher demand for AfT as measured by indicators of 'underperformance' in trade.[7]

One strand of the literature explores whether AfT positively affects exports from the donor country to the recipient country given that, until the early 1990s, over half of all bilateral aid was at least partly tied to donor exports. Using a gravity equation, Wagner (2003) shows that this form of trade was indeed boosted; but Osei *et al* (2004), using a gravity equation in first differences for a panel of four European donors and 26 African recipients, found an unstable and insignificant impact of aid on exports from donor to recipient. Recently, Nelson and Silva (2008) have used a more conventional gravity equation, including bilateral aid flows as a regressor (instrumented by their one-year lagged value), and have found a small but significant impact on trade flows from donor to recipient.

From a development perspective, only a few of the recent studies focus on the more relevant question of whether aid raises the export capacity of recipient countries. Calì and te Velde (2011) regress trading costs and the value of exports on lagged AfT disbursements and control variables, using data from the OECD's Creditor Reporting System that separately identifies aid-to-trade facilitation and infrastructure from aid to productive capacity.[8] Using a large panel of developing countries, Calì and te Velde address the possibility of endogeneity and measurement errors in AfT flows by instrumenting those with Freedom House's index of civil liberties. The message that emerges across their various specifications is that aid-to-trade facilitation and infrastructure seems to have a significant effect in reducing trade costs and in increasing export values, while aid to productive capacity is insignificant. When considering sectorally targeted aid, Calì and te Velde again find that aid to infrastructure has a significant impact on export values, but aid to productive capacity does not, controlling for country–sector fixed effects that account for comparative advantage differences.

Brenton and von Uexkull (2009) examine the response of product-level exports from developing countries to product-level export-development aid, combining mirrored product-level (HS4) export data with export-development aid data from the German cooperation agency GTZ and from the OECD/WTO

---

[7]Underperformance in trade is captured by multiple indicators. Countries that underperform in trade can be those in the lower two quintiles of performance measured along five dimensions: those experiencing relatively slow growth of exports of goods and services; those losing global market share; those suffering deterioration in competitiveness in existing markets; those exporting slow-growing products or to slow-growing markets; those over-reliant on only a few exports. Also, countries that underperform in trade are those that under-trade with bilateral partners, controlling for market size and distance, those with low-level scores on the World Bank logistics performance index for transport or for customs and on an indicator of peak tariffs.

[8]Trading costs are measured by the trading-across-borders indicators of the Doing Business database.

Trade Capacity Building Database for 48 developing countries. Using a matching difference-in-differences (DID) approach (discussed in Section 4) they show insignificant effects of contemporaneous and lagged aid on product-level exports after controlling for lagged exports, and country and year product fixed effects and eliminating outliers.[9] However, Brenton and von Uexkull do show strong positive effects in a simple comparison of product-level exports before and after receiving export-development aid. This finding suggests an important attribution problem—namely, export growth may not be due to the aid received but instead may reflect the fact that aid targets sectors with promising prospects. They go on to argue that, in evaluating the impact of technical assistance for exports, it is essential to identify what would have happened in the absence of the policy intervention. This is a primary concern in this chapter, and in this volume.

As the literature stands, it is fair to say that the effect of AfT on the export performance of beneficiary countries has not been established on the basis of aggregate numbers. Ferro *et al* advance the analysis of the effectiveness of AfT in Chapter 8 in this volume, revisiting the data from OECD's Creditor Reporting System. They exploit the differential intensities of service use across manufacturing sectors (based on input–output tables from the USA and Argentina) to evaluate the impact of AfT flows directed at five service sectors (transport, communications, energy, banking/financial services and business services) on the exports of downstream manufacturing sectors in 106 aid-recipient countries over the period 1990–2008. Their identification strategy aims at circumventing reverse-causality problems common in the AfT literature, and their results show that aid flows directed at the energy and banking sectors have a significant positive impact on downstream manufacturing exports.

3.2   *Evaluating National Trade Interventions*

A few recent cross-country studies suggest a positive impact of certain types of trade interventions, regardless of whether they are financed by donors or domestic government budgets. On export promotion, Lederman *et al* (2010a) examine the effectiveness of export promotion agencies (EPAs) based on a rich survey of EPAs across 88 developed and developing countries. The goals of EPAs are to help exporters understand and find markets for their products and services and can be divided into four categories (Lederman *et al* 2010a, pp 257–8):

---

[9]Their matching approach pairs each treatment country that receives export-development aid for a given product i to the country that is more similar to it in terms of its likelihood to export product i, where this likelihood is estimated based on observable country characteristics such as the level of development, factor endowments and climate conditions.

- country image building (advertising, promotional events, but also advocacy);
- export support services (exporter training, technical assistance, capacity building, including regulatory compliance, information on trade finance, logistics, customs, packaging, pricing);
- marketing (trade fairs, exporter and importer missions, follow-up services offered by representatives abroad);
- market research and publications (general, sector, and firm-level information, such as market surveys, online information on export markets, publications encouraging firms to export, importer and exporter contact databases).

For 21 of the 73 developing countries surveyed, Lederman *et al* find that EPAs receive budgetary support from multilateral donors such as the World Bank. They estimate the effect of EPAs' expenditures per capita on overall exports per capita at the country level, accounting for selection bias in survey responses and for potential reverse causality. Their main conclusion is that, on average, EPAs have a significant positive effect on exports. Their estimates also point to the importance of EPAs' services for overcoming foreign trade barriers and solving asymmetric information problems associated with exports of differentiated goods. In addition, they find evidence of strong diminishing returns, suggesting that 'small is beautiful' as far as EPAs are concerned. However, they acknowledge that cross-country regressions cannot fully capture the heterogeneity of policy environments and institutional structures in which EPAs operate; hence, more detailed studies or project-type analyses are needed to provide specific policy advice.

On the subject of trade facilitation, Helble *et al* (2009) examine the responsiveness of trade flows to various types of AfT—linked to reform of trade policy and regulation, trade development (productive capacity building) and economic infrastructure—using a gravity equation framework covering 167 importers (reporters) and 172 exporters (partners) during the 1990–2005 period. Their results indicate that relatively small amounts of aid targeted at trade policy and regulatory reform have a greater impact with respect to increased trade flows than aid for broad trade development assistance or infrastructure. Several recent papers point to the importance of internal barriers related to infrastructure and institutions, including logistics performance, as obstacles to developing countries' ability to trade and the volume of trade (*eg* Djankov *et al* 2010; François and Manchin 2006; Freund and Rocha 2010; Hoekman and Nicita 2008; Portugal-Pérez and Wilson 2010). More specific studies highlight the importance of reducing marketing, transport, and other intermediary costs in agricultural supply chains (Balat *et al* 2009; Diop *et al* 2005). Although these studies point out the relevance of increased donor

assistance to trade facilitation, they do not help delineate the policies and programmes that would be most effective in cutting trade costs.[10]

In their recent authoritative survey of the state-of-the-art literature on industrial policy, Harrison and Rodrigues-Clare (2010) conclude that empirical evidence on the effectiveness of various forms of industrial policy is scarce. They look at the case of East Asian countries where industrial policies based on use of production subsidies, subsidised credit, fiscal incentives and trade protection to foster particular sectors. From this, they claim that the available evidence does *not* answer the most important question: what was the effect of these industrial policies relative to the counterfactual situation where such intervention was absent?[11] In summary, there are no studies that can credibly credit industrial policies with bringing about East Asia's successful industrialisation experience. But Harrison and Rodrigues-Clare do make a tentative argument that industrial policies played a role in some countries' growth experiences based on two complementary ideas. First, the composition of a country's export basket—a tilt towards manufacturing or skill-intensive goods rather than primary products or raw materials—seems to matter for its long-run growth. Second, China's export basket in 1992 was much more sophisticated than would be expected given the country's per capita gross domestic product (GDP) and that could only be the outcome of its industrial policies (Rodrik 2006).[12]

Harrison and Rodrigues-Clare's literature survey concludes with an advocacy statement on the type of national trade-related assistance likely to be most successful: that which increases exposure to trade (such as export promotion) in contrast to that which limits trade (such as tariffs or domestic content requirements).[13] They also make a statement on the specifics of policy design, where they envision an increasing role for 'soft' industrial policies that deal directly with coordination problems, such as those that keep productivity low in existing or emerging sectors. These policies include programmes 'to

---

[10]As an example of the type of results in these papers, Portugal-Pérez and Wilson (2010) estimate the impact of aggregate indicators of 'soft' and 'hard' infrastructure on the export performance of 101 developing countries over the 2004–7 period. Their estimates show that trade-facilitation reforms, particularly investment in physical infrastructure and regulatory reform to improve the business environment, improve significantly export performance. Moreover, their estimates provide evidence that the marginal effect of infrastructure improvements on exports appears to be decreasing with per capita income.

[11]One empirical approach that has been followed in some studies is to examine whether the sectors that received most support from industrial policies are those that have grown most rapidly, but that approach does not address the counterfactual issue.

[12]This finding was based on the measure of sophistication of a country's exports basket developed by Hausman *et al* (2005) constructed using the level of GDP per capita associated with exports of different goods worldwide.

[13]The authors make this statement based on extensive cross-country and cross-sector evidence on trade and growth.

help particular clusters by increasing supply of skilled workers, encouraging technology adoption, and improving regulation and infrastructure' (Harrison and Rodriguez-Clare 2010, p 4112).[14] The problem with this statement is that it has a 'magic wand' aspect, because the survey includes little supporting evidence. In fact, the absence of evidence for the policy recommendations the survey offers is a reason for our effort to initiate new research on these issues.

3.3   Evaluating the World Bank's Trade Programmes

In principle, World Bank trade-related projects should be a key source of evidence on the effects of specific trade interventions, which could become the basis for further evidence-based policy advice. In practice, though, this is rarely the case. Few interventions have undergone rigorous impact evaluation.

An evaluation of World Bank financed trade-related assistance during the 1987–2004 period conducted by the Independent Evaluation Group (IEG) concluded that it helped countries liberalise their trade regimes—average tariffs fell and coverage of non-tariff barriers diminished—with positive effects on economic growth (IEG 2006). However, the evaluation also argued that assistance fell short of generating a strong export supply response. Many client countries, especially in Africa, could not diversify their exports and remained vulnerable to commodity price shocks.

IEG (2006) also discusses the performance ratings of World Bank AfT projects, which give a sense of their effectiveness in achieving their stated goals. The report shows that trade-related adjustment loans until 2004 performed better than other adjustment loans, whereas trade-related investment loans performed worse than other investment loans of the World Bank.[15] Moreover, according to the same evaluation, assistance on trade logistics—ports, customs and trade finance—and export incentives had a mixed record, though one that improved over time.

A review of the IEG ratings of recent investment projects and programmes on trade promotion, completed in 2007 (World Bank 2009), indicates that more than 85% were rated as having moderately satisfactory, satisfactory or highly satisfactory outcomes, which was higher than for projects in other

---

[14]The authors argue that an advantage of such 'soft' industrial policies is that they are generally compatible with the multilateral and bilateral trade agreements that developing countries have entered into in recent decades.

[15]Projects that focused primarily on trade liberalisation achieved the best performance ratings, whereas those related to private financing (such as export finance guarantees and export reinsurance) were the least successful. The superior performance of projects focusing on trade liberalisation is not surprising, as it reflects the relative legislative ease of putting in place the associated actions (*eg* reform of the tariff regime). In contrast, projects that focused on thematic areas related to key supply-side constraints that impose greater demands on institutional and administrative capacity, such as trade financing, are more difficult to implement.

areas.[16] Aid-for-trade projects also had higher estimated economic rates of return (around 32%) than other non-trade related projects (around 23.7%).[17]

While they provide valuable insights, the IEG evaluations of trade assistance offer limited evidence to support focused trade interventions. Moreover, the evaluation does not cover much of the recent increase in AfT assistance for export promotion and trade facilitation.

In search of evidence on the impact of such trade interventions, we conducted a thorough review of the evaluation methods for 85 World Bank trade-related investment lending projects undertaken during the 1995–2005 period. The source of data was the World Bank's Operations portal website and, in particular, the Project Appraisal Documents (PADs) and the Implementation Completion Reports (ICRs).[18] The evaluation methods used can be classified into five distinct categories:[19]

(a) only economic or financial internal rates of return, net present value or effectiveness calculations;
(b) beneficiary surveys and stakeholder workshops;
(c) both (a) and (b);
(d) both (a) and (b), with a comparison of beneficiaries to a control group;
(e) no formal evaluation methods used.

One key aspect to note is that the implementation of a beneficiary survey does not guarantee that a rigorous impact evaluation can be conducted, since in most cases the survey covers only outcomes pertaining to beneficiaries of the

---

[16]IEG assesses the performance of roughly one World Bank project out of four (about 70 projects a year), measuring outcomes against the original objectives, sustainability of results and institutional development impact.

[17]An economic rate of return is the discounted interest rate that would keep an agent indifferent between the choice of undertaking or not undertaking the project.

[18]We thank Vivian Agbhega for compiling the data for this review. The selection of trade-related projects followed the criteria used by Steven Gunawan in a study of 'Monitoring and Evaluation Lessons of Trade Projects' that served as background work for the 2011 World Bank Trade Strategy. The projects were filtered from the World Bank's Operations portal website according to the theme 'Trade and Integration', and fell within the following criteria: (i) approved only after 1995 due to obsolescence; (ii) IBRD/IDA-funded; (iii) closed. A total of 321 projects were filtered, out of which 144 were development policy loans and 177 were investment loans, and 30 investment lending projects had to be dropped since they lacked ICRs. A final set of 85 investment lending projects was obtained after excluding projects that did not have any trade components. The main documents used to extract information on the projects were PADs and ICRs. For each project we collected information on the types of intervention, the types of outputs and outcomes achieved, the evaluation methods employed, and the evidence or proof of causation of the impact of the project.

[19]A beneficiary survey consists of a formal survey of the entities that received assistance from the project, whereas a stakeholder workshop is a more informal way to collect information on the various entities affected by the project.
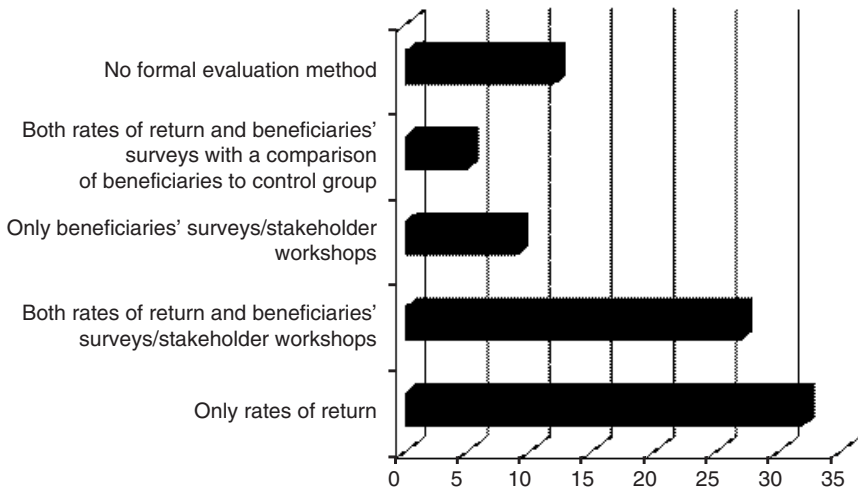
Figure 1.4: *Evaluation of World Bank trade-related projects 1995–2005.*
Source: authors' calculations based on data from the World Bank Operations portal website.

project, and no control group is covered (more details on these methods are provided in Section 4).

Figure 1.4 shows that evaluation using only economic or financial rates of return was the most common method for the trade-related projects, while 10% of the projects involved no formal evaluation method.[20] Included in the latter category is a trade competitiveness project that described the impact of the project in purely subjective terms:

> While the impact on the firms assisted had not yet been determined, a visit to two beneficiaries by a supervision mission confirmed that there had been an impressive impact on the firms' quality of products and skills.

Another example of the latter is a trade competitiveness project where the achievement of the overall goal was measured in terms of the higher average annual growth rate of exports during the project duration and increases in exports' share of GDP compared with the initial year of the project.

To be fair, task managers of trade-related projects are often candid about their project's achievements, writing in the ICR that observable results (particularly those relating to aggregate outcomes such as total exports) are not entirely the result of the programme alone, but rather the result of the work

---

[20]Our analysis of project ICRs did reveal, however, that the use of these methods is often handicapped by difficulty in quantifying some of the costs and benefits of the project. Some project ICRs explicitly say that certain benefits are not incorporated in net present value calculations due to their complexity.

Impact	Effects partly or exclusively attributable to the project
Outcomes	Welfare effects on target group directly attributable to the project
Outputs	Physical goods and services produced by the project
Activities	Actions and tasks carried out to transform inputs into outputs
Inputs	Financial, human, and material resources required

Impact evaluation

Implementation monitoring

Figure 1.5: *From inputs to impact.*

and resources of different institutions and sectors. The most striking fact in Figure 1.4 is that only 6% of projects (5 out of 85 projects) included a rigorous impact evaluation, involving a proper comparison of the outcomes of project beneficiaries with those of a control group. But even in such cases, the impact-evaluation method raised certain issues, which we discuss in the next section.

A clarification should be made at this point concerning the link between evaluation methods of projects and the monitoring and evaluation (M&E) framework.[21] M&E is an important part of the design and implementation of World Bank lending projects and is the reason why, as mentioned at the beginning of this section, we would expect to obtain evidence on the effects of certain types of trade intervention from project-level analysis. M&E is based on performance indicators capturing the outputs, outcomes and impact of a project (discussed in ICRs).

These performance indicator categories are thought to be related according to the scheme shown in Figure 1.5. The scheme makes clear the distinction between considering outputs in general, and going one step further and also considering outcomes and the impact attributable to the project *per se*. One common concern with the M&E framework for World Bank projects is that it often focuses too much on the monitoring part and not enough on the evaluation part. For example, most projects include exports as impact indicators but

---

[21] This discussion draws heavily on the aforementioned study by Steven Gunawan.

do not include a proper impact-evaluation strategy that allows for attribution to the project of an increase in exports.

In addition to World Bank investment lending projects, the World Bank also produces a large amount of analytical work—economic and sector work— where we could expect to find evidence that supports certain trade interventions. The key trade-related analytical pieces—diagnostic trade integration studies—do highlight the high costs of producing goods and services for export, and for delivering them to foreign markets, as being the major barriers to trade integration in less developed countries, and point to infrastructure as the most pressing constraint. But they do not inform the development community about which interventions work and which do not, and which interventions are most cost-effective.

4   CHALLENGES TO THE EVALUATION OF TRADE INTERVENTIONS

4.1   Striving for Internal and External Validity

The key problem that IE addresses is *attribution*: making sure that observed changes in outcome variables are caused by the programme or policy under evaluation and not by outside influences. Many outside influences can confound the identification of a programme or policy's impact. For instance, an export promotion scheme put in place in 2007 would see its positive impact confounded by the negative impact of the global crisis of 2008–9; a simple before–after comparison of outcomes is likely to suggest a *negative* impact of the programme.

In order to filter out these influences, we would want to know how beneficiary firms would have performed in the absence of the programme (presumably worse). But the data needed for this counterfactual does not exist, because firms cannot be both beneficiaries and non-beneficiaries at the same time. This missing data problem is solved by using as a counterfactual the performance of other firms that did not benefit from the programme. By analogy with medical sciences, where IE methods originate, beneficiaries are called the *treatment group* and non-beneficiaries the *control group*.[22]

The central idea of IE is best illustrated by a widely used technique called double-differences or difference-in-differences. Using this technique, the effect of a programme is assessed by comparing the performance of beneficiary firms before and after the treatment (first difference), and then benchmarking that difference by comparing it with the difference in performance

---

[22]A pedagogical reference to IE techniques can be found in Khandker *et al* (2010), which contains analytical guidance as well as case studies and STATA do-files. A formal treatment can be found in Blundell and Dias (2002).

Table 1.2: *Boundaries of impact evaluation.*

	Evaluation built into programme design	Evaluation not built into programme design
Naturally targeted (typically trade competitiveness related *eg* matching grants for producers for technology upgrading or export business plans; export credit guarantees for producers)	RCT is feasible; quasi-experimental methods are a possible alternative	RCT is infeasible; quasi-experimental methods are feasible
Naturally non-targeted (typically trade facilitation related, *eg* customs reform, port improvements), but also some trade competitiveness related (support for producer organisations or other institutional reforms)	RCT is typically infeasible; quasi-experimental methods are more appropriate; some methods of targeting can be introduced (phase-in, staggered implementation)	All IE methods are difficult; before–after comparisons may be only alternative

Notes: RCT, randomised controlled trial. Quasi-experimental methods are matching, difference-in-differences, instrumental variables or regression discontinuity design.

over the same period of non-beneficiary firms.[23] In our earlier example of an export promotion scheme put in place just before the onset of the global financial crisis, its confounding effect would be captured (and thus filtered out) by the decrease in the performance of non-beneficiary firms during the programme period. The programme's impact would then be measured by how much less badly beneficiary firms did than non-beneficiary ones.

As noted earlier, IE design relies on less-than-universal coverage, which provides a first categorisation into naturally targeted and naturally non-targeted programmes. Another useful distinction is whether an evaluation is built into programme design. In what follows, we consider each of the cases defined in Table 1.2 in the trade context, and discuss the extent to which IE methods can be applied to them. Anticipating our conclusions, our basic argument is

---

[23]This difference in performance is not a *ceteris paribus* effect: it picks up both direct programme effects and induced behavioural changes, which may work to either reinforce or weaken the programme's direct effect. For instance, a programme combining matching grants with technical assistance targeted at particular operations within the firm can trigger broader management improvements (a reinforcing influence) or partial waste of programme money through management slack (a mitigating influence). See Duflo *et al* (2008) for a discussion.

that the scope for IE in trade assistance projects is broader than might at first appear, provided that we are not wedded to a particular methodology (RCTs, for instance).

4.2 Naturally Targeted Interventions

Naturally targeted trade interventions include 'clinical' trade competitiveness programmes such as export promotion schemes through matching grants for supporting export business plans, through export-credit guarantees, or through firm-level technical assistance for technology upgrading, for acquisition of international quality certifications or to meet other product standards. Because these interventions operate at the level of the firm, non-assisted firms can in principle serve as the control group.

Randomised Control Trials

In naturally targeted interventions, *when evaluation is built into programme design*, a randomised controlled trial, sometimes called the 'gold standard' of IE, is the best option. It consists of drawing beneficiaries at random from a large pool of firms. By the law of large numbers, the average characteristics of beneficiaries will be the same as those of non-beneficiaries. Were this condition not met, there would be a selection bias; that is, the programme's impact would be confounded not by outside factors, as before, but by differences in individual characteristics.[24]

Randomisation can be carried out in many ways, some better than others. For instance, it is better to randomise among firms that have applied to a programme if the decision to apply correlates with unobserved characteristics that may affect performance, like managerial ability. Randomisation requires that the programme be designed for IE at the outset, and we will discuss later in this chapter the difficulties that randomisation encounters, both in general and in the context of trade-related assistance.

In Chapter 4 of this volume, Atkin and Khandelwal describe an ongoing project for an RCT to assist microenterprises in the handloom weaving sector in Akhmeem, Upper Egypt, to enter into export markets. The project's objective is to link those microenterprises to foreign buyers in the USA through the provision of three kinds of service:

- putting Egyptian producers in contact with design consultants to develop patterns that can appeal to the tastes of US consumers;
- marketing assistance with US buyers;
- general business training.

---

[24]In other words, the probability of getting treatment, conditional on the firm's characteristics, should be independent of the outcome.

The project's impact-evaluation design is simple: after drawing up a list of potentially viable producers/exporters in the sector and region, a random group of them will be given the opportunity to export to the US market with the help of the three services listed above. The data on both outcomes (export performance) and covariates (producer characteristics) will be generated through surveys conducted as part of the IE. A baseline survey will collect information on all viable exporters—both those that will benefit from this intervention and those not approached before the services are provided. Another survey will be conducted long enough after the intervention in order for the effects to be tangible.

The World Bank is considering implementing RCTs in some of its own projects as well, although plans at this stage are preliminary. Candidate projects include a customs border-post modernisation project at the border between the Democratic Republic of Congo and Rwanda, where petty traders on foot, mostly women, are regularly exposed to corruption and harassment. The project would involve some of the women's associations (with group randomisation) to designate customs brokers acting as shields between women and predatory customs officers. Another project involves the facilitation of payments for small cross-border transactions through branchless banking near the Cameroon–Chad border. Currently, all payments for such transactions are made in cash, which hampers trade. At the very least, the project would involve a natural experiment if branchless banking is allowed for traders on one side of the border but not on the other; in addition, the design may involve, in a pilot phase, selected access to non-cash payments for a randomly chosen treatment group.

One of the reasons why RCT is a preferred design for such experiments is that randomisation does away with the need for complex econometric techniques to control for selection in non-experimental settings. However, RCT is no silver bullet in small sample environments, as it relies on the law of large numbers to ensure that expected untreated outcomes are equal in treatment and control groups. In low-income countries, interventions sometimes target very small numbers of firms (McKenzie 2011).[25] For instance, the Pesticides Initiative Program (PIP), a European Union (EU) technical-assistance programme designed to help fruit and vegetable producers cope with EU standards, covers less than a few dozen firms in some African countries (Jaud and Cadot 2011). Randomisation is not an option in such environments. Quasi-

---

[25]McKenzie (2011) discusses the issue of small samples in World Bank private sector support programmes in Africa. None of those programmes has been subject to rigorous impact evaluations so far, but if such evaluations were to be conducted, researchers would be faced with a serious problem of power given the small number of enterprises assisted by the projects and their large degree of heterogeneity.

experimental methods may not do very well either, but if a cross-country sample is available with enough observations, econometrics may offer some scope to control for cross-country heterogeneity.[26] We will return to this small sample issue in the context of naturally non-targeted interventions in Section 4.3 when referring to the Cameroon customs project described by Cantens *et al* in Chapter 7.

In terms of practical feasibility, randomisation can be a hard sell with client governments. Duflo *et al* (2008) note that the spread of RCTs in health, education and poverty reduction programmes owes much to the collaboration with non-governmental organisation (NGOs), as collaboration with local authorities is still relatively rare. NGOs are much less involved in trade-related programmes than in other programmes, so the scope for RCTs may be inherently less, at least as long as the evaluation culture remains rare in public policy. Atkin and Khandelwal discuss in Chapter 4 how carrying out an RCT within the context of international trade depends crucially on finding a suitable local project partner who can provide the export promoting services to producers, and on convincing such a project partner of the feasibility and value of the randomisation procedure.

Randomisation does allow for flexibility, which may help make it acceptable. First, it does not need to cover all individuals. For instance, a programme can use standard selection methods to determine eligibility, and introduce randomisation among either all eligible firms or only 'marginal' ones. That is, very strong candidates can be taken in, very weak ones left out and only those in the middle subject to randomisation.[27] Lotteries are somehow more appealing than blind randomisation because they avoid the impression that something is hidden. Alternatively, 'encouragement' designs can be used, whereby some firms, chosen at random, get more information (*eg* phone calls) than others, raising the probability that they apply to a programme.[28]

Quasi-Experimental Methods

When evaluation is not built into programme design, RCT is not an option and quasi-experimental methods must be used, all relying on econometric tech-

---

[26]Randomisation across countries would be more difficult to implement than within a country and would not necessarily increase the test's power.

[27]We are grateful to David McKenzie for pointing this out to us.

[28]Duflo *et al* (2006) offer a recent example of an encouragement design. They tested whether seeing a neighbour use fertilisers would encourage other farmers to do the same. For each using farmer, they invited randomly chosen neighbours to attend a demonstration of fertiliser use. Although other farmers were also welcome to attend, the attendance rate was much higher in the subsample of invited ones, which was randomised. To our knowledge, no trade intervention has been evaluated with encouragement design.

niques to overcome selection bias.[29] The first is the difference-in-differences (DID) method briefly described above. By comparing differences in outcomes instead of comparing levels, DID controls for unequal performance levels of treatment and control groups not related to the programme. However, DID relies on the assumption of parallel trends and does not control for selection on observables (firm-level covariates).

The DID method can be improved by matching, provided that selection into the programme is based on observable characteristics. Matching controls for observed firm-level characteristics correlating with both programme participation and performance. However, it does not control for unobserved characteristics. The matching procedure evolves in two steps. First, firm-level covariates are used to predict the probability of getting (or enrolling into) the programme using a probit or logit regression. This predicted probability is called a propensity score. Second, the control group is formed by picking, for each treated firm, the untreated firms with the closest propensity score. For each treated firm, depending on the method, there can be either one matched control firm or several, using a weighted scheme.[30] Average outcomes (in first differences) are then compared between the treatment group and the matched control group.

The studies surveyed by Volpe Martincus in Chapter 2 are good illustrations of the use of quasi-experimental methods in the evaluation of trade assistance. These studies, recently carried out at the Integration and Trade Sector of the Inter-American Development Bank, use DID and matching-DID methods to assess the effectiveness of export promotion activities of PROMPEX/PROMPERU (Peru), PROCOMER (Costa Rica), URUGUAY XXI (Uruguay), PROCHILE (Chile), EXPORTAR (Argentina) and PROEXPORT (Colombia). They use rich and unique data sets for the six Latin American countries that combine firm-level customs data with covariates drawn from other national firm-level data sources and constitute the first rigorous micro-based evidence of the effects of export promotion.[31] The picture emerging from Volpe Martincus's survey is that export promotion was effective in facilitating export expansion for firms in Latin America, but primarily along the extensive margin. Firms

---

[29]How well quasi-experimental methods perform compared with randomisation has been a subject of intense scrutiny since the seminal paper of Lalonde (1986), with largely inconclusive results. Glazerman *et al* (2003) found that quasi-experimental methods produced substantially biased results compared with experimental ones in 12 replication studies of welfare and employment programmes in the USA. Cook *et al* (2006) found less clear-cut results for education programmes.

[30]The single-match method is called 'nearest neighbour'. Alternatively, we can use n nearest neighbours, or the entire sample of untreated firms with weights that decrease with distance from the treated firm's propensity score. This latter method is called 'kernel matching'. Many other refinements are possible.

[31]An alternative, more traditional route to the evaluation of export-promotion's effectiveness is the aforementioned cross-country study of Lederman *et al* (2010a).

exporting differentiated goods benefit more than those selling more homogeneous goods. Small and relatively inexperienced companies benefit more than larger and already established exporters. Finally, bundled services that provide support to firms throughout the export-development process appear to be more effective than isolated actions.

In Chapter 3 of this volume, Gourdon *et al* apply the same type of quasi-experimental methods to the evaluation of FAMEX, a World Bank-supported export promotion programme in Tunisia, which provided a mixture of counselling and matching grants to new exporters. The study exploits a customised firm-level survey to estimate the effects of FAMEX on the export performance of beneficiary firms at the intensive and extensive margins. Propensity-score matching DID estimates suggest a very large and statistically significant growth effect at the intensive margin: a 39% differential in terms of annual export growth compared to control firms over the 2004–8 period. The treatment effect at the extensive margin, in terms of products and destinations, is both smaller quantitatively (a 5% growth differential in the count of products and destinations for programme beneficiaries compared with control firms) and of marginal or no significance (at 10% confidence level for destinations and insignificant for products). In addition to the observed acceleration in export growth, Gourdon *et al* find a significant boost to employment growth: a 10% annual differential for programme beneficiaries, significant at the 5% confidence level. An original feature of their data set is that it covers service firms in addition to manufacturing firms, and they find considerably stronger effects for the former. One potential issue with their data is that the survey was conducted *ex post* (no baseline survey was conducted, as IE was not part of the programme design) so the data may suffer from recall bias. Preliminary results by Cadot *et al* (2001) based on an alternative source of data (customs data) suggest a smaller and non-persistent treatment effect.

Jaud and Cadot (2011) also apply quasi-experimental methods to assess the impact of PIP on the export performance of firms in Senegal's horticulture sector. Their results suggest that, while the programme had no significant effect on exports of fresh fruit and vegetables pooled over all products and destinations, it had a positive effect when considering exports to the EU.

Other quasi-experimental methods can address selection bias in the evaluation of the impact of a programme. One approach relies on instrumental variable (IV) estimation. This can be used when programme take-up is less than complete and thought to be correlated with unobserved individual characteristics influencing performance. In this case, eligibility can be used as an instrument for participation, provided that eligibility is truly exogenous (*eg* if there is randomisation of eligibility but programme take-up is incomplete or some participants drop out). This method is used in the context of naturally non-targeted interventions by Sequeira in Chapter 5 of this volume, as described in Section 4.3.

Another approach is regression discontinuity design, which makes use of breaks in eligibility to identify a programme's impact.[32] For instance, suppose that an export promotion programme targets small and medium-sized enterprises (SMEs) as defined by a cut-off level of sales. If the sample is large enough, outcomes can be compared for SMEs immediately below the cut-off (eligible) and for SMEs immediately above (ineligible), on the assumption that they are close enough in the characteristic upon which eligibility is defined to be good matches for each other, and most importantly that the cut-off rule is indeed enforced.[33]

4.3 Naturally Non-Targeted Interventions

Naturally non-targeted trade interventions mostly cover programmes that help reduce trade costs. These include trade-facilitation programmes such as upgrading of bottleneck infrastructures in ports, roads or railways, reforms of customs agencies and procedures, and some types of trade competitiveness programmes related to general improvements in the business environment or support to producer organisations. Because these interventions generally do not target micro entities and their direct beneficiaries are multiple and diffuse, the identification of a control group is difficult, and so they are less amenable to experimental or quasi-experimental design.

Considering 'hard' and 'soft' infrastructure-related trade-facilitation programmes, the two key constraints to estimating their effects are the endogeneity of programme placement and the absence of well-defined treatment and control groups. Thus, the pre-treatment unobservable characteristics that determine infrastructure placement and affect outcomes are likely to differ between treatment and comparison groups (where groups are, in this case, most likely to be locations). Randomisation in the context of large and sensitive hard transport infrastructure programmes is generally not feasible. This is also the case for soft trade-facilitation programmes relating to rules, regulations and government agencies dealing with the movement of cargo across borders that are often not amenable to random assignment at the micro level nor to the creation of comparison groups for the purposes of an IE.

For interventions such as customs reform, the only way to generate a control group is to introduce elements of targeting through progressive phase-in during a pilot phase, staggered, for example, across different border posts, or through selective implementation covering only some customs offices or officials, or by giving privileged access only to some firms or to some types of traded goods. For instance, a 'green channel' in customs, which is a speedy

---

[32]See Campbell (1969) for details; a survey can be found in Todd (2006).

[33]The issue of rule enforcement has been a controversial one, *eg* in the context of microcredit evaluation (Morduch 1998), but may be less fraught for firm-level trade interventions like support to SMEs.

clearance for trusted operators, can be restricted and randomly allocated in an early phase, using non-eligible operators as controls.[34] In this case, methods such as DID can in principle be applied using the locations initially not covered and customs offices, officials or firms like the control group for the targeted entities.

However, in many cases, during the pilot phase the control group will not be strictly comparable to the treatment group. For example, when a border modernisation programme is initially deployed in one border post, other border posts of different scale and product mixes serving other areas could serve as controls. It may then be necessary to use regression analysis to control explicitly for the heterogeneity in covariates in estimating differences in outcomes between treated and control border posts.

In some cases, policy design or implementation inadvertently creates the conditions necessary to perform evaluation through quasi-experimental methods: what economists call a 'natural experiment'. Datt and Yang exploit one such natural experiment in Chapter 6. The government of the Philippines used pre-shipment inspection (PSI) services to combat corruption in customs and increase import duty collections. The natural experiment arose from two conditions:

- imports from only *some* countries of origin were covered by PSI, which created a natural control group (imports from other countries);
- in 1990 the Philippines government decided to close a loophole whereby import transactions below a threshold of US$5,000 were exempted from PSI.

The loophole had enabled traders to slice shipments into small batches and under-invoice them without being detected. The customs reform consisted of lowering the threshold to US$500, so the period after 1990 can be considered a 'treatment period'. A DID equation can then be used to compare the evolution of outcomes before and after the reform for the treatment and control group of countries. The DID estimates show that, when inspections were expanded to lower-valued shipments, imports shipments were no longer mis-valued, but those from treatment countries shifted differentially to an alternative duty-avoidance method: shipping via duty-exempt EPZs. Thus, increased enforcement reduced the targeted method of duty avoidance, but led to substantial displacement to an alternative duty-avoidance method. Duty collection failed to rise, while importers incurred higher fixed costs as they relocated to EPZs. This chapter shows that, to be successful, anti-corruption reforms need to encompass a wide range of possible alternative methods of combatting illegal activity.

---

[34]This approach is similar to so-called 'pipeline' methods, where applicants are used as controls for beneficiaries.

In Chapter 5 of this volume Sequeira discusses a transport infrastructure project consisting of investments in a railway connecting the economic heartland of South Africa to the port of Maputo in Mozambique. Given the poor state of Mozambique's infrastructure after two decades of war—and considering budget constraints—the government had to be selective in its choice of infrastructure investments. They decided to rehabilitate the old-pre-colonial railway in the Maputo transport corridor (which would promote regional integration) rather than building an entirely new north–south connection, as was demanded by the Mozambican business class. As the layout of the old-pre-colonial railways had been designed to serve 19th-century mining companies, there is plausible exogenous variation in the emergence of the rehabilitated railway relative to the geography of manufacturing and retail firms at the time of rehabilitation of the railway.

The IE of this transport infrastructure project estimates the impact of railway rehabilitation on firm performance—namely, changes in transport costs for different firms and sectors, how firms respond to these changes, and what the spillover and network effects are across rail and road transport. To identify a causal relationship, the study will use a quasi-experimental method, IV, where the treatment, defined as changes in transportation costs, will be instrumented by the distance between a firm's location and a working railway station.

In addition, the study exploits the fact that other transport corridors in Mozambique developed at different speeds, and identifies two sets of control firms to match to the treated firms in the Maputo transport corridor:

- firms in the Beira corridor (that have access to a new port but no railway) and
- firms in the Nacala corridor (that have no access to a new port or railway).

To isolate the impact of the Maputo railway rehabilitation, the study will use a matching DID estimation that assumes that the only factor making the trajectories of these three sets of firms different during the sample period is that they were exposed to different transport choice sets. The impact of the Maputo railway rehabilitation is not yet known, since only the baseline survey information is available; a follow-up survey will be conducted in spring 2011.

Sequeira also discusses a 'soft' transport infrastructure project, focusing on corruption in Southern African ports.[35] By collecting original data on bribe payments made to customs officials and to port operators in the two competing ports of Durban and Maputo, the study is able to trace differences

---

[35]The project is described more extensively in Djankov and Sequeira (2010).

in bribe schedules to the organisational structure of each port. By observing how firms adapt their shipping and sourcing decisions to the type of corruption faced at each port—which enters into the calculation of the overall cost of using each port—the study estimates the impact of corruption at ports on the behaviour of South African firms. The estimates show that corruption imposes a distortion in terms of 'diversion', *ie* firms travel on average an additional 322 kilometres, more than doubling their transport costs, just to avoid 'coercive' corruption at a port.

This effect is only observed for firms facing a higher probability of being coerced into a bribe because of the kind of product they ship. Firms are willing to incur higher costs to avoid corruption because of an aversion to the uncertainty surrounding bribe payments at the more corrupt port (Maputo). The uncertainty in Maputo seems linked to the short time periods caused by high job turnover among customs officials. Firms also respond to different types of corruption by adjusting their sourcing decisions for inputs, domestically or internationally, since corruption at ports increases the cost of using the port and thus directly affects the relative cost of imports.

While this project is not an impact evaluation of an intervention to reduce corruption in ports, it provides two sets of valuable insights on such interventions because it considers the entire chain between competing port bureaucracies setting bribes and user firms making shipping and sourcing decisions. First, the study shows that, depending on the type of corruption that bureaucrats engage in, bribes can affect the deadweight loss, tariff revenue and the demand for the public service. In particular, corruption seems to reduce significantly demand for the Maputo port, stifling the returns to the massive investments in hard infrastructure of the corridor that have taken place in recent years. Second, policy changes to the organisation of ports and to the nature of the interaction between shippers and port officials could reduce corruption. Such changes include reducing the discretion of port officials in the clearance process, and eliminating face-to-face interactions between clearing agents and port officials.

Cantens *et al* describe in Chapter 7 of this volume a recent pilot for customs reform in Cameroon that involved the introduction of contracts with performance indicators for front-line customs inspectors in two of the country's customs bureaus (henceforth referred to as 'treated bureaus'). The performance indicators covered both trade facilitation and the fight against fraud and bad practices. Front-line customs inspectors with good performance would be rewarded with non-financial incentives such as congratulatory letters entered into their personnel files, easier access to the director general of customs, training courses and transfers to more attractive bureaus. Poorly performing inspectors would be sanctioned by eviction from bureaus with strong 'fiscal potential', that is, where the possibilities of earning money legally through disputed claims were high.

   This project is an interesting example of a trade intervention that in principle is naturally non-targeted, but where targeting could have been introduced by focusing on a subset of front-line customs inspectors. This could then have been an ideal setting to implement an RCT, whereby a subset of randomly chosen front-line inspectors would have been under performance contracts, while others would not. However, it was not possible to implement an RCT for several reasons. First, the seven customs bureaus in Cameroon are specialised (oil imports, special customs regimes related to public trends, transit, exports, bulk cargo and the two treated bureaus) and differ so much in customs practices that it would be difficult to make comparisons across bureaus. Hence, if anything, a bureau would need to be split into a treated group and a control group of front-line inspectors. But this was not feasible given a small sample problem: less than ten staff work in each bureau. Second, as is generally the case in projects funded by governments or international donors, the time devoted to the pilot project was limited. Thus, it was not possible to overcome the small sample issue by allowing for turnover within each bureau to artificially increase the number of treated and control officers. Moreover, since contract incentives were not financial, time was required to reward good performers (*eg* it would not have been feasible to appoint high-performing inspectors to better positions every six months).

   Therefore, the IE of the customs performance contracts project was conducted as a comparison of inspectors' behaviour before and after the project was implemented, without a defined control group, although the impact on clearance times was assessed using the bulk-cargo import bureau as a counterfactual. The estimated effects of the pilot performance contracts were positive surprisingly soon after the pilot was launched in mid 2009. Duties and taxes assessed increased despite a fall in the number of imported containers (likely to be linked to the financial crisis), and the tax yield of the declarations also rose. The performance contracts also affected clearance times, as the share of declarations treated within 24 hours increased more in the treated bureaus than in the counterfactual bureau, and the variance of clearance times decreased dramatically. The impact on disputed claims was equally interesting, with inspectors abandoning low-level disputed claims to focus on major ones, and the ratio of taxes adjusted to taxes assessed increased. Finally, the contracts also had a major impact in reducing costly practices. For instance, the number of litigious reroutings from the yellow channel (documents control) to the red channel (physical inspection) declined tremendously.

5   DATA ISSUES

5.1   *What Are We Measuring?*

The choice of performance measures is important not only to ensure that IE focuses on the appropriate indicators, but also because using IE can affect

the incentives of agents and programme managers in unintended ways. Performance indicators that strongly relate to targeted interventions in a causal sense are often too technical to be of interest from a broad policy perspective, whereas the highly aggregate indicators that interest policymakers are rarely faithful reflections of the effect of targeted interventions and projects. Thus, selecting performance indicators involves a trade-off between breadth and identification.

Much of the talk in AfT evaluation focuses on aggregate indicators such as national export performance or other macro variables. Although policymakers may find these broad indicators relevant, the causal link between them and the actual performance of trade interventions is tenuous, implying weak identification.

By contrast, M&E frameworks, developed to ensure project management and quality control, have used intermediate outcomes more directly linked to the projects themselves, like customs clearance times. In a causal sense, these measures are closer to project management but are likely to be narrow in scope. Deciding which approach is better depends on what the indicators are used for. If evaluation results are expected to feed into incentive structures for programme managers, identification is critical and breadth is secondary. In contrast, in order to catch the attention of policymakers, breadth matters more, possibly at the cost of weaker identification.

Impact evaluation does not escape this general trade-off between breadth and identification, but is typically located at the 'narrow' end of the spectrum, since it identifies changes in performance measures that are directly attributable to the project. For instance, when evaluating a customs modernisation programme, the performance measure is likely to be something like container dwell time, even though less quantifiable dimensions of customs performance, like security at the borders, may also matter.

But identifying and documenting the chain of causality from programme to ultimate outcomes can be challenging for some trade interventions. In trade-facilitation programmes it is not always clear what the micro-level mechanisms are by which transport costs reductions influence firms and households and, more generally, economic activity.

In addition, the use of IE can affect incentives in the long run. The focus on narrow, immediate performance outcomes may well lead to measurement biases or, even worse, create perverse incentives when used for monitoring and evaluation. For one thing, it can focus attention on readily measured outcomes at the expense of less easily measured ones.

Consider a customs modernisation programme. Using IE results to design reward schemes for customs officials might lead to over-emphasis on easy-to-measure reductions in clearance times, at the expense of the monitoring of suspect shipments. If, say, there is a low rate of smuggling illicit products, it may take time before the consequences of reduced monitoring get noticed: too long to show up in an IE.

Table 1.3: *Intermediate and ultimate performance outcomes.*

	Trade competitiveness	Trade facilitation
	Example of programme: matching grant to support firms access export markets	Example of programme: customs reform
Intermediate outcomes to understand the chain of causality from programme to outcomes	Exports, output, input choices at firm level	Customs or port clearance time and costs, incidence of illegal activity
Ultimate outcomes	Productivity, wages, employment at firm level	Trade volumes, customs revenue collected
Covariates to use as controls or to understand the heterogeneity of effects of programme	Firm-related industry, location, age, size, ownership, workforce details	Firm-related or customs office or official-related: location, education, age, contract

5.2   Getting the Data

Table 1.3 provides examples of intermediate and ultimate outcomes in the context of new-style trade interventions linked to trade competitiveness and trade facilitation.

The feasibility of rigorous impact evaluation hinges critically on data availability. Whether the IE is based on experimental (RCTs) or quasi-experimental design, it needs to include a baseline survey and at least one follow-up survey. If quasi-experimental methods are used, the baseline survey must include a rich set of covariates to estimate a (first-stage) selection regression. One of the advantages of RCTs, especially in developing countries, is that they are less demanding in terms of data; however, even with randomisation, firm-level covariates can be useful in verifying that the treatment and control groups are comparable in their observable characteristics. This is especially important for small samples. Moreover, the availability of a rich set of covariates allows for the analysis of heterogeneity in the effects of the programme.

The evaluations of the impact of trade-facilitation programmes—especially those related to infrastructure—suffer currently from a serious lack of micro-data on transport costs and prices before and after interventions take place. For these types of intervention, it is imperative to conduct baseline and follow-up surveys of programme beneficiaries and control groups even though they can be costly.

In addition, baseline and follow-up surveys may not be enough to assess a programme's impact. Consider, for example, the case of a one-year export-

promotion programme, where firms can enlist in any year between 2005 and 2009; and then suppose that a baseline survey is conducted in 2004 and a follow-up survey is conducted in 2010. For firms that enrolled in 2005, the follow-up survey will pick up outcomes four years after the treatment. By then, if the effects are transient, they may have vanished, and the follow-up survey will pick up heterogeneous effects (one year after treatment for firms enrolled in 2009, two years for those enrolled in 2008 and so on). Thus, although costly, it may be necessary to run repeated follow-up surveys year after year.

While World Bank projects typically have budgets for baseline data collection, these may not always be enough to gather the data needed for a proper IE evaluation after the project is completed. An alternative cost-efficient method is to use official pre-existing sources of data, provided that they are collected often enough and can be reconciled with programme data. For example, for an export promotion programme, customs records at the transaction/firm level can be used to measure outcomes such as growth in export value (the intensive margin), number of products and number of destinations (the extensive margin).[36]

The trade and integration unit of the World Bank Development Research Group is involved in a major data-collection exercise that may help IE of trade-related interventions in the next few years. As described in Freund and Pierola (2011), the exercise consists of the collection and compilation of the first ever database on exporter-level customs transaction data across countries and over time. Data has been obtained for 20 countries in Africa, Asia, Eastern Europe and Latin America, and negotiations are in progress to obtain data for 25 more countries. The database will include statistics on exporters' characteristics and behaviour by country, industry and destination market. The purpose of the database is to provide policymakers, development agencies, researchers and the public with a novel source of information to conduct analysis of export growth at the micro level and allow for the evaluation of programmes and policies affecting that growth.

Data on firm characteristics (covariates), used to control for selection bias, is typically hard to obtain. If an industrial survey is available in the country where the trade intervention is taking place, it can provide the required variables (*eg* location, age of the firm, education of its head, number of employees, foreign ownership). However, this requires that customs and industrial census data be merged, which raises confidentiality issues. When data is not available, an alternative is to conduct a 'retrospective survey', although this method may be biased. In its evaluation of Tunisia's export-promotion agency, Chapter 3 by Gourdon *et al* uses a combination of data from their own survey and from national sources (customs and national statistics institute).

---

[36]See Freund and Pierola (2010) and Lederman *et al* (2010b) for uses of such data in a non-IE context.

6   LOOKING AHEAD: CHALLENGES FACING IE OF TRADE ASSISTANCE

6.1   External Validity and Cost

One key concern with impact evaluations is that their external validity is an act of faith. When a programme is found to be effective (or ineffective), how do we know that the result would carry over to other programmes run in different environments?

As Rodrik (2008) and Ravallion (2008) argued, IE does not produce any harder evidence than traditional econometric methods. Instead, there is a trade-off in policy evaluation between external and internal validity. As traditional identification of causal effects through instrumental-variable strategies never completely eliminates confounding influences, these strategies always suffer from an internal-validity problem. However, when based on cross-country evidence, they pick up average effects that can be relatively stable—provided they are consistent with some sort of theory—because induction, even on cross-country samples, may fail to produce generalised results.

By contrast, IE purges confounding influences, but generates results that are empirical and case dependent. Such results may fail to carry over to different settings. Limited external validity of any study would not be a problem if we could replicate it easily. With enough replications, the sheer mass of evidence would provide the desired generality (although the method would still be inductive and would thus suffer from the general critism of inductive methods in science).

But some kinds of IE can be costly. For instance, the World Bank reckons that household surveys cost on average US$300 per household. At that rate, a baseline and final survey of 500 households would cost US$300,000. This is a lot for studies for only internal validity.[37] However, costs can often be contained by working with local institutions, which has the added advantage of building capacity in a key area.

Some trade-related programmes target limited numbers of firms, so their evaluation is less costly than that of poverty-reduction programmes. For instance, in a middle-income country, the cost of surveying 500 firms can be substantially lower than US$100,000. Moreover, the data often exist prior to and independently of the IE in the form of census or industrial surveys and customs records. In that case, the cost of the IE goes down dramatically.

---

[37]In their discussion of quasi-experimental versus experimental methods, Duflo *et al* (2008) make a noteworthy point about the commitment value of costly experimental design. It has often been argued, with some statistical support (see, for example, Ashenfelter *et al* 1999), that statistically significant results (positive impact in our setting) are more likely to get published, a so-called 'publication bias'. As experimental methods are costly and usually planned with donors, self-censure in the face of insignificant results is less likely to be feasible than when relatively low-cost quasi-experimental methods are used with publicly available data. In that sense, IEs may be less affected by publication bias.

The problem then is no longer one of cost but of securing buy-in from the agencies possessing the data so that they share it.

6.2 Can Treatment Effects Be Used to Justify Government Intervention?

Externalities can bias treatment effects by blurring or magnifying the difference in outcomes between treatment and control groups. In the context of policy evaluation, this raises a deep issue, as externalities are the basic justification for government intervention.

One key assumption of both experimental and quasi-experimental methods is that the control group is not 'polluted' by the treatment group, lest the comparison of outcomes be biased. A classic case in economics occurs when general equilibrium effects transmit the benefits conferred on beneficiaries to non-beneficiaries or, alternatively, penalise them, say, through rising input prices.

In the evaluation of trade-related programmes of limited scale, such as export promotion or trade facilitation, general equilibrium effects may not be critical. However, spillovers may be present through other channels; for example, an export promotion programme may have 'demonstration effects' yielding valuable information on the viability of destinations or product markets that can be easily imitated by non-participants. In such circumstances, the estimated treatment effect will be biased downward because the difference in outcomes between the treatment and control groups will measure only the purely private effect.

Conversely, a programme to upgrade one border post may induce traffic shifting from other, untreated border posts. The volume of trade going through the treated border post will then be increased by the substitution from traffic that normally goes through other ones, and as 'control' border posts see their traffic go down, using them as controls will result in a bias in the estimated treatment effect of the programme.

Not surprisingly, treatment effects may be contaminated by information externalities. After all, even the most rigorous RCTs used to test the effectiveness of drugs can also be affected by informational biases, as is the case when individuals in the control group observe that they do not suffer a drug's side-effects while individuals in the treatment group do, and as a result infer that they received the placebo instead of the treatment. But in economics, the presence of externalities takes on special importance because it plays a key role in justifying government intervention. If the benefits of a programme whose costs are borne by taxpayers were internalised by the beneficiaries, surely those beneficiaries ought to pay for it and there would be no justification for public intervention. In contrast, if a programme generated spillovers so powerful that no treatment effect was detectable, *ie* the control group indirectly benefits as much from the programme as the beneficiaries—then the

argument for a public intervention could be right, as beneficiaries would be willing to pay *nothing* for the treatment.[38]

Thus, seeking to justify government-financed programmes solely on the basis of treatment effects may not just be affected by bias, it may be altogether wrongheaded. In the export promotion scheme example, what IE would be measuring is only the private-good dimension of the intervention; the public-good dimension would be left unevaluated.

It is thus important to disentangle whether a no-effect finding is due to externalities or to programme ineffectiveness. This may call for an independent effort, aside from the IE itself, to detect the presence of externalities. For instance, a regression of outcomes of untreated individuals might be run on some continuous measure of exposure/closeness to treated individuals, to see if more, or closer, treated neighbours raise the outcome of untreated ones.

Alternatively, we can include this same measure of exposure to (other) treated individuals in the DID equation and interact it with the treatment to see if the treatment is more powerful on individuals 'surrounded' (in some economic sense) by other treated individuals. These methods are inspired by measures of contagion used in epidemiological studies.

In contrast with medical sciences, however, in social sciences the mechanisms by which contagion takes place are largely unknown. Baseline surveys may help in understanding and identifying channels through which future programme benefits might spread from one firm to another, *eg* professional association memberships, personal contacts and so on.

6.3   Mainstreaming IE into Trade Interventions

In spite of the challenges, rising demands for results and accountability from donors and clients alike require that AfT evaluation strategies need more ambition and rigour. Implementing agencies should no longer be content with traditional methods based on output monitoring and before–after comparisons. Output monitoring is largely introspective, relying on measures defined by the task managers and therefore liable to biases, while before–after comparisons are vulnerable to confounding influences.

The basic problem faced in the evaluation of a policy, programme or project impact is *attribution*. Are the observed changes in the performance of treated entities really attributable to the intervention under consideration, or do they reflect a fortuitous combination of effects? IE methods—developed outside of the social sciences but widely adopted in the evaluation of poverty-reduction, health and education programmes—provide a generally accepted answer to the problem of attribution.

But trade interventions have so far escaped the rising tide of evaluation methods. And, as this book tries to show, there is no justification for this

---

[38]This point was made to the authors by Daniel Lederman.

'trade exceptionalism'. IE techniques are many and provide numerous flexibilities for use in the case of interventions not 'naturally targeted' at a defined group of treated individuals.

As the authors have experienced in their campaign for greater recourse to IE techniques in trade, the key barriers to progress are not conceptual. Rather, they concern incentive issues, as IEs are costly, burdensome, lengthy and not necessarily aligned with project managers' incentives. For example, World Bank projects to assist private sector firms in Africa last, on average, five years, which would imply that, if their IE involved an RCT, many years would need to elapse for the projects to show results (McKenzie 2011). These many years would go well beyond a project manager's horizon.

However, researchers would not need to wait until completion of the project to evaluate its effects; rather, results one or two years after the project could be assessed and used to guide the implementation of the project in the subsequent years.

The weakness of current evaluation practice can be illustrated no better than by this critical assessment, found in the Implementation Completion Report of a recent World Bank project in the area of export promotion:

> Although the design of the M&E system was appropriate, both Bank and Government project teams had difficulty measuring the achievements of the project using the broad indicators cited in the PAD. [B]y current standards, they were insufficient and incomplete.
>   … M&E, particularly important as a learning objective, was weak. It was slow to start and did not deliver. The M&E staff … lacked the capacity and experience to carry out the monitoring activities, and the Unit was unable to carry out baseline and impact surveys of randomly selected farmers in both project and non-project areas, *ie* survey to gauge key interest groups' response to the outputs generated by the pilot activities. The M&E Unit's ability to collaborate with other implementing agencies to collect information and data was also ineffective. Implementing partners did not regard the M&E exercise as a learning process but instead, conducted their promotion activities without consulting or collaborating with the M&E unit.

As the reviewers noted, the learning function of evaluation escapes implementing agencies, being overshadowed by the 'monitoring' function.

In order to overcome these hurdles, several avenues must be explored. First, the burden imposed on project managers should be relieved by making impact evaluation a separate exercise, carried out by specialists, albeit in collaboration with project managers. Project managers should be involved at the right time, *ie* during project design, and from then on, as much as possible, left in peace. The World Bank has moved in this direction through the creation of the DIME unit, which provides expertise and help with IE financing.

At the same time, governments in the countries receiving trade assistance must buy into the process. This means sharing knowledge and building capacities for a proper interpretation of IE results and, in the long run, for govern-

ments to build their own IE capabilities as part of public-services delivery improvements.

Also, every effort should be made to reduce the cost of IEs. For small-scale activities, the cost of an IE can be as great as that of the activity itself. This is excessive. Local resources—in particular, universities and graduate students—should be involved, producing a double benefit: costs are reduced and local capacities are strengthened.

Finally, the exploitation of IE results should prioritise learning over monitoring. That is, donors and implementing agencies should tread cautiously in using IE results to frame incentive systems. Care is needed in the interpretation of IE results because premature conclusions could easily provoke a backlash and because a considerable accumulation of evidence is needed to yield truly valuable new knowledge.

Olivier Cadot is a Senior Trade Economist in the International Trade Department, Poverty Reduction and Economic Management Network at the World Bank.

Ana M. Fernandes is a Senior Economist in the Trade and International Integration Team in the Development Research Group at the World Bank.

Julien Gourdon is a Consultant in the Social & Economic Development Group for the Middle East and North Africa Region at the World Bank.

Aaditya Mattoo is the Manager of the Trade and International Integration Team in the Development Research Group at the World Bank.

REFERENCES

Ashenfelter, O., C. Harmon, and H. Oosterbeek (1999). A review of estimates of the schooling/earnings relationship. *Labour Economics 6*, 453–470.

Banerjee, A., S. Jacob, M. Kremer, J. Lanjouw, and P. Lanjouw (2005). Moving to universal education: costs and trade-offs. Mimeo, MIT, Cambridge, MA.

Banerjee, A., A. Amsden, R. Bates, J. Bhagwati, and N. Stern (2007). *Making Aid Work.* Cambridge, MA: MIT Press.

Banerjee, A., and E. Duflo (2008). The experimental approach to development economics. NBER Working Paper 14467.

Blundell, R., and M. Dias (2002). Alternative approaches to evaluation in empirical macroeconomics. CEMMAP Working Paper CWP10/02.

Brenton, P., and E. von Uexkuhll (2009). Product-specific technical assistance for exports; has it been effective? *Journal of International Trade and Economic Development* **18**, 235–254.

Bruhn, M. (2011). License to sell: the effect of business registration reform on entrepreneurial activity in Mexico. *Review of Economics and Statistics* **93**, 382–386.

Cadot, O., A. Fernandes, J. Gourdon, and A. Mattoo (2011). An evaluation of Tunisia's export promotion program. Mimeo, World Bank.

Calì, M., and D. te Velde (2011). Does aid for trade really improve trade performance? *World Development* **39**(5), 725–740.

Campbell, D. (1969). Reforms as experiments. *American Psychologist* **24**, 407–429.

Cook, T. D., W. Shadish, and V. Wong (2006). Within study comparisons of experiments and non-experiments: can they help decide on evaluation policy? Mimeo, Northwestern University, Evanston, IL.

Djankov, S., C. Freund, and C. Pham (2010). Trading on time. *Review of Economics and Statistics* **92**(1), 166-173.

Djankov, S., and S. Sequeira (2010). An empirical study of corruption at ports. Mimeo, London School of Economics.

Duflo, E., M. Kremer, and J. Robinson (2006). Understanding technology adoption: fertilizer in Western Kenya. Preliminary results from field experiments. Mimeo, MIT, Cambridge, MA.

Duflo, E., R. Glennerster, and M. Kremer (2008). Using randomization in development economics research: a toolkit. *Handbook of Development Economics* **4**, 389–392.

Freund, C., and M. Pierola (2010). Export entrepreneurs: evidence from Peru. World Bank Policy Research Working Paper 5407.

Freund, C., and M. Pierola (2011). Global patterns in exporter entry and exit. Mimeo, World Bank.

Gine, X., and I. Love (2010). Do reorganization costs matter for efficiency? Evidence from a bankruptcy reform in Colombia. *Journal of Law and Economics* **53**(4), 833–864.

Glazerman, S., D. Levy, and D. Myers (2003). *Nonexperimental Replications of Social Experiments: A Systematic Review.* Princeton, NJ: Mathematica Policy Research, Inc.

Harrison, A., and A. Rodriguez-Clare (2010). Trade, foreign investment, and industrial policy for developing countries, *Handbook of Development Economics*, Volume 5, Chapter 63. Amsterdam: Elsevier.

IEG (2006). Assessing World Bank support for trade, 1987–2004: an IEG evaluation. Report, World Bank, Washington, DC. http://go.worldbank.org/5T55SG8ZD1.

Jaud, M., and O. Cadot (2011). A second look at the pesticides initiative program: evidence from Senegal. World Bank Policy Research Working Paper 5635.

Khandker, S., G. Koolwal, and H. Samad (2010). *Handbook on Impact Evaluation.* World Bank.

Klapper, L., and I. Love (2010). The impact of business environment reforms on new firm registration. World Bank Policy Research Working Paper 5493.

Lalonde, R. (1986). Evaluating the econometric evaluations of training programs using experimental data. *American Economic Review* **76**, 602–620.

Lederman, D., M. Olarreaga, and L. Payton (2010a). Export promotion agencies revisited. *Journal of Development Economics* **91**, 257–265.

Lederman, D., A. Rodriguez-Clare, and D. Yi Xu (2010b). Entrepreneurship and the extensive margin in export growth: a microeconomic accounting of Costa Rica's export growth during 1997–2007. World Bank Policy Research Working Paper 5376.

Lopez-Acevedo, G., and M. Tinajero (2010). Mexico: Impact Evaluation of SME Programs Using Panel Firm Data. World Bank Policy Research Working Paper 5186.

McKenzie, D. (2010). Impact assessments in finance and private-sector development: what have we learned and what should we learn? *World Bank Research Observer* **25**, 209–233.

McKenzie, D. (2011). How can we learn whether firm policies are working in Africa? Challenges (and solutions?) for experiments and structural models. World Bank Policy Research Working Paper 5632.

Miguel, E., and M. Kremer (2004). Worms: identifying impacts on education and health in the presence of treatment externalities. *Econometrica* **72**, 159–217.

Morduch, J. (1998). Does microfinance really help the poor? new evidence from flag-ship programs in bangladesh. Princeton University, Woodrow Wilson School of Public and International Affairs, Research Program in Development Studies Working Paper 198.

Nelson, D., and S. J. Silva (2008). Does aid cause trade? Evidence from an asymmetric gravity model. University of Nottingham Research Paper 2008/21.

Osei, R., O. Morrissey, and T. Lloyd (2004). The nature of aid and trade relationships. *European Journal of Development Research* **16**, 354–374.

Rajan, R., and A. Subramanian (2008). Aid and growth: what does the cross-country evidence really show? *Review of Economics and Statistics* **90**, 643–665.

Ravallion, M. (2008). *Evaluation in the Practice of Development*. World Bank.

Rodrik, D. (2008). The new development economics: we shall experiment, but shall we learn? Mimeo, John F. Kennedy School of Government, Harvard University.

Rosenbaum, P., and D. Rubin (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika 70*, 41–55.

Savedoff, W. (2006). *The Evaluation Gap: An International Initiative to Build Knowledge.* Washington, DC: Center for Global Development.

Tan, H. (2009). Evaluating SME support programs in Chile using panel firm data. World Bank Policy Research Working Paper 5082.

Todd, P. (2006). Evaluating social programs with endogenous program placement and self selection of the treated. *Handbook of Development Economics* **4**, 3848–3891.

Wagner, D. (2003). Aid and trade: an empirical study. *Journal of the Japanese and International Economies* **17**, 153–173.

World Bank (2009). Unlocking global opportunities: the aid for trade program of the World Bank Group. Report. http://go.worldbank.org/CJVEJIR2J0.

World Bank (2011). *Leveraging Trade for Development and Growth: The World Bank Group Trade Strategy, 2011–2021*. World Bank.

2

Assessing the Impact of Trade Promotion in Latin America

CHRISTIAN VOLPE MARTINCUS[1]

1 INTRODUCTION

Factors shaping firms' (and thereby countries') trade performance can be classified into two interrelated groups: those affecting conditions under which production activities are developed within the countries, and those affecting conditions under which the output of these activities can be moved between countries.[2] The second set of factors can be generically bundled as *trade costs*. Trade costs are all the costs incurred in getting a good to the final user other than the marginal cost of producing the good itself (Anderson and van Wincoop 2004). In considering possible trade cost-related explanations of trade outcomes, tariffs and non-tariff impediments, as well as transport costs (both freight and time costs), appear as natural candidates.[3]

But other subtler, less studied barriers can also present deterrents to trade. Without denying the relevance of other obstacles, it can be argued that one such barrier to trade is lack of information. Information and the search for it play a vast role in economic life (Stigler 1961). Relevant, accurate and timely information is a key input to effective marketing decisions.[4] Given the

---

[1] The views and interpretation in this chapter are strictly those of the author and should not be attributed to the Inter-American Development Bank, its executive directors or its member countries. Other usual disclaimers also apply.

[2] The first set of factors has been examined extensively elsewhere (see, for example, Pagés-Serra 2010).

[3] A recent study shows that nowadays transport costs are higher than tariffs over several country and sector dimensions and appear to have a significant negative impact on both Latin American and Caribbean countries' total exports and their diversification (Mesquita Moreira *et al* 2008).

[4] Information is necessary for understanding the marketplace in which the firms intend to operate, and for monitoring changes in rapidly shifting business environments, designing reliable marketing plans and strategies, finding solutions to specific marketing problems such as changing prices, setting up distribution channels and choosing effective means for promoting products (Leonidou and Theodosiu 2004).

variety of business environments, the many factors to be considered when selling abroad and, in particular, the need to deal with situations not encountered in domestic operations, information is especially important for firms operating beyond national boundaries (Johanson and Vahlne 1977; Czinkota and Ronkainen 2001; Leonidou and Theodosiu 2004). Among other things, firms must know the formal export process at home, the different ways of shipping the merchandise and their associated costs, the potential markets abroad and their demand profile, the conditions for entering these markets and the channels available to raise awareness of their products and to market these products, and they must find specific business partners. In this latter regard, supply–demand signals cannot frequently be sent (or received) cost-effectively to (or from) potential exchange partners independently or via market mechanisms. Thus, firms pursuing cross-border economic opportunities must engage in a costly process of identifying and assessing the reliability, trustworthiness, timeliness and capabilities of these partners.[5]

In short, information gaps limit the ability of firms to learn about international trading opportunities and find a suitable trade partner, and in this way negatively affect exports.[6] In particular, exporters intending to enter a new market or expand foreign sales within an already served market are preceded by their reputation, which, in the absence of an identifiable brand name, largely depends on the perception of country of origin (Chisik 2003). This issue is especially important for firms from developing countries, whose products are more likely to be perceived as technologically less advanced and of poorer quality than those of peers from developed countries (see, for example, Chiang and Masson 1988; Egan and Mody 1992; Han and Terpstra 1988; Hudson and Jones 2003). This would especially be the case if consumers attach informational value to quantity and accordingly interpret low market shares as a signal of low quality (Caminal and Vives 1996).

How important are information obstacles? While, unlike other trade costs, such as tariffs and transport costs, there is no direct measure of the relative

---

[5]The difficulty of this search is determined by the extent to which economic opportunities and potential trading partners are geographically dispersed, whereas the importance of deliberations increases with the cost of reversing business decisions or their effects (Rangan and Lawrence 1999; Rangan 2000). Hence, the search and deliberation processes that must precede entry into a new export market usually require face-to-face contacts to coordinate business activities (Storper and Venables 2004). Interviews with managers of purchasing firms have confirmed the importance of face-to-face contacts. These executives put a high priority on capable management when deciding among alternative partners, and for this reason tend to visit their counterparts in their place of business before establishing a trade relationship (Egan and Mody 1992).

[6]See Rauch and Casella (2003); Suárez-Ortega (2003) and Chen (2004). Lack of information can be a particularly significant barrier to trade when uncertainty aversion is a factor. It has been shown that countries whose business communities are uncertainty averse, and thus more sensitive to informational ambiguity, trade disproportionately less with more distant partners with whom they are predictably less familiar (Huang 2007).

importance of those obstacles, indirect means, such as surveys to firms and inferences from econometric estimations, make it possible to arrive at some conclusions. Several survey-based empirical studies on the impact of alternative trade barriers in the USA, Europe and newly industrialised Asian countries indicate that lack of information is one of the most relevant export barriers, in terms of both frequency of occurrence and degree of severity.[7] In particular, most common export impediments are associated with identifying the initial contact and the marketing costs involved in doing business overseas, and also with difficulties in establishing initial dialogue with prospective customers or business partners and building relationships (Kneller and Pisu 2007).

Indications of the importance of information barriers can also be obtained from econometric studies (Rauch 1999; Rauch and Trindade 2002). The Chinese immigrant network is shown to have a larger trade-increasing effect for differentiated goods (*ie* products that are heterogeneous in both characteristics and quality) than for homogeneous goods.[8] If the trade-expanding effect of the network on the latter group of goods can be interpreted as the value of the network as informal contract enforcement, then the difference in observed impacts between these two goods classes may be taken to represent the value of market information, matching and referral services provided by the network, the implied information costs being approximately 6%.[9] Similarly, it has been shown that, after being intermediated in Hong Kong, mark-ups on re-exports of Chinese differentiated goods are 9–13% higher than those on homogeneous goods, which might be seen as the value of information-cost-reducing services provided by intermediating middlemen.[10]

Information impediments have motivated public interventions, which are commonly called *export promotion actions*. In fact, these interventions, and the formal entities responsible for them, *export promotion organisations*, are

---

[7]The studies use information directly obtained from the firms, primarily through mail surveys, and through personal and phone interviews (Leonidou 1995; see also Albaum 1983; Czinkota and Ricks 1983; Keng and Jiuan 1989; Katsikeas and Morgan 1994; Suárez-Ortega 2003; Leonidou 2004). In particular, limited information is often cited by exporters as a major barrier to both entering new export markets and expanding current export operations (Cavusgil and Naor 1987; Katsikeas 1994; Souchon and Diamantopoulos 1998).

[8]On the role of immigrant networks as information institutions that serve as nodes that help match buyers and sellers; see, for example, Gould (1994); Belderbos and Sleuwaegen (19980); Head and Ries (1998, 2001); Combes *et al* (2005); Herander and Saavedra (2005).

[9]By comparing the impact on trade of switching the Chinese network variable from zero to the sample mean for countries with strong Chinese immigrant links (*ie* both partners have more than 1% Chinese population), Anderson and van Wincoop (2004) calculate that the information-cost-reducing value of the network would be worth a 47% increase in trade. Assuming a value of elasticity of substitution of 8, the Chinese networks save an information cost worth 6%.

[10]Mark-ups are also higher for products with higher variance in export prices, for products sent to China to further processing, and for products shipped to countries that trade less with China (Feenstra and Hanson 2004).

almost ubiquitous (Rauch 1996). Economically, these actions might be—and have been—justified on the basis of market failures, primarily in the form of information externalities.[11] Given that it is difficult to exclude third parties from information and that its use is one of non-rivalry (*ie* use by one agent does not preclude its use by other agents), there is a potential for free riding on the successful searches of firms for foreign buyers.[12] These searches and the associated transactions then reveal information that may be used by other firms, which might eventually follow the pioneering firms without incurring the latter's costs.[13] As a result, the followers obtain important benefits from the first movers' initial investments and devalue the potential benefits from their searches (see, for example, Rauch 1996; Álvarez 2007). This is particularly true when companies attempt to enter a new export market or to trade a new product.[14] Private returns from these exporting activities would thus be lower than the corresponding social returns, and investment in their development would then be sub-optimally low (Westphal 1990).

However, the existence of a case for public intervention does not in itself mean that intervention is warranted. Its social costs need to be factored in. Besides the need to consider the implied opportunity costs, care must be taken not to underestimate obvious risks of resource diversion associated with potential rent-seeking activities as well as capture of the responsible agency by specific interest groups. Intervention would be advisable only if it would improve social welfare, *ie* if potential social benefits exceed corresponding social costs.[15]

As for benefits, activities undertaken by export promotion organisations can be viewed as a means of subsidising searches that counter the disincentives arising from potential free riding (Rauch 1996). These actions can help attenuate information problems. More precisely, trade assistance initiatives

---

[11]Information asymmetries on product quality may also create a case for trade policies (see, for example, Grossman 1989; Bagwell 1991). The same may also hold for externalities originated from managerial practices, training activities, technological change and production linkages (see, for example, Álvarez and López 2006; Edwards 1993; Feder 1983; Kessing 1967; Westphal 1990).

[12]Firms may learn about export opportunities from other firms through employee circulation, customs documents, customer lists and other referrals (Rauch 1996).

[13]Several studies present evidence on spillovers (Aitken *et al* 1997; Álvarez *et al* 2007; Greenaway *et al* 2004; Koenig *et al* 2010). Nevertheless, further research is required to determine the specific channels through which these spillovers take place (*eg* employment circulation) and thus whether they are actually spillovers and not potentially confounding effects, as well as the extent to which their magnitude can justify export promotion policies.

[14]See Hausmann and Rodrik (2003) and Álvarez *et al* (2007). In Hausmann and Rodrik's (2003) model, investment in developing new export activities is too low *ex ante* and entry is too high *ex post*.

[15]Determining the extent to which this is actually the case requires a study of its own.

can lower the fixed costs that firms incur when exporting for the first time and in entering specific new markets by reducing those costs associated with information gathering, *eg* carrying out overseas market studies on prices, product standards and potential buyers (Wagner 1995; Roberts and Tybout 1997). As a result, trade promotion organisations can potentially facilitate the internationalisation of companies and specifically their penetration into new country and/or product markets.

But is this actually the case? This simple answer to this question is that, so far, we do not know much. Information on how effectively programmes are managed by responsible entities seems necessary, since these programmes are costly and are just one of the possible alternative applications of generally scarce (and for the most part, public) resources. This chapter intends to present evidence in this regard.

Among other things, the effectiveness of export promotion will hinge upon the relevant macroeconomic and sectoral policies, the institutional attributes of the organisations (*eg* reporting schemes, norms that govern the selection and promotion of personnel, network of external offices, etc) and their incentives structures and the specific kinds of promotion activities performed and instruments applied. In this chapter we take the first two sets of factors—macroeconomic and sectoral policies and institutional attributes—as contextual conditioning elements to be controlled for, and take a detailed look at the impacts of programmes. Specifically, in this chapter we summarise the results of a series of assessments of the direct effects of these programmes on firms' export outcomes in several Latin American countries which were recently carried out by the Integration and Trade Sector of the Inter-American Development Bank (Volpe Martincus 2010). We first review the existing evidence and point out the main weaknesses of current measurement strategies; we then elaborate on how these can be overcome by applying econometric methods from the impact-evaluation literature on appropriate and comprehensive data sets. We close by discussing the potential limitation of this approach and advancing future lines of research in this area.

2   THE IMPACT OF TRADE PROMOTION ON FIRMS' EXPORT PERFORMANCE: WHAT DID WE KNOW?

There are two main sources of estimates of impacts of trade promotion actions on firms' export performance: organisations' own assessments and results from academic research. Most export promotion organisations generally rely on client satisfaction surveys and calculations based on firm-level customs data to assess the effects of their actions. Surveys primarily provide these organisations with qualitative indications on how they are doing. But the usefulness of this information is doubtful because evaluations based on non-objective data may be easily biased (see, for example, Klette *et al* 2000).

In addition, these studies are generally carried out by professional evaluators who work on commission and risk losing future clients if they provoke strong criticism.

In some cases, these surveys ask the managers of firms about the volume of incremental sales associated with the assistance received from the organisation. These quantitative measures of the effects of their activities also have several weaknesses (see, for example, Klette *et al* 2000). First, it could be presumed that the managers would exaggerate the size of the pay-off because that would increase chances that the programme would continue.[16] Second, individual case studies may have high marginal costs per case and may not be representative. Specifically, the response rate may be, and in fact is, markedly uneven and, on average, relatively low. Third, managers may not necessarily provide an accurate estimate of the pay-offs from a certain trade promotion activity because they must address counterfactual questions that are similar to those the econometricians must deal with, and may even have less information than the latter on the outcome of competing programmes and firms.

While lack of objective information is not an issue when comprehensive firm-level customs data are available, organisations with access to this data do not properly exploit them to overcome the limitations of survey-based evaluations. The most common practice can be called *direct imputation*. Here, entities directly take the sum of the values of exports or compute the change in this value for those firms that they have assisted, attributing the export outcomes of these firms—and the resulting expansion of national exports— as their contributions. These figures are likely to overestimate the impact of export promotion support, as it is implicitly assumed that these foreign sales or the increment of these sales would not have taken place in the absence of this support. This is evidently a questionable assumption.

On the other hand, until very recently, only a few studies in the empirical trade literature have used data at the firm level to evaluate more rigorously the impact of public policies on firm export behaviour. Two careful studies have been carried out focusing on activities undertaken by Chile's national export promotion organisation, PROCHILE (Álvarez and Crespi 2000; Álvarez 2004). These studies conclude that instruments managed by this organisation had a positive and direct effect on the number of destination markets to which firms export and, indirectly, after a period of four years, on product diversification. Furthermore, whereas trade shows and trade missions do not significantly affect the probability that firms become permanent exporters, exporter committees do have such effects. Results for the USA indicate that average states' expenditures on export promotion per firm do not significantly influence the probability that they will export (Bernard and Jensen 2004). In Ireland, grants

---

[16]On the other hand, in some countries it has been reported that sometimes exporters under-declare sales abroad, anticipating that this information might be used for tax purposes.

aimed at increasing investment in technology, training and physical capital, when large enough, appear to be effective in increasing exports of firms that are already exporting, but are not effective in encouraging new firms to enter international markets (Görg *et al* 2008). While insightful, all these studies concentrate just on the manufacturing sector or are based on small samples of firms.[17] Moreover, they do not fully identify the specific channels through which export promotion may affect exports.[18]

   Summing up, some organisations perform their own impact evaluations. In doing this, they rely primarily on information collected from firms participating in export promotion activities that they organise and/or automatically attribute to these activities the level and/or the change of the value of exports of firms receiving assistance. Both strategies have clear methodological flaws, which make their results highly questionable. Moreover, the existing academic literature virtually ignores developing countries and activities other than manufacturing, and here generally using only small samples of firms. Our understanding of the effects of trade promotion actions has been at least limited and, in particular, there has been a clear need to provide export promotion organisations with a set of analytical tools to help them evaluate these actions and better allocate funds to maximise their effectiveness. The subsequent sections introduce a first systematic attempt to fill these analytical and operational gaps.

---

[17]Specifically, Bernard and Jensen (2004) examine a sample of 13,550 US manufacturing plants over the period 1984–92, whereas Görg *et al* (2008) analyse a sample of 11,730 manufacturing firm–year observations in Ireland over the period 1983–2002 (*ie* an average of 587 firms per year). Álvarez and Crespi (2000) consider a sample of 365 Chilean firms out of a population of 7,479 exporting firms over the period 1992–6, while Álvarez (2004) investigates a sample of 295 Chilean manufacturing firms.

[18]Some papers examine the effects of regional and national expenditures on trade promotion on aggregate trade outcomes. States' export promotion spending has been reported to have positively affected total states' exports in the USA. In particular, it is estimated that an increase in manufacturing promotion expenditures of US$1 would generate an additional US$432 of manufacturing exports (Coughlin and Cartwright 1987). Recent evidence consistently shows that the size of the budget of export promotion organisations is positively related to countries' total exports in a cross-section of countries. For the median organisation, for each US$1 spent on trade promotion, exports would increase by US$40 (Lederman *et al* 2006). It has been also shown that official export credit guarantees are positively associated with the volume of exports in Germany (Moser *et al* 2008). Several other papers, mainly from the business economics literature, also investigate the impact of trade promotion on export performance. However, most of these contributions, besides their exclusive focus on developed countries, use highly specific, geographically and/or sectorally limited samples, or just look at a single or a few specific programmes. Hence, it would be virtually impossible to generalise from these analyses. In addition, endogenous selection of firms into trade assistance or its specific programmes is almost never taken into account. As a result, impact estimates are likely to be severely biased.

3   PRODUCING MORE RELIABLE ESTIMATES

Assessing the impact of public programmes is essentially a counterfactual analysis in which causal inference about the effect of these programmes requires determining how participants would have performed if they had not participated. The fundamental problem of impact evaluation is that, while *ex ante* each of the potential levels of exports is latent and could be observed, *ex post* only exports corresponding to participation or non-participation are observed. The other outcome is counterfactual and unobservable by definition, as is the difference between a firm's exports if it uses the services provided by the export promotion organisation relative to what its exports would be in the absence of these services. As a consequence, the counterfactual outcome must somehow be recovered from the data available. Constructing a valid control group to get a proper counterfactual may turn out to be a challenging task. The most obvious candidates are those firms that have not been served by the export promotion organisations. Thus, if we are interested in the average assistance effect on firms' total exports, the mean exports of those firms that have not been assisted by the organisation could be used.

However, firms receiving assistance can hardly be considered random draws, *ie* there may be non-random differences between assisted and non-assisted firms that may lead to potentially different export outcomes. Failure to account for these differences would clearly produce biased estimated impacts. In particular, if assisted firms are systematically better than non-assisted firms along specific dimensions not controlled for in the analysis, the estimates would overstate the causal effect of export promotion assistance.

In the empirical analysis whose results are presented below, observable differences among companies are accounted for using rich information on firm characteristics to reweight the unconditional differences among export outcomes. Nevertheless, as discussed later, upward biases are a potential risk inherent to these kinds of evaluation approaches, which unfortunately cannot be fully ruled out.

Alternative non-experimental methods have been proposed in the literature to control for firms' differing characteristics and thereby to construct the correct sample counterpart for the missing information on the outcomes had the firms not received trade support services. Two of these methods are *difference-in-differences* and *matching difference-in-differences*. The difference-in-differences estimator uses repeated observations on individuals (firms in this case) to measure the difference between the before and after change in exports for assisted firms and the corresponding change for non-assisted firms (Smith 2000; Jaffe 2002). The latter change serves here as an estimate of the true counterfactual, *ie* the export results that the firms in the treatment group would have achieved if they had not received trade promotion support, which makes it possible to identify temporal variations in outcomes that are not due to having received assistance (Abadie 2005). Therefore,

by comparing the aforementioned changes, the difference-in-differences estimator permits controlling for observed and unobserved time-invariant firm characteristics as well as time-varying factors common to both assisted and control firms that might be correlated with participation in export promotion programmes and export outcomes (see, for example, Galiani *et al* 2008).

Matching consists of pairing each assisted firm with the more similar members of the non-assisted group on the basis of their observable characteristics, and then estimating the impact of the assistance by comparing exports of matched assisted and non-assisted firms. This method is based on the main identifying assumption that selection into assistance occurs only on these observable characteristics of firms.[19] Due to data limitations, the analyst may not observe several characteristics. Consequently, systematic differences between the outcomes of assisted and non-assisted firms may persist even after conditioning on observable factors and the assumption that there is no selection on unobservables can be restrictive.

However, under certain conditions, selection on an unobservable determinant can be allowed for if matching is combined with difference-in-differences.[20] This is the *matching difference-in-differences* estimator (Heckman *et al* 1997, 1998; Abadie 2005; Smith and Todd 2005a). This estimator compares the before and after change in exports of assisted firms with that of matched non-assisted firms, so that imbalances in the distribution of covariates between both groups are accounted for and time-invariant effects are eliminated. Both procedures rely for identification on the assumption that there are no time-varying unobserved effects influencing selection into pro-

---

[19]See, for example, Heckman and Robb (1985) and Heckman *et al* (1998). Formally, matching is based on two assumptions. First, conditional on a set of observables, the non-treated exports are independent of the participation status (conditional independence assumption). The rationale is that firms that are similar in terms of the characteristics determining their selection into a trade assistance programme and potential export outcomes should have similar exports when participating, so that the differences in exports between participating and non-participating firms could be used as an estimate of the average effect of assistance if enough pairs of similar firms exist (Rubin 1974; Frölich 2004). Second, all firms have a counterpart in the non-assisted population, and any firm is a possible participant (common support). Together, both assumptions are called *strong ignorability*. Under these conditions, experimental and non-experimental analyses identify the same parameter. For additional details see, for example, Rosenbaum and Rubin (1983), Heckman *et al* (1997, 1998,1999), Angrist and Krueger (1999), Blundell and Costa Dias (2002) and Caliendo and Kopeinig (2008). Firms differ across multiple dimensions. Thus, matching firms may imply a potentially important dimensionality problem. In order to reduce this problem, matching is in general performed on the propensity to participate given the set of observable characteristics, or propensity score (Rosenbaum and Rubin 1983). Non-participants are then paired with participants that are similar in terms of this score according to a specific metric. See the appendix for further details.

[20]In particular, selection on an unobservable determinant is possible as long as this determinant lies on separable individual and/or time-specific components of the error term (Blundell and Costa Dias 2002).

motion activities and exports (Heckman *et al* 1997; Blundell and Costa Dias 2002).

The methods described above, as well as some variants of these methods, have been applied on firm-level export data from six Latin American countries: Peru, Costa Rica, Uruguay, Chile, Argentina and Colombia. For each country, the data set consists of two main databases. The first database has highly dis-aggregated export data at the firm level for four to eight years (depending on the country) over the period 2000–7 from the national customs agencies. Data is reported annually at the firm–product–country level to reveal how much of a certain product was exported by a given firm to a certain destination market in a particular year. Each record includes a firm's identifier, the product code (8-to-10-digit Harmonized System (HS)),[21] the country of destination and the export value in US dollars.[22] In all cases the sum of these firms' exports virtually adds up to the total merchandise exports as reported by the national central banks or the countries' national statistical offices. Hence, these data sets cover the whole population of exporters and not merely a sample of manufacturing firms. This is especially important for most Latin American countries as non-manufacturing activities still account for relatively large shares of total exports.

The second database consists of lists of firms assisted in each year of the respective sample periods, which were kindly provided by the export promotion organisations in the concerned countries. One organisation, PROEXPORT, has additionally furnished a list of companies using each of its main services. Finally, for some countries, additional data has been gathered on exporters, such as employment and location (Peru, Costa Rica and Argentina), starting data (Peru) and sales (Chile). These data, which are from the national tax or social security agencies (*eg* Peru's National Tax Administration Agency (SUNAT), Costa Rica's Social Security Administration (CCSS), Argentina's Federal Administration of Public Revenues (AFIP) and Chile's Internal Revenue Service (SSI)), allowed us to control for the influence of these variables.[23]

---

[21] The MCO Harmonized Commodity Description and Coding System.

[22] Unfortunately, we do not have the data needed to estimate and explicitly control for firms' total factor productivity. Nevertheless, if adding a new destination country or product requires incurring specific sunk costs of entry, then trading with a larger number of countries or a larger number of products will reflect higher productivity (Bernard *et al* 2006). Those export outcome indicators (lagged) are included in the propensity score underlying the estimates presented here. Hence, the role of productivity differences across (groups of) firms, and the possibility that the agency picks 'winners', may be at least partially accounted for.

[23] Thus, data on employment only cover formal employment. There is, of course, some risk of misreporting, which would generate measurement errors. As long as these errors are systematic across firms, they will be eliminated by the time differentiation implemented in the estimation methods used to carry out the evaluations.

These data have been used to assess the impact of trade promotion support on several firm-level export performance indicators, such as total exports, number of destination countries, number of products exported, average exports per country, average exports per product and average exports per country and product.[24] Therefore, the main aim of the assessment has been to estimate the *direct effect* of this support, and not to evaluate it from a social welfare point of view.[25] Moreover, since the required data were not available, the way export assistance affects other dimensions of firms' performance—such as productivity, total sales, or profits—could not be examined. For the same reason, it was not possible to analyse the impact of trade promotion activities on the overall firm extensive margin (*ie* the number of exporters).[26] In addition, indirect, sometimes non-pecuniary, effects from participating in export promotion activities, such as fairs and missions (*eg* testing the market for product acceptance, intelligence on competitors, morale of staff, etc), cannot be gauged easily and were not explicitly measured (see, for example, Bonoma 1983; Spence 2003; Seringhaus and Rosson 2005).

Finally, it should be stressed that the quantitative outcomes of the assessments are not always directly (perfectly) comparable across countries due to differences in sample periods, coverage of trade support data and sets of control variables (see the appendix for a description of the specific data set used in each country) and even in specific estimation methods.[27] Different estimation methods needed to be used in some cases to explore specific impacts of

---

[24]The primary focus is on contemporaneous effects. Notice, however, that there can also be lagged effects. For instance, business contacts obtained through participation in export promotion activities such as missions and fairs may take some time to materialise into concrete sales. Some evidence indicates that these effects are present. In the same vein, cumulative effects associated with self-learning and reputation building over time may also be present.

[25]To do so, we would need to contrast the social costs implied by trade promotion policies with the social benefit they may generate. This was beyond the scope of the study.

[26]Among other things, this analysis would require firm-level data on variables such as total sales and/or employment for both exporters and non-exporters and a list of non-exporting firms assisted by the export promotion organisations.

[27]The unavailability of some data prevented us from carrying out the same analysis in all cases. For example, comparisons of different programmes were only possible in the case of PROEXPORT, since data were not available for other countries' organisations. Similarly, examination of how effects of trade promotion vary with firm size as measured by employment could not be performed for Uruguay, Chile or Colombia. In addition, control variables vary from case to case: for instance, data on employment and age could be gathered for the entire population of Peruvian exporters, but similar data could not be obtained for Colombian or Uruguayan exporters. Admittedly, this might potentially create heterogeneous risks of overestimation of the true causal effects among countries. The size of the group of beneficiaries of export promotion programmes might also affect estimates, particularly in the cases of Chile and Colombia. These organisations assist a large proportion of exporters. As a consequence, trade support-related spillovers would be more likely and, *ceteris paribus*, so might therefore be an understatement of the effects of interest.

Table 2.1: *Empirical approach used in each case study.*

Export promotion organisation	Programme	Export outcomes*	Estimation method
PROMPERU	Single programme (binary participation status: 1 if a firm receives assistance from the organisation through one or more programmes described in Chapter 2, and 0 otherwise).	Total exports, number of destination countries, number of products exported, average exports per country, average exports per product, average exports per country and product, exports per country, number of products per country, exports per product, number of countries per product, exports per country and product.	Difference-in-differences, matching difference-in-differences and GMM system.
PROCOMER	Single programme (binary participation status: 1 if a firm receives assistance from the organisation through one or more programmes described in Chapter 2, and 0 otherwise).	Total exports, number of destination countries, number of products exported, average exports per country, average exports per product, average exports per country and product (both overall and by groups of firms exporting differentiated goods, reference-priced goods, homogeneous goods, differentiated and reference-priced goods, differentiated and homogeneous goods, reference-priced and homogeneous goods and differentiated, reference-priced and homogeneous goods).	Matching difference-in-differences.

trade promotion (*eg* average effects versus distributional effects; continuous export outcomes versus discrete export outcomes, etc). Furthermore, from an economic policy point of view, organisations operate in heterogeneous contexts and have different levels of resources and structures, including foreign offices (Volpe Martincus 2010). Hence, differences in estimated effects among

---

Finally, the coverage of export assistance data, in terms of firms included and programmes in which they participated, would also predictably influence estimated impacts.

Table 2.1: *Continued.*

Export promotion organisation	Programme	Export outcomes*	Estimation method
URUGUAY XXI	Single programme (binary participation status: 1 if a firm receives assistance from the organisation through one or more programmes described in Chapter 2, and 0 otherwise).	Total exports, number of destination countries, number of products exported, addition of a new export product, addition of a new differentiated export product, addition of new destination country, and addition of a new OECD destination country.	Matching difference-in-differences and endogenous switching binary response model.
PROCHILE	Single programme (binary participation status: 1 if a firm receives assistance from the organisation through one or more programmes described in Chapter 2, and 0 otherwise).	Total exports, number of destination countries, number of products exported, average exports per country, average exports per product, average exports per country and product, both overall and by deciles of these variables.	Semiparametric method for estimating quantile treatment effects (combined with difference-in-differences).
EXPORTAR	Single programme (binary participation status: 1 if a firm receives assistance from the organisation through one or more programmes described in Chapter 2, and 0 otherwise).	Total exports, number of destination countries, number of products exported, average exports per country, average exports per product, average exports per country and product, both overall and by firm size categories (small, medium, and large) as defined in terms of their number of employees.	Difference-in-differences, matching difference-in-differences, and double-robust estimation.

organisations should be interpreted with extreme caution, because these differences might be due to various factors, and it is not possible to clearly establish the extent to which these various factors are driving them.

4   THE IMPACT OF TRADE PROMOTION ON FIRMS' EXPORT PERFORMANCE: NEW EVIDENCE

This section presents the results of the evaluations of the effects of trade support programmes on firms' export performance in the six Latin American

Table 2.1: *Continued.*

Export promotion organisation	Programme	Export outcomes*	Estimation method
PROEXPORT	Multiple programmes (participation status specific to counselling services, trade agenda services; trade fair, shows, and mission services; or their alternative combinations).	Total exports, number of destination countries, number of products exported, average exports per country, average exports per product, average exports per country and product, both overall and comparing programmes to non-participation and to each other.	Multiple programme matching difference-in-differences.

*Since estimation methods work on differences over time to control for observed and unobserved time-invariant firm characteristics as well as time-varying factors common to both assisted and control firms that might be correlated with participation in trade promotion programmes and export outcomes, estimation effects are in fact measured on the growth of these variables, except for those discrete export outcomes examined in the case of Uruguay.

countries listed above. These effects can generally be expected to be heterogeneous along several dimensions. Specifically, their size is likely to be related to the severity of the information problems involved in the specific trading operations and/or faced by the firms carrying out these operations. This will be illustrated next using the experience of the following export promotion organisations:[28]

- PROMPEX/PROMPERU (Peru);
- PROCOMER (Costa Rica);
- URUGUAY XXI (Uruguay);
- PROCHILE (Chile);
- EXPORTAR (Argentina);
- PROEXPORT (Colombia).

4.1   *Peru: Extensive Margin or Intensive Margin?*[29]

Informational obstacles tend to be more important when firms attempt to increase their number of destination countries or the set of products they sell abroad (extensive margin or diversification) than when they seek to expand exports of goods they have already been trading and/or to countries that are already among their destination markets (intensive margin or deepening).

---

[28]Detailed information on these organisations can be found in Jordana *et al* (2010) and Volpe Martincus (2010). Tables 2.1 and 2.2 describe each of the programmes in detail.

[29]This subsection is based on Volpe Martincus and Carballo (2008).

Table 2.2: *Datasets.*

Country	Export data	Trade support data	Sample period
Peru	Firm exports disaggregated by product (10-digit HS) and destination country	All exporters/all programmes	2001–5
Costa Rica	Firm exports disaggregated by product (10-digit HS) and destination country	All exporters/all programmes	2001–6
Uruguay	Firm exports disaggregated by product (10-digit HS) and destination country	Exporters interacting closely with the organisation (*ie* face-to-face contacts)/primarily missions and fairs	2000–7
Chile	Firm exports disaggregated by product (8-digit HS) and destination country	All exporters/all programmes	2002–6
Argentina	Firm exports disaggregated by product (10-digit HS) and destination country	Exporters interacting closely with the organisation (*ie* face-to-face contacts)/primarily missions and fairs	2002–6
Colombia	Firm exports disaggregated by product (10-digit HS) and destination country	All exporters/all programmes	2003–6

Trade promotion programmes can have varying effects across firms' export activities. In particular, the effect of these programmes will predictably be larger on the extensive margin than on the intensive margin of exports. We focused on the case of PROMPEX (currently PROMPERU) to shed light on this issue.

The overall estimates suggest that participation in activities carried out by PROMPEX/PROMPERU has been associated with an increased rate of growth of firms' total exports, number of destination countries and number of products exported (see Figure 2.1). The rate of growth of exports was 17% higher for firms assisted by PROMPEX, while those of the number of countries and
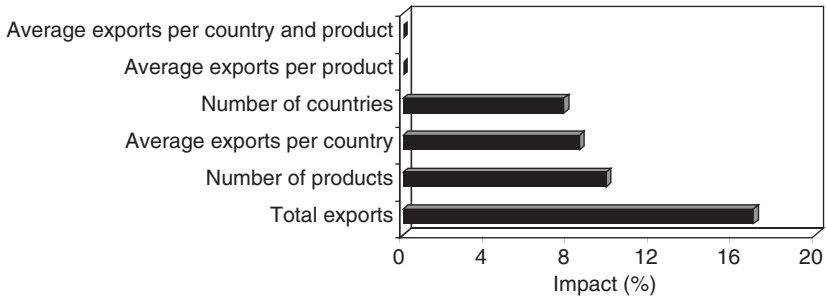
Figure 2.1: *Peru: Average export assistance effect on assisted firms.*
Source: authors' calculations based on data from PROMPEX (currently PROMPERU) and SUNAT. Note: the sample period is 2001–5. Statistically insignificant effects are reported as zero.

the number of products were 7.8% and 9.9% higher, respectively. Given a sample average annual growth rate of the number of products of 36.5%, the latter result implies that supported companies would have had a growth rate 3.6 percentage points higher than comparable non-supported companies.

In contrast, the impact on variables capturing export outcomes along the intensive margin was weaker and evidently less robust.[30] Hence, in line with prior expectations, export promotion assistance by PROMPEX/PROMPERU has helped Peruvian firms expand their exports, primarily through the expansion of the number of destination countries and the number of products exported.

4.2   Costa Rica: What Kind of Exports Does Export Promotion Promote?[31]

The degree of incompleteness of information can vary according to the nature of the goods traded. In particular, differentiated goods are heterogeneous in terms of both their characteristics and their quality. This interferes with the signalling function of prices, thus making it difficult to trade these goods in organised exchanges (Rauch 1999). Thus, information problems faced when trading differentiated products are more severe than those arising when trading more homogeneous goods.[32] Export promotion assistance may then have

---

[30]Trade promotion only seems to stimulate greater exports per country. This might be explained by the fact that an organisation can help firms obtain business contacts in new regions within countries that are already among their destination markets. However, this latter result was not as robust as the previous ones and does not survive all control exercises.

[31]This subsection is based on Volpe Martincus and Carballo (2011).

[32]Information problems can also become an important entry barrier in export markets in the case of experience goods, *ie* products whose quality is learned from consumption after purchase (Nelson 1970).
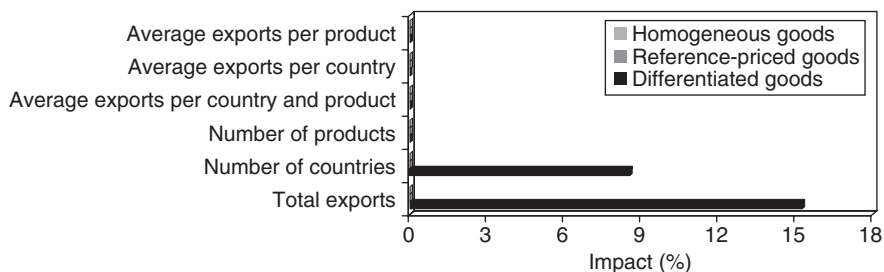
Figure 2.2: *Costa Rica: average export assistance effect on assisted firms by type of products.*
Source: authors' calculations on data from PROCOMER and CCSS. Note: statistically insignificant effects are reported as zero.

different effects on export performance, depending on the degree of differentiation of the products exported by the companies. Trade promotion actions can be expected to have a stronger impact on the extensive margin of firms exporting differentiated goods, *ie* on the introduction of additional differentiated products and/or the incorporation of more countries to the set of destinations to which these products are exported. We explored whether this was the case based on the experience of PROCOMER in Costa Rica.

The results from the impact-evaluation exercises indicate that firms already exporting only differentiated goods that participated in promotion activities organised by PROCOMER had higher rates of growth of exports and number of destination countries than comparable firms that were not assisted (see Figure 2.2). The rate of growth of exports was on average 15.3% higher for firms assisted by PROCOMER, while that of the number of countries was 8.5% higher.[33]

In contrast, assistance by PROCOMER does not seem to have resulted in higher export growth either on the intensive or on the extensive margin for firms that only exported reference-priced or homogeneous products.[34] Thus, evidence from Costa Rica confirms that the effects of trade promotion actions have been greater for export operations along the country extensive margin involving differentiated goods than for operations involving more homogeneous goods.

[33]Nonetheless, trade promotion actions do not seem to have affected the probability of start exporting differentiated goods.

[34]Effects on outcomes of firms that export alternative combinations of the different types of goods are not significant.
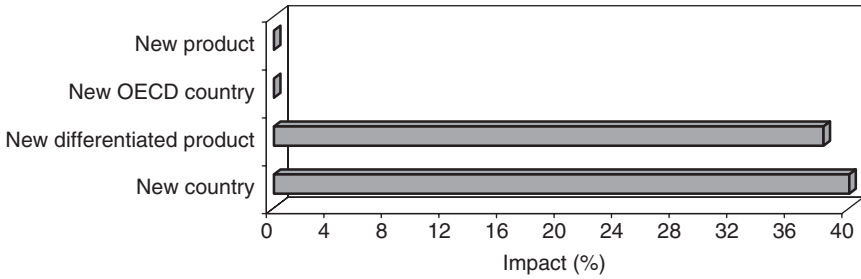
Figure 2.3: *Uruguay: export assistance effect on the probability of entering new country and product markets.*

Source: authors' calculations based on data from URUGUAY XXI. Note: the sample period is 2000–7. Statistically insignificant effects are reported as zero.

4.3 Uruguay: Does Export Promotion Help Enter New Markets? [35]

Previous estimates show how export promotion affected the number of destination countries and the number of products exported in general and for specific groups with different degrees of differentiation. However, the direct impact of trade promotion programmes on the probability that a firm will add an entirely new destination country or introduce a completely new export product into its export business activities has so far not been strictly evaluated.[36] We therefore evaluate the impact of export promotion on the entry into new markets by considering the case of URUGUAY XXI.

The estimates reveal that trade support has had a positive and significant impact on the probability of adding a new country—40% higher for firms supported by URUGUAY XXI (see Figure 2.3). Still, export support does not seem to have enabled firms to enter a new OECD country. In fact, positive significant impacts are only observed in the case of Latin American and Caribbean countries.[37]

---

[35] This subsection is based on Volpe Martincus and Carballo (2010a).

[36] Note that this is not necessarily the same as an overall increase in the number of markets in which firms operate, since such an increase might just as well result from simultaneously adding several markets and dropping others, potentially including some that could have been served in the past.

[37] A firm based in a developing country must undergo product and marketing upgrades to succeed in exporting to developed countries. Properly shaping the marketing strategy is an information-intensive activity. For instance, firms need to learn and understand the preferences of foreign consumers, the nature of competition in foreign markets, the structure of distribution networks and the requirements, incentives and constraints of distributors. These activities are intrinsically more difficult when exporting to more sophisticated markets (Artopoulos *et al* 2010). As a consequence, effects of export promotion might not be uniform across destination countries with different levels of development.
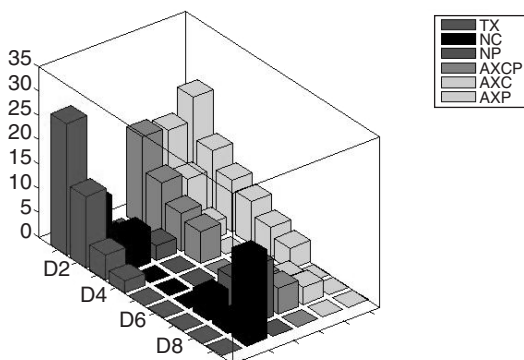
Figure 2.4: *Chile: assistance effect on assisted firms by export outcome deciles.*
Source: authors' calculations based on data provided by PROCHILE. Note: TX, total exports; NC, number of countries; NP, number of products; AXCP, average exports per country and product; AXC, average exports per country; AXP, average exports per product. Deciles are defined in terms of growth rates of these variables. The sample period is 2002–6. Statistically insignificant effects are reported as zero.

However, export promotion assistance does not appear to have had any impact on the probability that a firm would add new products in general. Positive effects have been confined to differentiated goods, where, as discussed above, information barriers to trade are higher. In this case, the assistance effect on assisted firms was 38.2 percentage points, *ie* the probability of introducing these goods was 38.2% higher for firms participating in trade promotion programmes. Export supporting activities by URUGUAY XXI seem to have been effective in helping Uruguayan firms penetrate new destination countries, especially Latin American and Caribbean markets, and introduce new differentiated products.

4.4   Chile and Argentina: What Are the Distributional Effects of Export Promotion?[38]

The relative importance of information-related obstacles faced when operating in foreign markets is different for firms of different degrees of export involvement and different sizes (see, for example, Diamantopoulos *et al* 1993; Naidu and Rao 1993; Czinkota 1996; Moini 1998). It has been shown that the frequency of firms indicating the aforementioned barriers as difficulties to exporting declines with the experience of firms in international markets, which suggests that there is a process whereby firms learn how to deal with export barriers through direct experience in these markets (Kneller and Pisu 2007). Similarly, the literature agrees that smaller firms face greater limitations than larger firms in trading across borders (see, for example, Roberts and Tybout 1997; Bernard and Jensen 1999, 2004; Wagner 2001, 2007; Álvarez 2004).
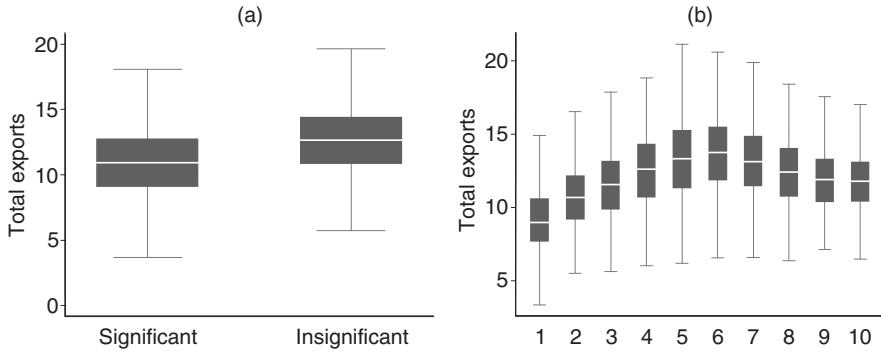
Figure 2.5: *Chile: distribution of exports over significance groups and deciles defined in terms of export growth.*
Source: Our calculations based on data provided by PROCHILE. Note: the sample period is 2002–6.

These differences across firm sizes may be related to heterogeneity in the scale and resources to perform the gathering activities to access to relevant information (*eg* through market studies), and more generally in the ability to cope with the sunk cost associated with the penetration of new foreign markets, such as those originated in setting up an export department (or redesigning products for foreign customers). Public programmes aimed at addressing information problems can therefore be expected to change with the stages of firms' internalisation process and their size categories. In particular, these programmes are likely to have greater impacts on export outcomes of relatively inexperienced and smaller companies.

We next discuss the evidence on these effects based on the experiences of PROCHILE in Chile and EXPORTAR in Argentina.

According to the results of our analysis, trade promotion assistance by PROCHILE has had a significant impact on the lower tail of the distribution of export growth rates, namely, in the first to fourth deciles (see Figure 2.4). The impact was the strongest in the lowest decile, and it monotonically decreased from the second decile to the fourth decile. Moreover, significant effects were observed in both tails of the distribution (first to third and seventh to ninth deciles) of the growth rate of the number of countries. Furthermore, while the average assistance effect on the number of products was virtually zero, significant positive impacts were identified in specific parts of the relevant distribution. As with the case of the number of countries, these impacts were concentrated at the lower and upper ends of the distribution (second to third and seventh to eighth deciles).

In order to identify which kinds of firms were benefiting from these programmes, we need to look back at the previous export levels of the different
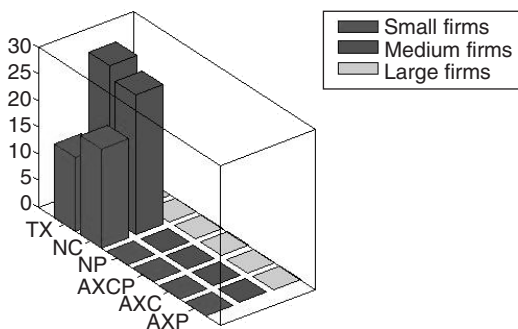
Figure 2.6: *Argentina: average assistance effect on assisted firms by size category.*
Source: Our calculations based on data provided by UMCE-SICP, EXPORTAR and AFIP.
Note: TX, total exports; NC, number of countries; NP, number of products; AXCP, average exports per country and product; AXC, average exports per country; AXP, average exports per product. Effects correspond to first-time assistance. Statistically insignificant effects are reported as zero. Small firms: 1–50 employees; medium-sized firms: 51–200 employees; large firms: more than 200 employees. Sample period is 2002–6.

groups of firms.[39] The distribution of exports for the set of firms with significant impacts is below that of the set of firms with no significant impacts (see Figure 2.5(a)).

More interestingly, the distribution of exports for the (two) group(s) of firms where the strongest effects were detected is located below those for the groups of firms where weaker or no significant effects were registered (see Figure 2.5(b)). This indicates that smaller exporters benefited proportionally more from trade promotion activities than did larger exporters.

In Argentina, the estimates consistently suggest that the positive effects of export promotion programmes administered by EXPORTAR on total exports and number of destination countries was stronger for small and medium-sized firms, as defined in terms of their number of employees. Based on the estimated impacts of the first assistance, for small firms that had participated in these programmes, growth rates of exports and number of destination markets have been 13.9% and 18.5% higher, while for medium-sized participating firms they have been 28.7% and 26.4% higher, respectively, in both cases relative to comparable non-participating counterparts (see Figure 2.6).[40]

---

[39]We estimated the distributions of firms' total (lagged) exports both aggregating over deciles of the distribution of their growth rates where significant and non-significant effects of trade promotion have been found, and for each decile of the distribution of first-differentiated total exports. These distributions are shown as box plots in Figure 2.5.

[40]In principle, firms can be assisted several times over the years. The results shown in the figure correspond to the effects of the services received by the firms the first year they were supported.
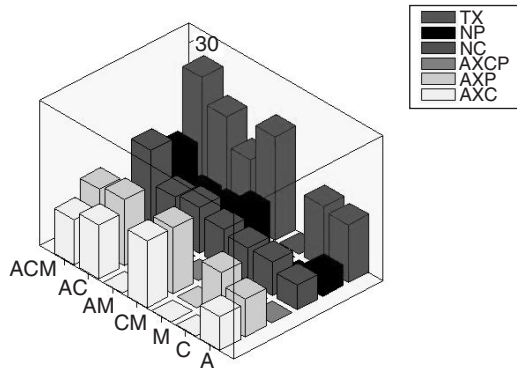
Figure 2.7: *Colombia: average effect of export assistance programmes on assisted firms relative to non-participation.*

Source: authors' calculations based on data from PROEXPORT. Note: the figure reports the effect of each export promotion programme relative to non-participation. C, counselling services; A, trade agenda services; M, trade fair, shows and mission services; TX, total exports; NP, number of products; NC, number of countries; AXCP, average exports per country and product; AXC, average exports per country; AXP, average exports per product. Sample period is 2003–6. Statistically insignificant effects are reported as zero.

For large firms, no significant impacts on exports were observed. Due to their inherent characteristics and their antecedents in international markets, these firms could improve their performance without the support of the organisation.

As expected, the benefits from trade promotion programmes managed by PROCHILE and EXPORTAR seem to have been greater for relatively inexperienced and smaller firms as measured by their past total exports and their number of employees, which are precisely those companies that face the greatest challenges in overcoming informational barriers.

4.5   Colombia: Which Export Promotion Programmes Are Most Effective?[41]

Export promotion policies consist of a variety of programmes. Although all programmes share the common aim of improving the export performance of firms, they may differ significantly in terms of effectiveness. This can be explained by differences in the degree of correspondence between companies' specific needs and the specific support provided by the organisation and the relative intensity of synergic effects linked to the combination of services. Gauging the relative effectiveness of these programmes is extremely important for assessing whether trade promotion activities are effectively

---

[41] This subsection is based on Volpe Martincus and Carballo (2010c).
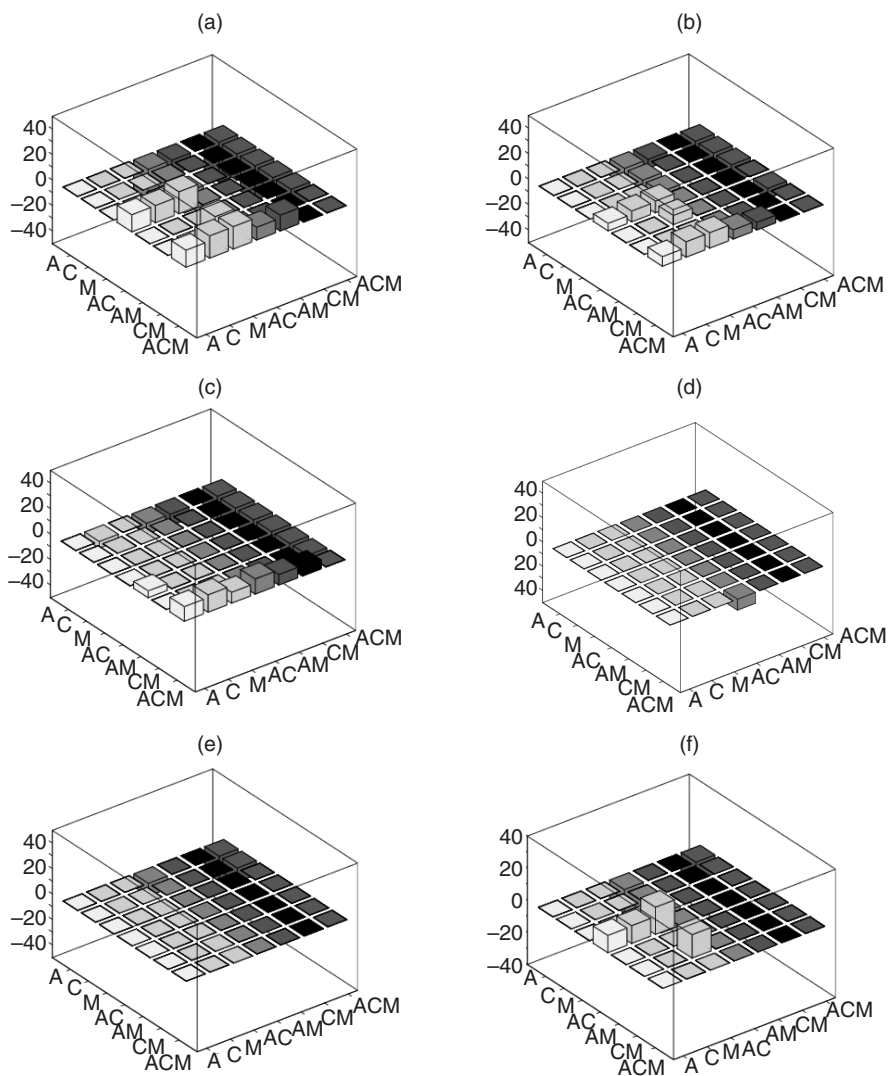
Figure 2.8: *Colombia: average effect of export assistance programmes on assisted firms relative to each other. (a) Total exports; (b) number of products; (c) number of countries. (d) Average exports per product and country; (e) average exports per product; (f) average exports per country.*

Source: authors' calculations based on data from PROEXPORT. Note: the figure reports the effect of each export promotion programme relative to each other. C, counselling services; A, trade agenda services; M, trade fair, shows and mission services. The sample period is 2003–6. Statistically insignificant effects are reported as zero.

targeted—in the sense that firms that use a certain service perform better than if they had used another service—or whether some services are consistently better than others. This information can be valuable in guiding the allocation of public funds devoted to trade promotion in order to maximise their impact and thereby improve existing policies. The case study on Colombia examined these programme-specific effects.

The results from our assessment indicate that a combination of the three basic services—counselling, missions and fairs, and trade agenda—systematically performed better. These bundled services seem to have been associated with better export outcomes, primarily in terms of total exports and number of destination countries, relative to both the non-assistance situation and each of the services individually considered (see Figures 2.7 and 2.8). Firms combining these three services have had significantly higher export growth along the country and product extensive margins than if they had used each of these services separately. For these firms, the growth rate of exports was on average 17.7% higher, the number of countries was 11.7% higher and the number of products was 11% higher. These firms also exhibit a higher growth of the number of destination countries (on average, 9.4% higher), when compared with a situation in which they had used alternative combinations of two of these three services.

In contrast, trade promotion programmes, either individually or bundled, do not appear to have had significant impacts on average exports per country, average exports per product and average exports per country and product. This is consistent with findings in other countries in the region.

To sum up, an examination of the different promotion programmes carried out by Colombia's PROEXPORT reveals that bundled services combining counselling, trade agenda and trade missions and fairs—which provide exporters with a comprehensive support throughout the process of starting export businesses and building up buyer–seller relationships with foreign partners—are more effective than isolated assistance actions such as trade missions and fairs alone.

Recapitulating

We have applied econometric methods already employed in other public policy fields on highly disaggregated firm-level export and trade support data in order to analyse the impact of trade promotion programmes managed by relevant entities in several Latin American countries on various dimensions of firms' export performance. From this in-depth analysis at least four general conclusions can be made.

First, trade assistance effects are predictably greater on the extensive margin of firms' exports, *ie* when firms attempt to increase the number of destination countries and/or to expand the set of goods exported and,

specifically, when they seek to enter an entirely new country or product market.[42]

Second, export promotion actions are more likely to generate larger export gains to the extent that products traded are more differentiated, because information barriers are more important in these cases.

Third, due to the greater limitations they face in accessing relevant export information, small firms with limited previous involvement in international markets can be expected to benefit more from export assistance.

Fourth, bundled support services provided throughout the export process, from the beginning of the commercial contacts to the establishment of the business relationships, seem to be more effective in enhancing firms' export perspectives than individual actions.

Do the previous findings mean that larger companies expanding their exports of reference-priced goods in their current destination markets should not be supported? Not necessarily. For instance, these firms might generate positive external reputational effects that benefit trading initiatives of other firms. Remember that these and other indirect effects have not been explicitly considered in these evaluations, although ideally they should be taken into account for computing cost-effectiveness ratios.

These facts should therefore be interpreted as general criteria that, along with others to be developed through further research and after factoring in implied relative costs, could be used in designing trade support programmes to maximise their impact.

5 WHAT SHOULD BE CONSIDERED WHEN INTERPRETING THE PREVIOUS FINDINGS?

As mentioned above, current impact-evaluation practices of many trade promotion organisations have methodological flaws. These evaluations are likely to misrepresent the real contribution these organisations make to the companies' export growth and thereby to that of the countries. In fact, previous estimates indicate that, over the sample period, the latter strategy would on average overestimate these contributions by 5.9 times for PROMPEX (currently PROMPERU); 9.4 times for PROCOMER; 7.3 times for URUGUAY XXI; 14.3 times for PROCHILE; 5.2 times for EXPORTAR; and 3.7 times for PROEXPORT. Hence, the outputs of these evaluations are not adequate for guiding the strategies and activities of these organisations as well as the allocation of their scarce

---

[42]A simple portfolio argument suggests that if covariance of firm sales across countries is not perfect, then spreading these sales over a larger number of countries will be associated with more stable total sales. This can be expected to result in less likelihood of business failure and of abandoning international markets. The final outcome may be increased firm survival (Volpe Martincus and Carballo 2009).

resources to these activities to maximise their influence on their countries'
export development.

While the methods used to generate previous results allow for substantial
improvements in the accuracy of impact estimates, they are not exempted
from limitations. Caution is hence required when interpreting these estimates.
First, identification is based on the assumption that time-varying unobserved
effects do not affect selection into support and exports. Admittedly, there
might be time-variant firm-specific factors that lead to improved export per-
formance and are not observable to us. This is, for instance, predictably the
case with productivity. While we generally include some firm-level variables
that at least partially account for productivity differences across groups of
firms, these remain imperfect controls. If, after conditioning on these vari-
ables, productivity still plays a significant role in determining programme par-
ticipation, and is accordingly higher for firms that receive export assistance
from the organisations, then these procedures would overestimate the true
causal effects of this assistance on export performance. Unfortunately, this
possibility cannot be excluded. But, given the process of selection into trade
support, the extension of our sample period and the covariates included, we
would expect that, if any, this is not a serious problem.

Second, these evaluation exercises take the main assumptions of the Roy–
Rubin model for granted (Roy 1951; Rubin 1974). Cross-market and gen-
eral equilibrium effects are assumed away.[43] However, these assumptions
are likely to be violated in many contexts. This might happen, for instance,
when estimating the effects of foreign acquisitions on wages (see, for exam-
ple, Girma and Görg 2007). Evaluation of export promotion policies is, of
course, not an exception. As we have said, there may be information exter-
nalities associated with exporting activities.[44] If these spillovers were linked
to participation in specific export promotion actions, then the outcome dif-
ferences between assisted and non-assisted firms corrected by observable
heterogeneity across these groups would underestimate the true impact of
these actions (see, for example, Heckman *et al* 1999; Miguel and Kremer 2004;
Ravallion 2008). Under perfect contemporary dissemination of information
across firms, this impact would not be statistically different from zero and
could accordingly not be identified. Yet, there might also be negative (pecu-
niary) externalities in the form of increased competition. Firms receiving trade
assistance (as well as their followers) may penetrate particular country and/or
product markets, thereby potentially eroding the position of other domestic
firms already serving these markets. According to informal tests performed in

---

[43] Hence, outcomes for each firm do not depend on the overall level of participation in
the activities performed by the agency (Heckman *et al* 1998).

[44] Álvarez *et al* (2007) report evidence in favour of the existence of such spillovers in the
case of Chile. They find that the probability of firms introducing given products to new
countries or different products to the same countries increases with the number of firms
exporting those products and to those destinations, respectively.

some of the case studies, neither self-discovery nor competition effects seem to seriously threaten the validity of the estimation results reported in this chapter. Nonetheless, these phenomena deserve to be explored more thoroughly in future research.

Third, in several policy areas, interventions are multiple. Support to companies is, of course, not an exception. In some countries, firms may get assistance from different public and private entities. Regrettably, in these cases there is no unified register of firms benefiting from various support measures. Thus, it is not possible to account for the influence of interventions other than trade promotion, with the result that these actions become an unobserved factor. If this factor is time invariant over the sample period, its impact will be automatically controlled for by the estimation procedures, which identify the effects of interest based on the time variation. If firms' participation status in other assistance programmes is instead a time-varying variable, we are back to the first scenario described above. Two extreme cases can be considered. If all firms assisted by export promotion organisations are also simultaneously receiving support through programmes managed by other public or private agencies, and if these programmes have significant effects on firms' export performance, then the estimated impacts will overestimate those of trade promotion activities and will instead reflect the effects of the combined assistance. In contrast, if no company participating in these activities is simultaneously a beneficiary of other support initiatives, then, as long as these are effective, estimates will understate their true incidence on firms' export outcomes. We will return to this issue in the next section.

6   THE WAY AHEAD

Periodic evaluations are necessary for well-informed policy decisions and, in particular, an indispensable component of the dynamic adjustment of the organisations to the evolving needs of their services users. The adequacy of the design of export promotion activities and the robustness of the evaluations of their effectiveness depend on the access to a set of information that allows for a precise knowledge of the firms. Unfortunately, gathering such information is extremely difficult, even for the export promotion organisations themselves, in some cases because from a legal point of view these are private entities.

These organisations need to gain better access to relevant data stored by other public agencies, including national bureaus of statistics, and improve collaboration to generate new data, all under conditions that ensure strict confidentiality. Here, strengthened cooperation between public and private entities would be desirable. The availability of richer and consistent databases would make it possible to go beyond the analysis of the effects of trade promotion actions on the primary variable of interest: exports. This would facilitate

the examination of their impact on other measures of firm performance such as productivity.[45]

Export promotion policies are just one subset of public policy instruments that may affect countries' profiles in international trade. Strictly conceived, they reduce information and trade costs, thus enabling existing firms entering international markets, as well as current exporters, to diversify their external sales of the goods they already produce across destination markets. Other specific public policies, some of which are becoming increasingly interconnected with export promotion in developed countries, would also predictably have impacts on countries' international trade. Thus, business development support throughout the process of establishing a new company may result in the emergence of new firms producing new goods, and therefore in product diversification and export diversification (for example, when properly combined with trade promotion).

Provided that the aforementioned databases include lists of beneficiaries of different public support programmes (*eg* export promotion and innovation promotion), it would be possible not only to determine how to improve coordination under the prevailing conditions, but also to carry out reliable evaluations that consider the existence of other assistance initiatives in which companies participate and assess complementarities and synergies among them. Insights into these potential interdependencies would be valuable for designing policy instruments and establishing their components and sequencing. These consistent data on the various relevant programmes would allow for evaluations of their relative merits in terms of a common metric and would help policymakers better allocate resources.

The results presented in this chapter are based on non-experimental methods. Social experiments can generate further and, under certain conditions, more robust insights into the effects of trade assistance, and therefore appear as the natural next alternative strategy to pursue.

Christian Volpe Martincus is a Senior Economist in the Integration and Trade Sector at the Inter-American Development Bank.

REFERENCES

Aakvik, A., T. Holmas, and E. Kjerstad (2003). A low-key social insurance reform-effects of multidisciplinary outpatient treatment for back pain patients in Norway. *Journal of Health Economics* **22**, 747–62.

---

[45]Although still it is disputed in the empirical literature, exporting and expanding exports along the country and product extensive margins may generate positive externalities (*ie* learning-by-exporting effects), thereby resulting in increased productivity. In this way, trade support may potentially end up boosting firms' productivity.

Aakvik, A., J. Heckman, and E. Vytlacil (2005). Estimating treatment effects for discrete outcomes when responses to treatment vary: an application to Norwegian vocational rehabilitation programs. *Journal of Econometrics* **125**, 15–51.

Abadie, A. (2005). Semiparametric difference-in-differences estimators. *Review of Economic Studies*, **72**(1), 1–19.

Abadie, A., and G. Imbens (2006). On the failure of the bootstrap for matching estimators. NBER Technical Working Paper 325.

Aitken, B., G. Hanson, and A. Harrison (1997). Spillovers, foreign investment, and export behavior. *Journal of International Economics* **43**, 103–132.

Albaum, G. (1983). Effectiveness of government export assistance for US smaller-sized manufacturers: some further evidence. *International Marketing Review* **1**(1), 68–75.

Álvarez, R. (2004). Sources of export success in small- and medium-sized enterprises: the impact of public programs. *International Business Review* **13**, 383–400.

Álvarez, R. (2007). Explaining export success: firm characteristics and spillover effects. *World Development* **35**, 377–393.

Álvarez, R., and G. Crespi (2000). Exporter performance and promotion instruments: Chilean empirical evidence. *Estudios de Economía* **27**, 225–241.

Álvarez, R., and R. López (2006). Is exporting a source of productivity spillovers? CAEPR Working Paper 2006-012.

Álvarez, R., H. Faruq, and R. López (2007). New products in export markets: Learning from experience and learning from others. Mimeo, Indiana University.

Anderson, J., and E. van Wincoop (2004). Trade costs. *Journal of Economic Literature* **42**, 691–751.

Angrist, J., and A. Krueger (1999). Empirical Strategies in Labor Economics, in O. Ashenfelter and D. Card (eds), *Handbook of Labor Economics*, Elsevier.

Artopoulos, A., D. Friel, and J. Hallak (2010). Challenges of exporting differentiated products to developed countries: the case of SME-dominated sectors in a semi-industrialized country. IDB Working Paper 166.

Athey, S., and G. Imbens (2006). Identification and inference in nonlinear difference-in-differences models. *Econometrica* **74**, 431–497.

Bagwell, K. (1991). Optimal export policy for a new-product monopoly. *American Economic Review* **81**(5), 1156–1169.

Barrios, S., H. Görg, and E. Strobl (2003). Exporting firms' export behaviour: R&D, spillovers, and destination market. *Oxford Bulletin of Economics and Statistics* **65**, 475–496.

Belderbos, R., and L. Sleuwaegen (1998). Tariff jumping DFI and export substitution: Japanese electronic firms in Europe. *International Journal of Industrial Organization* **16**, 601–638.

Bernard, A., and B. Jensen (1999). Exceptional exporter performance: cause, effect, or both? *Journal of International Economics* **47**, 1–25.

Bernard, A., and B. Jensen (2004). Why some firms export? *Review of Economics and Statistics* **86**, 561–569.

Bernard, A., S. Redding, and P. Schott (2006). Multi-product firms and product switching. NBER Working Paper 12293.

Bertrand, M., E. Duflo, and S. Mullainathan (2004). How much should we trust difference-in-differences estimates? *Quarterly Journal of Economics* **119**, 249–275.

Blundell, R., and S. Bond (1998). Initial conditions and moment restrictions in dynamic panel data models. *Journal of Econometrics* **87**, 115–143.

Blundell, R., and M. Costa Dias (2002). Alternative approaches to evaluation in empirical microeconomics. CEMMAP Working Paper CWP10/02.

Bonoma, T. (1983). Get more out of your trade shows. *Harvard Business Review*. http://hbr.org/1983/01/get-more-out-of-your-trade-shows/ar/1.

Branch, A. (1990). *Elements of Export Marketing and Management*. London, Chapman and Hall.

Caliendo, M., and S. Kopeinig (2008). Some practical guidance for the implementation of propensity score matching. *Journal of Economic Surveys* **22**, 31–72.

Caminal, R., and X. Vives (1996). Why market shares matter: an information-based theory. *Rand Journal of Economics* **27**, 221–239.

Cavusgil, S., and J. Naor (1987). Firm and management characteristics as discriminators of export marketing activity. *Journal of Business Research* **15**(3), 221–235.

Chen, N. (2004). Intra-national versus international trade in the European Union: why do national borders matter? *Journal of International Economics* **63**, 93–118.

Chen, S., R. Mu, and M. Ravallion (2009). Are there lasting impacts of aid to poor areas? *Journal of Public Economics* **93**, 512–528.

Chiang, S., and R. Masson (1988). Domestic industrial structure and export quality. *International Economic Review* **29**, 261–270.

Chisik, R. (2003). Export industry policy and reputational comparative advantage. *Journal of International Economics* **59**, 423–451.

Combes, P., M. Lafourcade, and T. Mayer (2005). The trade-creating effects of business and social networks: evidence from France. *Journal of International Economics* **66**, 423–451.

Coughlin, C., and P. Cartwright (1987). An examination of state foreign export promotion and manufacturing exports. *Journal of Regional Science* **27**, 439–449.

Czinkota, M. (1996). Why national export promotions. *International Trade Forum* **2**, 10–13.

Czinkota, M., and D. Ricks (1983). The use of multi-measurement approach in the determination of company export priorities. *Journal of the Academy of Marketing Science* **11**, 283–291.

Czinkota, M., and I. Ronkainen (2001). *International marketing* (Mason, OH: Thomson South-Western).

Diamantopoulos, A., B. Schlegelmich, and Y. Tse. (1993). Understanding the role of export marketing assistance: empirical evidence and research needs. *European Journal of Marketing* **27**, 5–18.

Djebbari, H., and J. Smith (2008). Heterogeneous impacts in PROGRESA. *Journal of Econometrics* **145**, 64–80.

Edwards, S. (1993). Openness, trade liberalization, and growth in developing countries. *Journal of Economic Literature* **31**, 1358–1393.

Egan, M., and A. Mody (1992). Buyer–seller links in export development. *World Development* **20**, 321–334.

Feder, G. (1983). On exports and economic growth. *Journal of Development Economics* **12**, 59–73.

Feenstra, R., and G. Hanson (2004). Intermediaries in entrepot trade: Hong Kong re-exports of Chinese goods. *Journal of Economics and Management Strategy* **13**, 3–35.

Firpo, S. (2007). Efficient semiparametric estimation of quantile treatment effects. *Econometrica* **75**, 259–276.

Frölich, M. (2004). Programme evaluation with multiple treatments. *Journal of Economic Survey* **18**, 181–224.

Galiani, S., P. Gertler, and E. Schargrodsky (2008). School decentralization: helping the good get better, but leaving the poor behind. *Journal of Public Economics* **92**, 2106–2120.

Girma, S., and H. Görg (2007). Evaluating the foreign ownership wage premium using a difference-in-differences matching approach. *Journal of International Economics* **72**, 97–112.

Görg, H., M. Henry, and E. Strobl (2008). Grant support and exporting activity. *Review of Economics and Statistics* **90**, 168–174.

Gould, D. (1994). Immigrant links to the home country: empirical implications for US bilateral trade flows. *Review of Economics and Statistics* **76**, 302–316.

Greenaway, D., N. Sousa, and K. Wakelin (2004). Do domestic firms learn to export from multinationals? *European Journal of Political Economy* **20**, 1027–1043.

Grossman, G. (1989). Promoting new industrial activities: a survey of recent arguments and evidence. Mimeo, Woodrow Wilson School, Princeton University.

Hall, P. (1998). *Cities in Civilization* (Oxford: Blackwell).

Han, C., and V. Terpstra (1988). Country-of-origin effects for uni-national and bi-national products. *Journal of International Business Studies* **19**, 235–255.

Hausmann, R., and D. Rodrik (2003). Economic development as self-discovery. *Journal of Development Economics* **72**, 603–633.

Head, K., and J. Ries (1998). Immigration and trade creation: econometric evidence from Canada. *Canadian Journal of Economics* 31, 47–62.

Head, K., and J. Ries (2001). Overseas investment and firm exports. *Review of International Economics* **9**, 108–122.

Heckman, J. (1981). Heterogeneity and state dependence, in S. Rosen (ed), *Studies on Labor Markets*. The University of Chicago Press.

Heckman, J., and R. Robb (1985). Alternative methods for evaluating the impact of interventions, in J. Heckman and B. Singer, *Longitudinal Analysis of Labor Market Data*. New York: Wiley.

Heckman, J., H. Ichimura, and P. Todd (1997). Matching as an econometric evaluation estimator: evidence from evaluating a job training programme. *Review of Economic Studies* **64**, 605–654.

Heckman, J., H. Ichimura, J. Smith, and P. Todd (1998). Characterizing selection bias using experimental data. *Econometrica* **66**, 1017–1098.

Heckman, J., R. LaLonde, and J. Smith (1999). The economics and econometrics of active labor market programs, in O. Ashenfelter and D. Card (eds), *Handbook of Labor Economics*. Elsevier.

Herander, M., and L. Saavedra (2005). Exports and the structure of immigrant-based networks: the role of geographic proximity. *Review of Economics and Statistics* **87**, 323–335.

Holland, P. (1986). Statistics and causal inference. *Journal of the American Statistical Association* **81**, 945–960.

Huang, R. R. (2007). Distance and trade: disentangling unfamiliarity effects and transport cost effects. *European Economic Review* **51**, 161–181.

Hudson, J., and P. Jones (2003). International trade in 'quality goods': signalling problems for developing countries. *Journal of International Development* **15**, 999–1013.

Imbens, G. (2004). Nonparametric estimation of average treatment effects under exogeneity: a review. *Review of Economics and Statistics* **86**, 4–29.

Imbens, G., and J. Wooldridge (2008). Recent developments in the econometrics of program evaluation. IZA Discussion Paper 3640.

Jaffe, A. (2002). Building program evaluation into the design of public research support programs. *Oxford Review of Economic Policy* **18**, 22–34.

Johanson, J., and J. Vahlne (1977). The internationalization process of the firm: a model of knowledge development and increasing foreign market commitments. *Journal of International Business Studies* **8**, 23–32.

Jordana, J., C. Volpe Martincus, and A. Gallo (2010). Export promotion organizations in Latin America and the Caribbean: an institutional portrait. IDB Working Paper 198.

Katsikeas, C. (1994). Export competitive advantages: the relevance of firm characteristics. *International Marketing Review* **11**(3), 33–53.

Katsikeas, C., and R. Morgan (1994). Differences in perceptions of exporting problems based on firms size and export experience. *European Journal of Marketing* **28**, 17–35.

Keng, K., and T. Jiuan (1989). Differences between small and medium sized exporting and non-exporting firms: nature or nurture. *International Marketing Review* **6**(4), 27–40.

Kessing, D. B. (1967). Outward-looking policies and economic development. *The Economic Journal* **77**, 303–320.

Klette, T., J. Moen, and Z. Griliches (2000). Do subsidies to commercial R&D reduce market failures? Microeconomic evaluation studies. *Research Policy* **29**, 471–495.

Kneller, R., and M. Pisu (2007). Export barriers: what are they and who do they matter to? Mimeo, University of Nottingham.

Koenig, P., F. Mayneris, and S. Poncet (2010). Local export spillovers in France. *European Economic Review* **54**, 622–641.

Koenker, R., and G. Bassett (1978). Regression quantiles. *Econometrica* **46**, 33–50.

Lach, S. (2002). Do R&D subsidies stimulate or displace private R&D? Evidence from Israel. *Journal of Industrial Economics* **50**, 369–390.

Lechner, M. (2002). Program heterogeneity and propensity score matching: an application to the evaluation of active labor market policies. *Review of Economics and Statistics* **84**, 205–220.

Lederman, D., M. Olarreaga, and L. Payton (2006). Export promotion agencies: what works and what doesn't. World Bank Policy Research Working Paper 4044.

Leonidou, L. C. (1995). Empirical research on export barriers: review, assessment, and synthesis. *Journal of International Marketing* **3**, 29–43.

Leonidou, L. C. (2004). An analysis of the barriers hindering small business export development. *Journal of Small Business Management* **42**, 279–302.

Leonidou, L. C., and M. Theodosiu (2004). The export marketing information system: an integration of the extant knowledge. *Journal of World Business* **39**, 12–36.

Meyer, B. D. (1995). Natural and quasi-experiments in economics. *Journal of Business and Economic Statistics* **13**, 151–161.

Miguel, E., and M. Kremer (2004). Worms: identifying impacts on education and health in the presence of treatment externalities. *Econometrica* **72**, 159–217.

Mesquita Moreira, M., C. Volpe Martincus, and J. Blyde (2008). *Unclogging the arteries: the impact of transport costs on Latin American and Caribbean trade.* Special Report on Integration and Trade. David Rockefeller Center for Latin American Studies, Harvard University, Cambridge, MA.

Moini, A. (1998). Small firms exporting: how effective are government export assistance programs? *Journal of Small Business Management* **36**, 1–15.

Moser, C., T. Nestmann, and M. Wedow (2008). Political risk and export promotion: evidence from Germany. *The World Economy* **31**, 781–803.

Naidu, G., and T. Rao (1993). Public-sector promotion of exports: a need-based approach. *Journal of Business Research* **27**, 85–101.

Nelson, P. (1970). Information and consumer behavior. *Journal of Political Economy* **78**, 311–329.

Pagés-Serra, C. (ed) (2010). *The Age of Productivity: Transforming Economies from the Bottom Up.* Washington, DC: Inter-American Development Bank.

PROEXPORT (2008). Prepare a su empresa para un evento internacional.

Rangan, S. (2000). Search and deliberation in international exchange: microfoundations to some macro patterns. *Journal of International Business Studies* **31**, 205–222.

Rangan, S., and R. Lawrence (1999). Search and deliberation in international exchange: learning from international trade about lags, distance effects, and home bias. NBER Working Paper 7012.

Rauch, J. E. (1996). Trade and search: social capital, sogo shosha, and spillovers. NBER Working Paper 5618.

Rauch, J. E. (1999). Networks versus markets in international trade. *Journal of International Economics* **48**, 7–35.

Rauch, J. E., and A. Casella (2003). Overcoming informational barriers to international resource allocation: prices and ties. *The Economic Journal* **113**, 21–42.

Rauch, J. E., and V. Trindade (2002). Ethnic Chinese networks in international trade. *Review of Economics and Statistics* **84**, 116–130.

Ravallion, M. (2008). Evaluating anti-poverty programs, in R. Evenson and P. Schultz (eds), *Handbook of Development Economics*. Amsterdam: North-Holland.

Roberts, M. J., and J. R. Tybout (1997). The decision to export in Colombia: an empirical model of entry with sunk costs. *American Economic Review* **87**, 545–564.

Robins, J., and A. Rotnitzky (1995). Semiparametric efficiency in multivariate regression models for repeated outcomes in the presence of missing data. *Journal of the American Statistical Association* **90**, 106–121.

Rosenbaum, P. R., and D. B. Rubin (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika* **70**, 41–55.

Rosenbaum, P. R., and D. B. Rubin (1985). Constructing a control group using multivariate matched sampling methods that incorporate the propensity score. *American Statistician* **39**, 33–38.

Roy, A. D. (1951). Some thoughts on the distribution of earnings. *Oxford Economic Papers* **3**, 135–146.

Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and non-randomized studies. *Journal of Educational Psychology* **66**, 688–701.

Rubin, D. B. (1980). Comment on 'Randomization analysis of experimental data: the Fisher randomization test'. *Journal of the American Statistical Association* **75**, 591–593.

Seringhaus, R., and P. Rosson (1990). *Government Export Promotion: A Global Perspective*. London: Routledge.

Seringhaus, R., and P. Rosson (2005). An analysis model of performance measurement of international trade fair exhibitors. Mimeo, Wilfred Laurier University, Waterloo, OT.

Smith, J. (2000). A critical survey of empirical methods for evaluating active labor market policies. *Zeitschrift für Volkswirtschaft and Statistik* **136**, 247–268.

Smith, J. A., and P. E. Todd (2005a). Does matching overcome Lalonde's critique of nonexperimental estimators? *Journal of Econometrics* **125**, 305–353.

Smith, J. A., and P. E. Todd (2005b). Rejoinder. *Journal of Econometrics* **125**, 365–375.

Souchon, A., and A. Diamantopoulos (1998). Information utilisation by exporting firms: conceptualisation, measurement and impact on export performance. In S. Urban and C. Nanopoulos (eds), *Information Management*. Wiesbaden: Gabler Verlag.

Spence, M. M. (2003). Evaluating export promotion programmes: UK overseas trade missions and export performance. *Small Business Economics* **20**(1), 83–103.

Stigler, G. J. (1961). The economics of information. *Journal of Political Economy* **69**, 213–225.

Storper, M., and A. Venables (2004). Buzz: face-to-face contact and the urban economy. *Journal of Economic Geography* **4**, 351–370.

Suárez-Ortega, S. (2003), Export barriers: Insights from small and medium-sized firms. *International Small Business Journal* **21**, 403–419.

Tanner, J. (1995). Curriculum guide to trade show marketing. Center for Exhibition Industry Research, Baylor University, Waco, TX.

Volpe Martincus, C. (2010). Odyssey in international markets: an assessment of the effectiveness of export promotion in Latin America and the Caribbean. Special Report on Integration and Trade. Inter-American Development Bank, Washington, DC.

Volpe Martincus, C., and J. Carballo (2008). Is export promotion effective in developing countries? Firm-level evidence on the intensive and extensive margins of exports. *Journal of International Economics* **76**, 89–106. (See also IDB Working Paper 201.)

Volpe Martincus, C., and J. Carballo (2009). Survival of new exporters in developing countries: does it matter how they diversify? IDB Working Paper 140.

Volpe Martincus, C., and J. Carballo (2010a). Entering new country and product export markets: does export promotion help? *Review of World Economics* **146**, 437–467. (See also IDB Working Paper 203.)

Volpe Martincus, C., and J. Carballo (2010b). Beyond the average effects: the distributional impacts of export promotion programs in developing countries. *Journal of Development Economics* **92**, 201–214. (See also IDB Working Paper 204.)

Volpe Martincus, C., and J. Carballo (2010c). Export promotion: bundled services work better. *The World Economy* **33**, 1718–1756. (See also IDB Working Paper 206.)

Volpe Martincus, C., and J. Carballo (2011). Export promotion activities in developing countries: what kind of trade do they promote? *Journal of International Trade and Economic Development*, http://www.informaworld.com/10.1080/09638199.2010.500741. (See also IDB Working Paper 202.)

Volpe Martincus, C., J. Carballo, and P. Garcia (2011). Public programs to promote firms' exports in developing countries: are there heterogeneous effects by size categories? *Applied Economics*, http://www.informaworld.com/10.1080/00036846.2010.508731. (See also IDB Working Paper 205.)

Wagner, J. (1995). Exports, firm size, and firm dynamics. *Small Business Economics* **7**, 29–39.

Wagner, J. (2001). A note on the firm size export relationship. *Small Business Economics* **17**, 229–237.

Wagner, J. (2007). Exports and productivity: a survey of the evidence from firm-level data. *The World Economy* **30**, 60–82.

Westphal, L. E. (1990). Industrial policy in an export propelled economy: lessons from South Korea's experience. *Journal of Economic Perspectives* **4**, 41–59.

Young, S. (1995). Export marketing: conceptual and empirical developments. *European Journal of Marketing*, **29**(8), 7–16.

APPENDIX: EMPIRICAL METHODOLOGY

In this appendix we briefly explain the main estimation methods used to generate the estimates reported in the chapter.[46] Let Y_{it} be (the natural logarithm of) firm i's total exports in year t.[47] Each year firm i may either participate in export promotion programmes ('1') or not participate in these programmes ('0'), but not both. Hence, firm i has two potential export outcomes, Y_{it}^1 and Y_{it}^0, which correspond to the participation and non-participation states, respectively. Further, let D_{it} be an indicator codifying information on assistance by the export promotion organisation. Specifically, D_{it} takes the value 1 if firm i has been assisted by the organisation in year t and 0 otherwise.[48] In this case, firm i's observed export outcome can be expressed as follows:[49]

$$Y_{it} = D_{it} Y_{it}^1 + (1 - D_{it}) Y_{it}^0 \qquad (2.1)$$

and the impact of trade support is therefore given by $\Delta Y_{it} = Y_{it}^1 - Y_{it}^0$. The fundamental problem of causal inference is that it is impossible to observe Y_{it}^1 and Y_{it}^0 for the same firm. Hence, the population of companies is generally used to learn about the properties of the potential outcomes and compute an average treatment effect. More specifically, when participation in the programmes under consideration is voluntary, it seems more relevant to determine their effects on those firms that participated and accordingly an *average treatment effect on the treated* is estimated:

$$\gamma = E(Y_{it}^1 \mid D_{it} = 1) - E(Y_{it}^0 \mid D_{it} = 1) = E(\Delta Y_{it} \mid D_{it} = 1), \qquad (2.2)$$

where $E(Y_{it}^1 \mid X_{it}, D_{it} = 1)$ is the expected (average) exports of those firms that have received export support and $E(Y_{it}^0 \mid X_{it}, D_{it} = 1)$ are the expected exports of these firms had they not been received this support. The parameter γ then measures the average rate of change in exports between these trade support statuses (Lach 2002).

Difference-in-differences and *matching difference-in-differences* are alternative methods for generating an appropriate sample counterpart for the missing information on outcomes had the firms not been assisted, and thereby to

---

[46]An explanation of the methods used in robust check exercises such as the dynamic panel data estimator proposed by Blundell and Bond (1998) or double-robust estimation (see, for example, Robins and Rotznisky 1995; Imbens 2004; Imbens and Wooldridge 2008; Chen *et al* 2009) can be found in the respective background technical papers (see http://hqdni09/research/book_detail.cfm?lang=en&pub_id=B-648).

[47]The use of the (natural) logarithm is partially motivated by the scale problem originating in the fact that our binary variable D does not capture the size of the assistance (Lach 2002). The presentation hereafter focuses on firms' total exports, but *mutatis mutandis* also applies to measures of export performance along the extensive margin and the intensive margin.

[48]We use the terms assistance, support, participation and treatment interchangeably.

[49]This is the potential outcomes framework due to, among others, Fisher (1935), Roy (1951) and Rubin (1974).

compute the effect of this assistance. Both procedures rely for identification on the assumption that there are no time-varying unobserved effects influencing selection into trade promotion programmes and exports (see, for example, Blundell and Costa Dias 2002).

Difference-in-Differences

In general, in order to calculate standard errors, a regression approach is used (Ravallion 2008). Thus, assuming that the conditional expectation function $E(Y \mid X, D)$ is linear and that unobserved characteristics, μ_{it}, can be decomposed into

- a firm-specific fixed-effect, λ_i,
- a year, common macroeconomic effect, ρ_t and
- a temporary firm-specific effect, ε_{it},

leads to the following error-components specification:

$$Y_{it} = X_{it}\theta + \gamma D_{it} + \lambda_i + \rho_t + \varepsilon_{it}. \tag{2.3}$$

This equation is estimated on the whole sample and, to create a common before-treatment period, on the subsamples formed by those firms that were never previously treated (thus yielding the effect of the first assistance) or those that were not assisted in the previous period (Lach 2002). Further, estimation of Equation (2.3) can potentially be affected by severe serial correlation problems (Bertrand *et al* 2004). Standard errors are then estimated allowing for an unrestricted covariance structure over time within firms, which may differ across them (Bertrand *et al* 2004).

A common treatment-effect (*ie* $\gamma = \gamma_i$ for all i) assumption underlies Equation (2.3). However, effects can vary across groups of firms. More formally, they are likely to be heterogeneous by observed covariates. Under heterogeneity, the correct specification of the estimating equation would be (Djebbari and Smith 2008)

$$Y_{it} = X_{it}\theta + (\gamma + \gamma_X X_{it})D_{it} + \lambda_i + \rho_t + \varepsilon_{it}. \tag{2.4}$$

Further, Equation (2.3) assumes linearity. This may lead to inconsistency as a consequence of potential misspecification (see, for example, Meyer 1995; Abadie 2005). Matching difference-in-differences does not impose this functional form restriction in estimating the conditional expectation of the outcome variable and therefore generate estimates that are robust to these potential specification errors.

Matching Difference-in-Differences

Formally, the estimator is given by

$$\hat{\gamma}^{\mathrm{MDID}} = \sum_{i \in \{I^1 \cap S*\}} \left\{ \Delta Y_{it} - \sum_{j \in \{I^0 \cap S*\}} W_{ij} \Delta Y_{jt} \right\} w_{ij}, \tag{2.5}$$

where I^0 (I^1) is the set of control (treatment) firms, S^* is the common support, W is the weight placed on comparison observation j for firm i and w accounts for the re-weighting that reconstructs the outcome distribution for the treated sample. The weights W depend on the cross-sectional matching estimator employed. Three alternative methods based on different metrics are generally used in the evaluation studies: the nearest neighbour, the radius and the kernel estimators.[50] Note that, in general, in order to reduce the dimensionality problem of matching, this is generally performed on the propensity to participate given the set of observable characteristics X, or *propensity score*: $P(X_i) = P(D_i = 1 \mid X_i)$ (Rosenbaum and Rubin 1983).

Since the propensity score is in fact based on fitting a parameter structure (probit or logit), its success in balancing the values of covariates between matched treatment and comparison groups needs to be tested. The quality of matching is thus evaluated using several alternative tests such as the stratification test; the standardised differences test; the t-test for equality of means in the matched sample; the test for joint equality of means in the matched sample or Hotelling test; and the pseudo R^2 along with the likelihood ratio test of joint insignificance of regressors in the propensity score before and after matching (see, for example, Smith and Todd 2005b; Girma and Görg 2007; Caliendo and Kopeinig 2008). Finally, the significance of estimated impacts is assessed using analytical, bootstrapped and subsample-based standard errors (Heckman *et al* 1998; Smith 2000; Abadie and Imbens 2006).

Multiple Programme Matching Difference-in-Differences

Interestingly, matching difference-in-differences can also be used to assess the relative effects of different trade assistance initiatives. Let export promotion policy be a bundle of S different programmes. There are accordingly $(S + 1)$ different mutually exclusive states (treatments) whose respective outcomes are denoted by $\{Y^0, Y^1, \ldots, Y^S\}$ and where outcomes correspond to a specific measure of export performance. Thus, Y_i^s is (the natural logarithm of) firm i's total exports if this firm is assigned to programme s. Similarly, Y_i^r is (the natural logarithm of) firm i's total exports if this firm is assigned to programme r, and so forth. In this case

$$\hat{y}^{s,r} = E(Y^s \mid D = s) - E(Y^r \mid D = s), \qquad (2.6)$$

where $D \in \{0, 1, \ldots, S\}$ is a variable indicating participating in a particular programme and $y^{s,r}$ is the expected (average) effect of programme s relative to programme r for a firm randomly drawn from the population of firms

---

[50]See, for example, Smith and Todd (2005a) for a formal definition of these estimators.

participating in programme s.[51] It can be shown that this average effect can be expressed as follows:

$$\hat{y}^{s,r} = E(Y^s \mid D = s) - E_{Pr|sr(X)}\{E[(Y^r \mid P^{r|sr}(X), D = r) \mid D = s]\}, \quad (2.7)$$

where

$$P^{r|sr}(x) = P^{r|sr}(D = r \mid D = r \vee D = s, X = x)$$

$$= \frac{P(D = r \mid X = x)}{P(D = s \mid X = x) + P(D = r \mid X = x)}.$$

In order to identify $y^{s,r}$, only information from the subsamples of participants in programmes s and r is required. When all values of s and r are of interest, binary conditional probabilities can be modelled and estimated separately over the $S(S-1)/2$ subsamples or the complete choice problem can be formulated in a model and estimated on the full sample with a multinomial probit (Lechner 2002). The methods described above produce estimates of average treatment effects. When, instead, we are interested in the distributional impacts of trade promotion, *quantile treatment effects* need to be estimated.

Quantile Treatment Effects

Formally, quantile treatment effects on the treated effects are given by

$$\Delta_{\tau|D=1} = q_{1,\tau|D=1} - q_{0,\tau|D=1}$$
$$= \inf_q\{\Pr[Y(1) \leqslant q] \geqslant \tau\} - \inf_q\{\Pr[Y(0) \leqslant q] \geqslant \tau\}, \quad (2.8)$$

where $\tau \in (0, 1)$ and inf denotes inverse function.

Under the conditional independence assumption and the common support condition assumptions, a consistent estimator of the quantile treatment effect on the treated can be obtained as the difference between the solutions of two minimisations of sums of weighted check functions (Firpo 2007):

$$\hat{\Delta}_{\tau|D=1} = \hat{q}_{1,\tau|D=1} - \hat{q}_{0,\tau||D=1}$$
$$= \operatorname{argmin}_q \sum_{i=1}^{N} \hat{\omega}_{1,i|D=1}\rho_\tau(Y_i - q) - \operatorname{argmin}_q \sum_{i=1}^{N} \hat{\omega}_{0,i|D=1}\rho_\tau(Y_i - q),$$
$$(2.9)$$

where the check function $\rho_\tau(\cdot)$ evaluated at the real number of a is $\rho_\tau(a) = a(\tau - 1\{a \leqslant 0\})$ and the $\hat{\omega}$s are the individual weights given by (Koenker and

---

[51] Notice that $y^{s,s} = 0$. In addition, if participants in programmes s and r differ in a non-random way, *ie* they systematically differ over the distribution of their characteristics, and programme effects vary with these characteristics, then the treatment effects on the treated are not symmetric, *ie* $y^{s,r} \neq y^{r,s}$.

Bassett 1978)

$$\hat{\varpi}_{1,i|D=1} = \frac{D_i}{\sum_{i=1}^{N} D_i} \tag{2.10}$$

$$\hat{\varpi}_{0,i|D=1} = \left[\frac{\hat{p}(X_i)}{(1 - \hat{p}(X_i))} \right] \left[\frac{1 - D_i}{\sum_{i=1}^{N} D_i} \right]. \tag{2.11}$$

In this context, selection on an unobservable determinant can be allowed for as long as we assume that this determinant lies on a separable individual specific component of the error term, *ie* using as outcome variable the first (logarithmic) difference of exports (Blundell and Costa Dias 2002). It should be noted that, in doing so, this procedure yields estimates of the impact of trade promotion actions across quantiles of the distribution of the growth rates of exports. In order to gain insights on effects of trade promotion actions across quantiles of the distribution of export levels, we have to compare the distribution of (lagged) export levels corresponding to firms in quantiles of the distribution of first-differentiated exports registering assistance effects of different magnitude.

Estimating Treatment Effects with Dichotomous Outcome Variables

Procedures such as difference-in-differences and matching difference-in-differences work well with continuous export performance measures along the extensive margin such as the (growth of the) number of export destinations and the number of products exported. However, with binary outcomes, standard procedures can lead to predictions outside the allowable range, and giving up the additivity assumptions to avoid potential misspecification without imposing additional assumptions may result in non-identification of the counterfactual distribution of outcomes (Athey and Imbens 2006).

As a consequence, to assess whether export promotion activities actually help firms reach new destination countries or introduce new export products, alternative estimation methods need to be used. A specific strategy has been recently proposed to address this issue (Aakvik *et al* 2005). This strategy consists of specifying and estimating an endogenous switching binary response model where selection into export promotion programmes and export outcomes are jointly determined and unobservables are generated by factor structures. Formally, assume the following export outcome equations of the assistance and non-assistance states and the following decision rule for using this assistance, respectively:

$$Y_{1i}^* = X_i \beta_1 + U_{1i},$$

$$Y_{1i} = \begin{cases} 1 & \text{if } Y_{1i}^* \geqslant 0, \\ 0 & \text{otherwise}, \end{cases}$$

$$Y_{0i}^* = X_i\beta_0 + U_{0i},$$

$$Y_{0i} = \begin{cases} 1 & \text{if } Y_{0i}^* \geqslant 0, \\ 0 & \text{otherwise}, \end{cases}$$

$$D_i^* = Z_i\beta_D + U_{Di},$$

$$D_i = \begin{cases} 1 & \text{if } D_i^* \geqslant 0, \\ 0 & \text{otherwise}, \end{cases}$$

where Y_{1i}^* is a latent index of adding a new country when receiving support and Y_{0i}^* is the corresponding latent index when not receiving support; X_i is a vector of observed random variables; β_0 and β_1 are sets of parameters; U_{0i} and U_{1i} are unobserved random variables with $U_{0i} \neq U_{1i}$, so that idiosyncratic gains from assistance are allowed for each firm; D_i^* is a latent index that determines whether a firm is assisted or not; Z_i is a vector of observed random background variables that determine selection into these programmes, such that those variables included therein but not included in X_i provide an identifying exclusion restriction; β_D is a set of parameters; and U_{Di} are unobservables (Aakvik *et al* 2003).

Unobserved heterogeneity is assumed to follow a factor structure and enter into the selection as well as the outcome equations (Heckman 1981; Aakvik *et al* 2005):

$$U_{Di} = \alpha_D\theta_i + \varepsilon_{Di}, \tag{2.12}$$

$$U_{1i} = \alpha_1\theta_i + \varepsilon_{1i}, \tag{2.13}$$

$$U_{0i} = \alpha_0\theta_i + \varepsilon_{0i}, \tag{2.14}$$

where θ_i is an unobserved firm-specific time-invariant factor and ε_D, ε_1, ε_0 are independent with respect to each other and of the exogenous variables in the model (Aakvik *et al* 2003). The αs are factor loadings in each equation that capture potential correlations among their error terms. In this case, the effect of the assistance by the organisation on assisted firms is given by

$$\Delta^{TT}(x, z, D = 1)$$
$$= E(\Delta \mid X = x, Z = z, D = 1)$$
$$= \Pr(Y_1 = 1 \mid X = x, Z = z, D = 1) - \Pr(Y_0 = 1 \mid X = x, Z = z, D = 1)$$
$$= \frac{1}{F_{U_D}(z\beta_D)}[F_{D,1}(z\beta_D, x\beta_1) - F_{D,0}(z\beta_D, x\beta_0)]$$
$$= \frac{1}{E(\Phi(z\beta_D/\sqrt{2}))} \int [\Phi(x\beta_1 + \alpha_1\theta) - \Phi(x\beta_0 + \alpha_0\theta)]\Phi(z\beta_D + \theta)\varphi(\theta)\,\mathrm{d}\theta.$$
$$\tag{2.15}$$

Since θ is not observed, it is integrated out assuming that $\theta \coprod (X, Z)$. The likelihood function for this one-factor model integrating out θ has the follow-

ing form:

$$L = \prod_{i=1}^{N} \int \Pr(D_i, Y_i \mid X_i, Z_i, \theta)\varphi(\theta)\, d\varphi,$$

where $\Pr(D_i, Y_i \mid X_i, Z_i, \theta_i) = \Pr(D_i \mid Z_i, \theta_i)\Pr(Y_i \mid D_i, X_i, \theta_i)$. This function's parameters can be estimated by maximum likelihood and the significance of the implied export support effect can be assessed based on bootstrapped standard errors.

Can Matching Grants Promote Exports? Evidence from Tunisia's FAMEX II Programme

JULIEN GOURDON, JEAN MICHEL MARCHAT, SIDDHARTH SHARMA
AND TARA VISHWANATH[1]

1 INTRODUCTION

A number of recent firm-level studies suggest that government export promotion activities, typically implemented through export promotion agencies, can have a positive impact on export performance (see, for example, Álvarez and Crespi 2000, 2004; Volpe Martincus and Carballo 2008; Görg *et al* 2009). National export promotion agencies have been part of most countries' national export strategy for a long time. While several past studies criticised their efficiency in developing countries, recent research (Lederman *et al* 2010) suggests that, despite heterogeneity across regions, levels of development and types of instruments proposed, they have—on average—an important and statistically significant impact on exports. These findings are broadly consistent with the common view that firms need assistance in reaching new export markets because of high information-related entry costs.

However, much remains to be learned about the effectiveness of specific export promotion instruments used by these agencies (Lederman *et al* 2010). Most export promotion agencies employ a wide array of instruments, ranging from the provision of public goods, such as export markets' analysis, to technical and financial assistance to individual firms. Participating firms are often assisted through multiple instruments, which cannot be disentangled in the available data. As a result, we know little about their relative effectiveness.

A recent study of Peru's national export promotion agency by Volpe Martincus and Carballo (2008) finds evidence that firms supported by its activities

exhibit significantly better export performance than comparable unsupported firms. But, like most such agencies, Peru's agency uses multiple instruments: it trains inexperienced exporters on the export process, marketing and negotiations; it produces markets' analysis; it provides information on trade opportunities as well as specialised counselling and technical assistance on how to take advantage of these opportunities; it coordinates and supports (and in some cases co-finances) firms' participation in international trade missions and trade shows; and it arranges meetings with potential foreign buyers. In the absence of data on the types of support received by individual firms, only the combined impact of this heterogeneous bundle of promotional activities could be measured. Another recent study, on US manufacturing plants by Bernard and Jensen (2004), measures the impact of state export promotion expenditures, but also lacks information on the exact instruments funded by those expenditures. While they are informative for gauging the usefulness of information-related support activities, the policy recommendations emerging from such studies can suffer from a lack of specificity.

In this chapter, we describe the results from an evaluation of a specific type of export promotion instrument: a matching grant. Starting in 2005, Tunisia's export promotion agency provided—through the FAMEX II—more than 1,000 firms with export-development assistance on a cost-sharing basis. Under the terms of this programme, the agency would meet 50% of the cost of export-development plans proposed by eligible firms approved for funding after a committee review.[2] We examine the impact of this FAMEX II programme using firm-level data collected through a purposely designed survey. The evaluation of the impact of this matching grant was conducted on the suggestion of World Bank colleagues who needed to draw lessons from FAMEX II for the preparation of a new export-development programme for Tunisia.

To be eligible for a FAMEX II grant, firms proposed export-development plans had to specify whether the set of activities aimed at developing new export products, new export markets and/or export skills in the case of first-time exporters. Broadly speaking, eligible activities were those which addressed informational constraints to entering export markets, such as market research, training and consulting. Thus, FAMEX was more focused than general-purpose grants for export-development investment, which include technology and physical capital investment such as those studied in Görg *et al* (2008) for firms in Ireland. Yet, being a grant, it was also less restrictive than programmes that directly provide services such as training: firms applying to FAMEX II had some flexibility in defining their export-development plan. Thus, we could expect FAMEX II to have had very different impacts on firms from the direct provision of consulting and other non-financial interventions, which have been the subject of most previous studies.

---

[2]FAMEX: Fonds d'Accès aux Marchés d'Exportation or 'Export Market Access Fund'.

Another distinguishing feature of FAMEX is the non-reimbursable co-financing that justifies its characterisation as a 'matching grant'. Firms applying to FAMEX knew that they would need to commit some of their own resources. Intuitively, this should have led to a better and more homogenous pool of applicants and better use of the funds compared with a pure grant. Indeed, this hypothesis is a key reason why matching grants have become an increasingly popular instrument for the development of small and medium-sized firms. But there is little rigorous evidence so far on their effectiveness (McKenzie 2010). Our study is one of the first using detailed micro-data to examine the effects of a matching grant programme for firms.

The main methodological challenge in identifying the impact of a firm-level assistance programme such as FAMEX II involves the measurement of a counterfactual: what would have been the outcomes for programme participants in the absence of the programme? Since this cannot be observed, we have to estimate the counterfactual through the outcome of a comparison or control group of firms that did not participate in the programme. The ideal control group should have no systematic differences from the programme participants in terms of determinants of programme outcomes. This ideal is achieved in experimental evaluation designs in which the assignment to the programme is randomised across firms. Since this chapter deals with an *ex post* evaluation, this was not possible. To our knowledge, this is a constraint shared by all existing studies of export promotion schemes.

We employed the next-best alternative, matching differences-in-differences (DID) estimation. This type of estimation has been used in several recent firm-level studies such as Volpe Martincus and Carballo (2008), Görg *et al* (2008) and Arnold and Javorcik (2009). Programme participants are paired with, or 'matched' to, observably similar firms and the changes in their outcomes before and after the programme are compared. The matching ensures that differences between the participant group and the control group that are due to observable characteristics are controlled for. By measuring the change in outcomes, the method also controls for time-invariant unobservable differences between the two groups.

The matching DID results suggest that FAMEX II had positive impacts on export growth. The estimated average annual growth rate of export values during the programme period 2004–8 is approximately 38.9% higher for FAMEX II participants than for the control group. The estimates suggest that FAMEX II improved the extensive margin of export performance. The estimated average annual growth in the number of exported products is 5% higher for the FAMEX II participants, and the corresponding estimate for the number of export destination countries is 4.5%. Sensitivity tests show that the estimation of linear differences-in-differences regressions gives qualitatively similar results to the matching DID estimates. The estimated impacts of FAMEX II on total firm sales and employment are weak, suggesting some reallocation between exported and non-exported products within supported firms. Our

data do not, unfortunately, allow us to examine if this was accompanied by increased profits.

Our analysis produces two novel and interesting findings. First, estimates show that FAMEX II had a disproportionate impact on first-time exporters, *ie* firms that started exporting after receiving FAMEX II assistance. Second, our estimates show that export promotion grants can be effective in promoting exports by services firms. The inclusion of services firms in the analysis is a novel contribution of our study. Together, these findings suggest that the informational costs of reaching new export markets are higher for services firms and for first-time exporters.

Interestingly, the estimates of the impact of FAMEX II on export growth are higher than those reported in previous studies of export promotion agencies. This could be because FAMEX was a matching grant, whereas most previous studies examined programmes that provided a mix of largely non-financial support, such as consulting, and a relatively limited grant element, such as funding for foreign trips. The cost-sharing element might have led to a self-selection of better-performing firms into the programme. Another possibility is that the FAMEX II programme spent more per recipient than other programmes: the grant could go up to US$100,000 per firm, and the average size of the grants actually disbursed was US$70,000. Lacking similar details on the programmes examined in previous studies, it is impossible to pursue this argument in depth. At the very least, this finding underscores the need to focus more on the specifics of export promotion instruments in future studies.

We would like to emphasise the methodological challenges faced in this and similar studies, in the hope that it will lead to more serious consideration of *ex ante* evaluation design of trade-related programmes and policies. Some of the key information, such as initial sales and employment, was drawn from retrospective data collection and suffered from some non-response and potential recall bias. Moreover, as explained in greater detail in Section 3, *ex post* evaluation techniques cannot eliminate bias arising from differences in the distribution of time-varying unobservable across the participants and the control group. To give an example of a potential overestimation bias, it could be that the programme selected firms that would have done better than others even without the matching grant (even after conditioning on observables). While we know that there was an application review process, we can only control for observable rejection crudely. A contrasting possibility is that the programme selected those firms most 'in need' of support, in which case we might have underestimated the impact. Yet another concern is that the programme could have had spillover effects; indeed, informational spillovers across exporter have long been theorised about and empirically tested for, and are a key economic justification for subsidies to exporters (see, for example, Krugman 1992; Roberts and Tybout 1997). However, lacking external infor-

mation on the nature of spillovers in our sample, we are uncertain of what they imply for our differences-in-differences estimates.[3]

The rest of this chapter is organised as follows. In Section 2 we describe Tunisia's export promotion scheme. The methodology is described in Section 3. In Section 4 we describe the survey data, while in Section 5 we present the main results. Section 6 concludes.

2 TUNISIA'S EXPORT PROMOTION MATCHING GRANT SCHEME

Matching grants (MGs) are a short-term and temporary mechanism that partially finance activities to promote improvements in the private sector. Over the years, MGs have become a common tool in private sector development (PSD) projects. Since 1994, the World Bank has financed a total of 37 projects with MG components, of which 22 were still active in 2008 (World Bank 2009).

Despite being commonly used, matching grants face criticisms. These focus on 'additionality' (funding activities that firms would have financed anyway) and 'selectivity' (failure to establish a distinction between private benefits and broader economic benefits) issues (Biggs 1999), or the sustainability of the support (Phillips 2001).

Although there have been assessments of MGs, there have been few attempts to assess the instrument with recent impact-evaluation (IE) techniques, or, for that matter, many PSD types of interventions. This is changing (McKenzie 2010) as demonstrated by a recent wave of IEs in PSD areas such as SME support programmes (Tang 2009), rainfall insurance (Giné and Yang 2009) and regulatory reforms (Bruhn 2008).

Since 2000, two export development projects have been implemented in Tunisia. In addition to activities in the area of trade facilitation and the establishment of a pre-shipment export guarantee facility, both projects implemented a matching grant scheme. The FAMEX II programme was jointly financed by the Tunisian government and by the World Bank within the framework of the Second Export Development Project (2005–11).[4] FAMEX II had three components:

- building the capacity of export associations and chambers of commerce to provide export assistance to their members (mainly SMEs);
- strengthening the export consulting sector;
- assisting individual enterprises through matching grants.

This study is focused on the firm-level intervention component of the programme, FAMEX II.

---

[3]One concern is that there may have been spillovers from participants to control firms, which would attenuate the estimated impacts.

[4]Tunisia's first export market access fund (FAMEX I) operated between 2000 and 2004.

FAMEX II began in 2005 and provided firms with export-marketing assistance on a cost-sharing basis. The stated objective of the export-marketing assistance was to help firms start exporting if they had little or no export experience, and either diversify their markets or develop new products if they did have export experience. The instrument was a non-reimbursable matching grant covering 50% of expenditures, up to a maximum of US$100,000 per firm, awarded to selected firms for eligible activities undertaken within the framework of an export-development plan. The cost-sharing design was intended to help ensure sound project design and implementation and reduce the likelihood of misallocation of funds. It was expected that the commitment of firms to their projects would be significantly increased because of the cost-sharing nature of FAMEX II.

In order to receive a FAMEX II matching grant, interested firms had to prepare an export-development plan specifying a set of activities focused on developing new export products, new export markets and/or first-time exporting. Applicants had to demonstrate that they had given serious consideration to the feasibility of the proposed activities. The export plan was assessed by a review committee composed of senior experts from the management team, and the review process included detailed interviews with the applicants. Upon receiving approval for a grant, firms were required to sign a letter of agreement binding them to present defined deliverables for evaluation by the management team. FAMEX II management also helped successful applicants refine their export plan, and facilitated technical assistance and training during the implementation of the export plan.

Firms had to be larger than a specified size to be eligible for the grant. The minimum thresholds for eligibility were about US$140,000 and US$70,000 in sales, respectively, for manufacturing and services firms. Data from the official registry of firms in Tunisia indicate that, as of 2004, nearly 6,000 manufacturing and 20,000 services firms (out of a total of 60,000 and 436,000, respectively) were eligible to apply to FAMEX II. By December 2009, 1,710 applications were submitted to FAMEX II, representing nearly 7% of all eligible firms. Perhaps not surprisingly, services firms were significantly less likely to express interest in exporting than manufacturing firms: fewer than 3% of eligible services firms applied to FAMEX II. Moreover, firms that were already exporting were more likely to apply: of the 2000 eligible manufacturing firms that were already exporting in 2004, about 20% applied to FAMEX II. By December 2009, 1,231 applications (72% of all applicants) had been accepted into the programme.

We consider 2008 as the last year of the analysis, given that the survey data collection began in 2009 at a time when the accounting information for that year was not yet available. Moreover, in view of the extraordinary circumstances following the 2009 downturn, it seems appropriate to take 2008 to be the last year of our study. Hence, we focus on firms that had applied and completed their FAMEX II-funded programme by 2008. Hence, we exclude from the

analysis 429 of the successful applicants who, as of 2008, were still in the middle of their proposed export plan (largely because they had applied towards the end of the sample period). We also exclude from the main analysis some successful applicants who dropped out of the FAMEX II programme without disbursement ('drop-outs'), either because they were deemed by the management to make insufficient progress, or because they voluntarily changed their plans.[5] Thus, the universe of FAMEX II 'treated' firms for the empirical approach based on survey data—including firms which received the grant and firms whose proposed export plan was deemed to have been completed by 2008—consists of 336 firms.

3 EMPIRICAL STRATEGY

We evaluate the impact of FAMEX II by comparing outcomes before and after the programme across Tunisian firms that receive this matching grant support (the 'treatment' group) and firms that are similar but did not receive the grant (the 'control' group). Ideally, these groups should have been 'identical' at the start of the programme, in the sense that outcomes were expected to be same across the treatment and control groups *in the absence of the grant.*

Formally, let D_i be an indicator for a firm i receiving the treatment (FAMEX II support). That is, $D_i = 1$ if firm i received support and $D_i = 0$ otherwise. Let Y_i^1 be an outcome of interest (*eg* total exports) for firm i in period t. Let Y_i^0 be defined as the outcome for firm i had it not received FAMEX support. We are interested in estimating the difference between the expected (average) outcome for treated firms ($D_i = 1$) when they received FAMEX support and their expected outcome in the absence of treatment:[6]

$$\mu = E(Y_i^1 \mid D_i = 1) - E(Y_i^0 \mid D_i = 1)$$

Since $E(Y_i^0 \mid D_i = 1)$ is not observed, the treatment effect is identified through $E(Y_i^0 \mid D_i = 0)$, the observed average outcome of firms which did not get treated. The best-case scenario is one where the outcome without treatment can be assumed to be on average unrelated to assignment to treatment. That is, $E(Y_i^0 \mid D_i = 0) = E(Y_i^0 \mid D_i = 1)$. This is ensured if assignment to treatment is random.

This best-case scenario is ruled out in our case, as acceptance into FAMEX was not random: the most promising applicants were systematically more likely to have been accepted by the senior expert team managing the matching grant. In this context, we have to use non-experimental (or *ex post*) evaluation techniques. Such techniques essentially require the identification of a control

---

[5]However, these 'drop-outs' will be considered as part of the treatment group in some robustness checks shown in Section 5.

[6]In the literature, this is known as the *average effect of treatment on the treated* (ATT).

group that is 'similar' to the treatment group, and then compare the treat-
ment and control groups after controlling for differences along observable
dimensions (such as size, age, sector and the prior exporting status of firms).
Implicit behind this approach is the assumption that, once these observables
are controlled for, all of the difference in outcomes across these two groups
of firms is due to their different FAMEX statuses. Formally, we aim to estimate

$$\gamma = E(Y_i^1 \mid X_i, D_i = 1) - E(Y_i^0 \mid X_i, D_i = 0),$$

where X_i is a set of observable firm characteristics. The identification assump-
tion is that $E(Y_i^0 \mid X_i, D_i = 0) = E(Y_i^0 \mid X_i, D_i = 1)$.

We implement this evaluation approach using matching combined with
differences-in-differences estimation, a technique that has been used in a
number of recent firm-level studies (Volpe Martincus and Carballo 2008;
Arnold and Javorcik 2009). The differences-in-differences estimation implies
that we compare the *change* in outcomes before and after FAMEX II across the
treatment and the control groups. First-differencing, *ie* the consideration of
the change in outcomes, accounts for unobserved fixed differences between
the two groups of firms. In our preferred strategy, treated firms are 'matched'
with control firms in order to balance the two groups on observable vari-
ables. This is done though propensity score matching (PSM), which involves
matching firms on the probability of treatment based on observables.[7] The
underlying assumption is that, conditional on the propensity score, there are
no time-varying unobserved determinants of the outcome that differ system-
atically across control firms and treatment firms.[8]

As a sensitivity check, we show that the results from linear-regression-
based differences-in-differences estimation are similar to those from the PSM
approach. Specifically, we regress the change in outcome on observable base-
line characteristics X_i and an indicator for treatment:

$$\Delta Y_i = \alpha + \beta X_i + \gamma D_i + e_i.$$

The treatment effect is given by γ. Note that fixed differences across treatment
and control groups are accounted for because the outcome is specified in

---

[7] Specifically, every treatment firm was compared with a weighted average of control
firms: the higher the weights, the closer the control firm's propensity scores to that of the
treatment firm in question.

[8] Rosenbaum and Rubin (1983) show that adjusting solely for differences between
treated and control units in the propensity score removes all biases associated with differ-
ences in observables (covariates). Heckman *et al* (1997) provide more evidence on matching
differences-in-differences estimators, decomposing the sources of bias using comparisons
with experimental estimates. Matching methods eliminate the estimation bias due to dif-
ferent distributions of the propensity score across control and treatment groups, as well
as that due to non-overlapping supports of the distributions. However, the bias due to
differences in the distribution of time-varying unobservable is not eliminated.

differences. The assumption for unbiased estimation of the treatment effect is that the treatment assignment is uncorrelated with the error term e_i.

We also considered using an alternative quasi-random evaluation technique (regression discontinuity approach), but found it to be infeasible. This approach would have involved comparing accepted firms with those that were just below some threshold for acceptance, the assumption being that the latter are a close approximation to the ideal control group. Regression discontinuity relies on being able to identify a sufficient number of 'marginal' applicants: those considered just good enough to receive the grant according to some cut-off, with some not winning the grant due to limited funds. This was not possible in our case, since we had no objective rating of applications to FAMEX II that could provide for such threshold.[9] Rejected applications were judged by the FAMEX II the senior expert team to have a lower chance of success than accepted applications, but there is no information on how inferior they were judged to be. Moreover, there were too few rejected applications to yield a sufficient sample of marginal applicants.

It is important to note that our control group includes both rejected applicants and firms which have never applied to FAMEX II. Including non-applicants in the control group helps to control for time-varying unobserved determinants that are common to FAMEX II recipients and other firms (conditional on observables).

We could make a case for keeping only rejected applicants to FAMEX II in the control group: they are more easily comparable with treated firms in the sense that both groups expressed an interest in the programme and prepared an export-development plan for cost-sharing. But the fact that rejected applications were rated worse than successful applications suggests that the rejected firms are likely to have had worse outcomes than the treated firms even in the absence of FAMEX II. Thus, comparing the group of treated firms with a control group including only rejected firms could overstate the impact of the FAMEX II programme. Nevertheless, we present in Section 5 results from a specification where treated firms are compared with rejected firms alone, with the caveat that this specification suffers from small sample size bias.

Including non-applicants to FAMEX II in the control group leads to another set of methodological questions, the crucial unknown being their reason for not applying. If those firms did not apply due to lack of awareness about the programme, then there is no obvious bias in using them as the control group. But there are other potential explanations which, if true, could lead to biases in the estimated effects of FAMEX II. First, some firms with export-development plans might not have applied to FAMEX II because they did not need the grant since they faced no constraints in financing fully their plans (they were able to

---

[9]If applicants to FAMEX had been given objective scores and if a sufficient number of applicants had been rejected because they were rated just below the acceptance threshold, the regression discontinuity approach could have been followed.

rely on retained earnings or bank credit, for example). Comparing the treated firms with such firms could underestimate the effect of the programme, since such firms are likely to have done better than the treated firms in the absence of the programme.

Second, some firms might not have applied because they had no interest in developing their exports. This could lead to an overestimation of the impact of the FAMEX II programme, since such firms would have experienced lower export growth than the treated firms even in the absence of the programme. Since the expected net bias is indeterminate, and since limiting the comparison group to the rejected firms would not necessarily avoid biases, we decided that there was more to gain from including non-applicants in the control group than not doing so. But the potential biases should be kept in mind when interpreting the results.

4   FIRM-LEVEL SURVEY DATA

We collected primary data on a sample of FAMEX II recipients and non-recipient firms through a firm-level survey conducted in 2009. This effort was necessary, given that the FAMEX II programme was implemented without any baseline data collection on eligible Tunisian firms. Specifically, our survey covered a sample of 420 firms allocated evenly between FAMEX recipients (treated firms) and non-recipients (the control group). In addition, the survey also covered random samples of 40 firms rejected by FAMEX and of 40 firms that dropped out of FAMEX, leading to a total sample size of 500 firms.

The analysis of prior data from Tunisia's national statistical institute (L'Institut National de la Statistique (INS)) on formal sector firms indicated that, on average, FAMEX II recipients differed from other firms along observable dimensions. For example, FAMEX II recipients were more likely to be exporting before the programme and their distribution across industrial sectors was different. This implied that a random sample of non-recipient firms would have differed systematically from a random sample of recipients. Hence, in order to improve the efficiency of our estimation, we attempted to 'balance' the control and treatment samples *ex ante* in terms of observable characteristics. That is, we wanted to ensure that the control sample of non-recipients would look similar—in terms of the distribution of key observable characteristics—to a random draw of FAMEX II recipients. We implemented this through stratified random sampling. The universe of non-recipient firms taken from the INS 2007 census of firms was grouped into strata based on three key observable characteristics: size, prior exporting status and sector. Every stratum was allocated a sample size proportional to the number of FAMEX II recipients in the corresponding recipient stratum. These strata allocations were filled through random sampling within every stratum.

The firm-level survey collected data on export volumes, products and destinations, as well as sales and employment and other firm characteristics (age,

type of ownership, tenure of the firm manager) for two years: 2004 and 2008. Information on location and sector was also collected for each firm. In the absence of baseline data, the survey had to rely on retrospective information for 2004. This contributed to a less than 100% response rate on some key questions such as those on the number of employees and sales in 2004. Given that those retrospective variables were essential for estimating the propensity of receiving a FAMEX II grant, non-response reduced the usable data set to 435 firms. While we have no reason to believe that the non-response to retrospective information was systematically correlated with unobserved determinants of export performance, we also cannot discard it as a potential source of bias.

After estimating the propensity scores that will be described in Section 5, in order to better balance the treatment and control groups on observables we dropped from the sample the FAMEX firms whose observable characteristics were too extreme and prevented them from being matched to any control firms (*ie* firms outside the region of common support in the propensity score matching). This resulted in a final sample of 428 firms (195 FAMEX firms and 202 untreated firms), for which we present descriptive statistics below.

The distributions of the treatment and control subsamples in terms of their sector, location, employment and sales categories are presented in Table 3.6. By design, the stratified random sample should have resulted in a balance between the subsamples along the stratification variables. But, as described above, the final usable sample was a subset of the intended sample, and this reduced the extent of balance *ex post*. Thus, while the sectoral distributions of FAMEX II firms and control firms are shown to be quite similar, they exhibit some minor differences, such as a lower representation of chemicals and a higher representation of other services in the FAMEX II subsample. Approximately 29% and 19% of the FAMEX II and control subsamples, respectively, consist of services firms. Similarly, there are minor and non-systematic differences across FAMEX II and control subsamples in terms of their distributions across sales and employment categories. Since location was not used for stratification, there is a larger difference in the distribution of firms by location: compared with the control group, FAMEX II firms are more concentrated in Tunis and less so in the eastern region.

Table 3.7 presents the means of other characteristics for the two subsamples. On average, most characteristics, such as the year in which the firm was set up, the years since the current owner has been managing the firm, the fraction of firms which started exporting over the sample period, the fraction of firms which export all their output, and the years of first export, are similar across FAMEX II firms and control firms. However, the control group includes more non-exporters than the treated group. Another difference concerns the share of domestic capital versus foreign capital: FAMEX firms have less foreign capital than control firms. This suggests one explanation for why firms approached FAMEX: they have less access to foreign sources of funds or to foreign networks which could help them exploit new export opportunities.
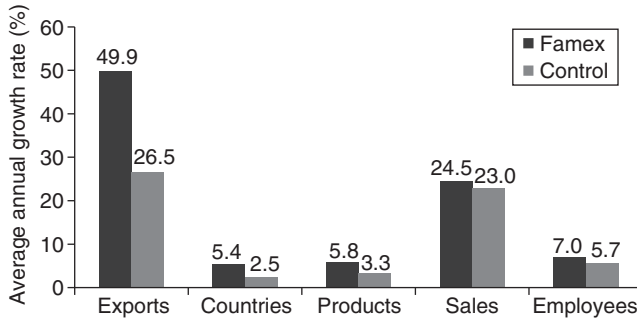
Figure 3.1: *Average annual growth rate over 2004–8.*
Source: authors' estimates based on data from the firm-level survey. Note: for each variable X, the annual rate of growth displayed is computed as $((\exp(Z))^{1/4} - 1) \times 100$, where Z is the change in log exports between 2004 and 2008 averaged across all firms in a group (FAMEX or control).

5   RESULTS

Figure 3.1 presents the raw (unmatched) differences in the growth rate of exports and other key outcomes across the treatment and control groups of firms between 2004 and 2008. FAMEX II recipients experienced on average faster growth in the total value of exports (the intensive margin), as well as in the number of exported products and export destination countries (the extensive margin). Specifically, the average annual growth rate of the total value of exports was almost 23 percentage points higher for FAMEX firms than for control firms. However, none of the differences are statistically significant, as revealed by Table 3.8, where these estimates and their *t*-statistics are shown. Qualitatively similar results are obtained when the sample is restricted to including only firms already exporting in 2004. However, in that case the estimated differences are smaller than those for the full sample, suggesting that the FAMEX II programme had a disproportionate impact on new exporters.

To pursue further the differences between treatment and control firms highlighted by Figure 3.1, we estimate a probit regression to explain the propensity of Tunisian firms to receive a FAMEX II grant during 2005–8. The regressors considered are firm characteristics in the baseline year 2004: age; location; sector; number of employees; sales; exporting status; share of domestic capital; the number of years the current manager has been with the firm. The probit estimation results are presented in Table 3.8. The noteworthy findings are that older firms, larger firms, firms with a higher share of domestic capital, firms located in Tunis and firms that were not exporting in 2004 are significantly more likely to receive a FAMEX II grant. The predicted probabilities using the estimated coefficients in Table 3.8 give the propensity score for each firm that is used for PSM and the identification of a control group of firms that is similar to the treated firms.
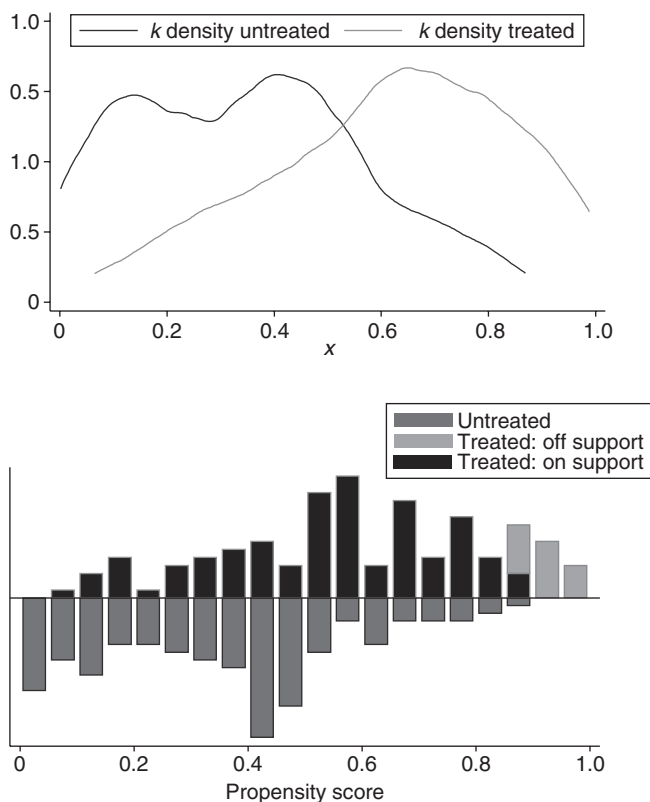
Figure 3.2: *Distribution of propensity score for FAMEX (treated) firms and control (untreated) firms.*
Source: authors' estimates based on data from the firm-level survey.

The distributions of the generated propensity scores within the treated and control groups are shown in Figure 3.2. As expected, the distribution of propensity scores for the treated group is to the right of the control group's distribution. Critically, this figure shows that the distributions have a large common support, which implies that most treated firms can be matched to one or more control firms on the nearness of propensity scores. Firms which are not on this common support—those with extreme scores—are dropped from the sample and are not included in the matching DID estimation as they cannot be 'matched' to firms in the control group. We also note that, since control firms with higher predicted propensity to be FAMEX II recipients were on average 'nearer' to more of the treated firms, the PSM gave more weight to such firms than did the unmatched comparison. Intuitively, these are the reasons why the results from the matching DID approach should be an improvement over results from the unmatched comparison shown in Figure 3.1.

Table 3.1: *Impact of FAMEX on export outcomes with matching difference-in-difference method: growth over 2004–8.*

Variable	Treated group	Control group	Difference	t-statistic
Change in log exports	1.686	0.373	1.313	2.13
Change in log number of exported products	0.226	0.056	0.170	1.64
Change in log number of export destinations	0.225	0.069	0.155	1.94

Note: the comparison is made for a matched sample of 150 FAMEX firms and 169 control firms.

Table 3.1 shows the matching DID estimates (and corresponding t-statistics) for the impact of FAMEX II on Tunisian firms.[10] The matched estimates are markedly higher than the unmatched estimates. For example, the matching-DID estimate of the differences in the change in log total exports across treated and control firms over the 2004–8 period is 1.313, compared with the raw difference of 0.68 (see Table 3.8). This estimated effect of FAMEX II of 1.313 implies that the average annual growth rate of total exports is 38.9% higher for the FAMEX treatment group than for the matched control group.[11] Unlike the raw difference in Figure 3.1 and Table 3.8, this matched difference is statistically significant at the 5% confidence level. Figure 3.3 plots the matching DID impact of FAMEX II on annual growth rates of total exports, the number of exported products and of destinations served by Tunisian firms.

Table 3.1 and Figure 3.3 suggest that FAMEX II improved the extensive margin of Tunisian firms' exports between 2004 and 2008. The estimated average annual growth in the number of exported products is approximately 5% higher for treated firms, though this impact is not statistically significant. The estimated average annual growth in the number of export destination countries is 4.5% higher for treated firms, and it is statistically significant at the 10% confidence level.

The matching-DID estimates suggest that FAMEX II succeeded in its objective of export promotion. Once again, we note that care should be taken in drawing causal inferences from these results. The validity of such inference depends on the underlying identification assumption that, conditional on the propensity score, unobserved determinants of export growth did not differ systematically across the treatment and control groups. Interestingly, the esti-

---

[10]The matching estimator used for the PSM is the kernel estimator.

[11]The estimate of 1.313 points implies that the growth rate in the level of exports during that period was $[\exp(1.313) - 1] \times 100 = 271\%$ higher for firms assisted by FAMEX, corresponding to a 38.9% higher average annual growth rate $(((2.71 + 1)^{1/4} - 1) \times 100)$.
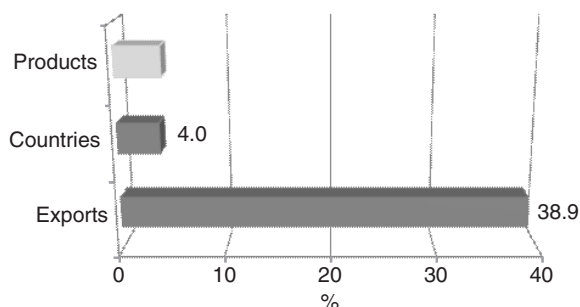
Figure 3.3: *Impact of FAMEX on average annual growth rates over 2004–8 (in per cent). Source: authors' estimates based on data from the firm-level survey. Note: the darker bars indicate the statistically significant impacts of FAMEX.*

Table 3.2: *Impact of FAMEX on other outcomes with matching difference-in-difference method: growth over 2004–8.*

Variable	Treated group	Control group	Difference	*t*-statistic
Change in log sales	0.923	0.481	0.442	0.97
Change in log number of employees	0.282	−0.066	0.349	2.84

Note: the comparison is made for a matched sample of 145 FAMEX firms and 152 control firms.

mates in Table 3.1 are rather high compared with those obtained in previous studies, using similar techniques, of the impact of export promotion agencies in Latin America as discussed by Volpe in Chapter 2 in this volume. This could be because FAMEX II offered funding for market development, whereas most export promotion agencies offer only consulting services. Moreover, it could be that the cost-sharing design of FAMEX II did an effective screening job. Another difference is that prior studies have focused on manufacturing firms, whereas in the case of Tunisia about 30% of FAMEX II beneficiaries were services firms.

In addition to export outcomes, we examined the impact of FAMEX II on sales and employment, and the estimates are presented in Table 3.2. The results are mixed: the estimated impact on the growth rates of both firm sales and firm employment are positive, but only the latter is statistically significant. This could be because of an important degree of non-response on sales and employment, or because retrospective sales data are less reliable. The impact of FAMEX II on average annual growth rates is 10.8% for sales and 9.1% for employment, which is markedly lower than that for the total value of exports. This could be because only about 30% of FAMEX II recipients exported all of their output. Another possible explanation for the weaker results for sales and employment is that FAMEX II recipients did not expand their total

Table 3.3: *Impact of FAMEX on export outcomes with matching difference-in- difference method including drop-outs in treatment group: growth over 2004–8.*

Variable	Treated group	Control group	Difference	t-statistic
Change in log exports	1.509	0.412	1.097	1.84
Change in log number of exported products	0.193	0.062	0.130	1.31
Change in log number of export destinations	0.204	0.051	0.152	1.97

Note: the comparison is made for a matched sample of 178 FAMEX firms (including drop-outs) and 164control firms.

production, but rather reallocated it towards exports. Unfortunately, we are unable to test whether this resulted in higher profits for treated firms, since we do not have reliable expenditure data.

As mentioned earlier, some successful FAMEX II applicants dropped out of the programme without disbursement, either because they were deemed by the expert team to have made insufficient progress, or because they voluntarily changed their plans. Being unsure of the reason for drop-out, we did not include them in our main analysis. To the extent that not using the FAMEX II grant was the firm's decision, based on a revaluation of its plan, it can be argued that those firms should be part of the treatment group. Hence, Table 3.3 presents alternative results where 40 FAMEX II drop-outs are added to the treatment group.[12] The inclusion of drop-outs lowers the estimated impact of FAMEX II, though the estimates remain statistically significant. This indicates that drop-outs performed significantly worse than other FAMEX II recipients. The problem with drawing a clear inference from this result is that if dropping out was due to a late rejection from the expert team, then there is a case for treating the drop-outs like other rejects, *ie* as part of the control group. Under that assumption, the estimated impact of FAMEX II would in fact increase.

As a sensitivity test, we estimate the impact of FAMEX II through difference-in-differences ordinary-least-square (OLS) regressions. The first-differenced outcome (for example, the change in the log of exports between 2008 and 2004) is regressed on an indicator for being a FAMEX II recipient and on firm characteristics, as discussed in Section 3. For comparability with the matching DID estimation, the same characteristics as those used to estimate the propensity scores are considered, and firms that are not on the common support are dropped from the estimating sample. The OLS estimates of the

---

[12]The sample size of drop-outs is proportional to their share in the universe of successful applicants.

Table 3.4: *OLS estimation of difference-in-differences regressions: growth over 2004–8.*

	Dependent variable:					
	excluding firms that drop out of FAMEX			including firms that drop out of FAMEX		
	Change in log exports (1)	Change in log exported products (2)	Change in log export destinations (3)	Change in log exports (4)	Change in log exported products (5)	Change in log export destinations (6)
FAMEX	1.325 (0.524)**	0.176 (0.097)*	0.156 (0.070)**	1.038** (0.507)**	0.127 (0.092)	0.132* (0.068)*
Number of observations	319	319	319	342	342	342
R^2	0.304	0.174	0.185	0.304	0.176	0.171

Note: robust standard errors in parentheses; *significant at 10%; **significant at 5%; ***significant at 1%. The regressions include all the independent variables included in the probit for the propensity to receive FAMEX assistance shown in Table 3.9. The sample includes only firms in the common support.

impact of FAMEX II shown in Table 3.4 are consistent with the matching DID estimates.

In the regression of growth in the value of exports shown in column (1) of Table 3.4, the coefficient on the FAMEX II indicator is significant at the 5% confidence level and implies that the impact of FAMEX II on the average annual growth rate of total exports between 2004 and 2008 was 39.3%, which is very close to the matching DID estimate. The same holds for the OLS estimate of the impact of FAMEX II on the number of exported products and the number of destinations served. Unreported OLS estimates of the impact of FAMEX II on sales and employment are positive and close to the matching DID estimates, but are significant only for employment. When drop-outs are included in the group of FAMEX II recipients in columns (4)–(6) of Table 3.4, the estimated effects of FAMEX decline in value but remain significant.

While in Table 3.4 we present an average impact across all Tunisian firms, it is possible that FAMEX II assistance had heterogeneous impacts across firms. We examine whether the impact differs across sector (manufacturing versus services) and according to the firm's exporting status prior to 2005. These differential impacts are estimated by including an interaction term between the FAMEX II treatment dummy and indicators for the firm's sector and the firm's prior exporting status in an OLS difference-in-differences regression. The estimates are presented in Table 3.5 and should be compared with columns (1)–(3) in Table 3.4 (where FAMEX II drop-outs are excluded). The results suggest that the impact of FAMEX II was significantly higher for firms in services and for new exporters, the latter being firms which started exporting after receiving

Table 3.5: *OLS estimation of difference-in-differences regressions with interactions: growth over 2004–8 with interactions.*

| | Dependent variable: | | | | | |
| | excluding firms that drop out of FAMEX | | | including firms that drop out of FAMEX | | |
	Change in log exports (1)	Change in log exported products (2)	Change in log export destinations (3)	Change in log exports (4)	Change in log exported products (5)	Change in log export destinations (6)
FAMEX × manufacturing	0.901 (0.542)*	0.146 (0.102)	0.080 (0.071)			
FAMEX × services	5.968 (1.300)***	0.597 (0.159)***	0.943 (0.248)***			
FAMEX × existing exporter				0.530 (0.528)	0.090 (0.110)	0.074 (0.077)
FAMEX × new exporter				3.420 (1.178)***	0.430 (0.170)**	0.363 (0.146)**
Number of observations	319	319	319	319	319	319

Note: robust standard errors in parentheses; *significant at 10%; **significant at 5%; ***significant at 1%. The regressions include all the regressors included in the probit for the propensity to receive FAMEX assistance shown in Table 3.9. The sample includes only firms in the common support. Existing exporter indicates that the firm was exporting prior to 2005.

FAMEX II assistance.[13] These results suggest that the informational costs of reaching new export markets are higher for services firms and for first-time exporters.

6   CONCLUSION

In this chapter, we presented the results from an *ex post* evaluation of Tunisia's export promotion programme for firms. Using detailed firm-level data collected through a purposely designed survey we employed efficient non-parametric matching techniques to compare FAMEX II programme recipients to observably similar non-recipients, and found that the former experienced significantly better export growth after receiving the export assistance.

Our results indicated that the matching grant programme served to increase the value of exports as well as to expand the extensive margin of exports, namely new exported products and new destinations served between 2004 and 2008. Thus, financial assistance can help firms in developing countries develop export markets. Moreover, the results suggest that such grants can help both manufacturing and services exporters and are particularly useful to encourage first-time exporters.

Our study differs from most previous studies in this area in that it measures the impact of a single, well-defined export promotion instrument (a matching grant), as opposed to a mix of instruments. Similarly focused studies can help build information on the relative effectiveness of such instruments, leading to more specific policy recommendations. For example, our results suggest that the FAMEX II grant worked best for firms which were exporting for the first time. It may be that other instruments are more appropriate for experienced exporters who want to enter new markets or start exporting new products.

Our mixed results for firm-level outcomes such as sales and employment prevent us from drawing any concrete conclusions on whether the FAMEX II programme helped Tunisian firms grow overall.[14] In testing for these impacts, we were constrained by the difficulty of collecting retrospective sales and expenditures data through firm surveys. As is the case with many similar programmes, while grant recipients were obligated to provide data for programme assessment, non-recipients had no such obligation and were reluctant to provide some of the key quantitative information. This experience

---

[13]Table 3.10 presents an alternative OLS specification which, in effect, replicates the matching DID approach by using weighted outcomes of untreated firms (weighted by their distance to the treated firm in the propensity score). In that case the average FAMEX impact estimate is identical to the matching DID estimate. The estimated differential impacts by sector and prior exporting are similar to those in Table 3.4, although the magnitudes of the differences are smaller.

[14]Moreover, since we could not capture potential spillover effects, we could not do a social cost–benefit analysis of the programme.

suggests that the assessment of firm-level interventions would be markedly easier if data obligations were extended to all applicants.

Indeed, an important recommendation is that incorporating an evaluation strategy at the start of programmes can have large returns for policy learning, especially given the growing importance of export promotion programmes. For one, data quality is likely to be greatly improved by conducting a baseline survey of target firms. *Ex post* evaluation techniques cannot ensure that estimation bias due to unobservable differences across the comparison groups has been fully removed. An *ex ante* emphasis on rigorous evaluation can significantly enhance the feasibility of experimental or quasi-experimental methods.

Finally, we would like to draw the reader's attention to the question of programme take-up. In Tunisia, less than 10% of eligible firms applied for the matching grant over the five-year period between 2004 and 2008. In fact, this rate was below 3% for services firms. Even among firms already exporting, only 20% applied to FAMEX II. This might suggest either a lack of capacity or a lack of interest for the vast majority of firms in receiving assistance to develop an export business plan. But it could also be that most firms face other types of constraints to exporting. All these possibilities need further consideration by export promotion agencies in their programmes.

Julien Gourdon is a Consultant in the Social & Economic Development Group, Middle East and North Africa Region at the World Bank.

Jean-Michel Marchat is a Senior Private Sector Development Specialist in the Finance and Private Sector Development Unit, Middle East and North Africa Region, at the World Bank.

Siddharth Sharma is a Young Professional in the Economic Policy and Debt Department, Poverty Reduction and Economic Management Network at the World Bank.

Tara Vishwanath is the Lead Economist, in the Poverty Reduction and Economic Management Network, Middle East and North Africa Region at the World Bank.

REFERENCES

Álvarez, R., and G. Crespi (2000). Exporter performance and promotion instruments: Chilean empirical evidence. *Estudios de Economía* **27**, 225–241.

Álvarez, R. (2004). Sources of export success in small- and medium-sized enterprises: the impact of public programs. *International Business Review* **13**, 383–400.

Arnold, J., and B. Javorcik (2009). Gifted kids or pushy parents? Foreign acquisitions and plant productivity in Indonesia. *Journal of International Economics* **79**, 42–53.

Bernard, A., and B. Jensen (2004). Why some firms export? *Review of Economics and Statistics* **86**, 561–569.

Biggs, T. (1999). A microeconometric evaluation of the Mauritius Technology Diffusion Scheme (TDS), RPED Paper 108. World Bank, Washington, DC.

Bruhn, M. (2008). License to sell: the effect of business registration reform on entrepreneurial activity in Mexico. Policy Research Working Paper 4538. World Bank, Washington, DC.

Giné, X., and D. Yang (2009). Insurance, credit, and technology adoption: field experimental evidence from Malawi. *Journal of Development Economics* **89**, 1–11.

Görg, H., M. Henry, and E. Strobl (2008). Grant support and exporting activity. *Review of Economics and Statistics* **90**, 168–174.

Heckman, J., H. Ichimura, and P. Todd (1997). Matching as an econometric evaluation estimator: evidence from evaluating a job training programme. *Review of Economic Studies* **64**, 605–654.

Krugman, P. (1992) *Geography and Trade* (Cambridge, MA: MIT Press).

Lederman, D., M. Olarreaga, and L. Payton (2010). Export promotion agencies revisited. *Journal of Development Economics* **91**, 257–265.

McKenzie, D. (2010). Impact assessments in finance and private-sector development: what have we learned and what should we learn? *World Bank Research Observer* **25**, 209–233.

Mills, G. (2006). Matching grant schemes: what they are, why they exist and how they work, ITC Position Paper.

Phillips, A. (2001). Implementing the market approach to enterprise support: an evaluation of ten matching grant schemes. Policy Research Working Paper 2589. World Bank, Washington, DC.

Roberts, M., and J. Tybout. (1997). An empirical model of sunk costs and the decision to export. *American Economic Review* **87**, 545–564.

Rosenbaum, P., and D. Rubin (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika* **70**, 41–55.

Tang, H. (2009). Evaluating SME support programs in Chile using panel firm data. Policy Research Working Paper 5082. World Bank, Washington, DC.

Volpe Martincus, C., and J. Carballo (2008). Is export promotion effective in developing countries? Firm-level evidence on the intensive and extensive margins of exports. *Journal of International Economics* **76**, 89–106.

World Bank (2009). *Matching Grants: A Review of Matching Grants Systems in Private Sector Development Projects*. Washington, DC: World Bank.

APPENDIX

Table 3.6: *Main characteristics of FAMEX firms and control firms.*

	Sector				
	Agro industry	Textile & apparels	Paper, wood, furniture	Chemicals	Metals
FAMEX firms	18	48	14	18	11
Control firms	23	51	15	26	13

	Sector					
	Machine & equipment	electric	Retail, hotel, transport	IT services	Other services	Total
FAMEX firms	25	4	13	14	30	195
Control firms	26	9	12	10	17	202

	Location				
	Tunis	Grand Tunis	East	Rest of Tunisia	Total
FAMEX firms	52	76	63	4	195
Control firms	26	67	95	14	202

	Employment					
	[1,9]	[10,49]	[50,99]	[100,199]	$\geqslant 200$	Total
FAMEX firms	24	66	34	24	47	195
Control firms	25	82	37	28	30	202

	Sales (in TND)					
	<500m	[500m–1M]	[1M–2M]	[2M–5M]	>10M	Total
FAMEX firms	46	19	18	36	25	195
Control firms	46	32	22	33	9	202

Note: except where noted, the firm characteristics refer to year 2004. The 195 FAMEX firms and 202 control firms are part of the common support identified by propensity score matching. 500m stands for 500,000 TND.

Table 3.7: *Additional characteristics of FAMEX firms and control firms.*

	Age (years)	Share domestic capital (%)	Share foreign capital (%)	Owner is manager (%)	New exporter (%)
FAMEX firms	20.4	91.7	7.4	56.4	11
Control firms	20.3	78.6	19.8	51.8	13.1

	Stop exporting (%)	Non exporter (%)	Export all production (%)	Exporter in 2004 (%)	Exporter in 2008 (%)	Year of first exports
FAMEX firms	4	8	33	83	91	1997
Control firms	7.6	22	30	66	72	1996

Table 3.8: *Unmatched differences in growth over 2004–8 for several outcomes.*

Variable	Treated group	Control group	Difference	*t*-statistic
Change in log exports	1.624	0.944	0.680	1.31
Change in log number of exported products	0.227	0.130	0.097	1.09
Change in log number of export destinations	0.211	0.098	0.113	1.63
Change in log sales	0.879	0.829	0.050	0.14
Change in log number of employees	0.271	0.223	0.048	0.51

Table 3.9: *Probit regression for the propensity to receive FAMEX assistance: survey data.*

	Dependent variable: FAMEX status
Age	1.141
	(0.512)**
Age squared	−0.201
	(0.096)**
Grand Tunis	−0.314
	(0.234)
Rest of Tunisia	−1.398
	(0.368)***
East of Tunisia	−0.818
	(0.229)***
Textiles, apparels and shoes	0.626
	(0.325)*
Paper, wood and furniture	0.392
	(0.276)
Chemicals	0.549
	(0.347)
Metals	0.135
	(0.314)
Machine and equipment	0.596
	(0.365)
Electrical	0.241
	(0.289)
Retail and transport services	−0.264
	(0.526)
IT services	0.370
	(0.369)
Other services	0.116
	(0.389)

Note: robust *t*-statistics in parentheses; *significant at 10%; **significant at 5%; ***significant at 1%. Age, years since the owner is at the firm; number of employees and sales refer to 2004. The omitted sector is agro-industry, the omitted location is Tunis and the omitted categories in terms of employment and sales are the smaller size categories.

Table 3.9: *Continued.*

	Dependent variable: FAMEX status
Share of domestic capital	0.346
	(0.074)***
Years since manager is at the firm	−0.269
	(0.107)***
10–49 employees	0.104
	(0.249)
50–99 employees	0.102
	(0.295)
100–199 employees	−0.167
	(0.318)
More than 200 employees	0.757
	(0.290)***
Sales of 500,000 to 1 million	−0.478
	(0.262)*
Sales of 1 million to 2 million	−0.244
	(0.285)
Sales of 2 million to 5 million	−0.047
	(0.257)
Sales of 5 million to 10 million	0.329
	(0.327)
Sales of more than 10 million	−0.297
	(0.216)
Non-exporter in 2004	−0.903
	(0.182)***
Number of observations	378

Note: robust *t*-statistics in parentheses; *significant at 10%; **significant at 5%; ***significant at 1%. Age, years since the owner is at the firm; number of employees and sales refer to 2004. The omitted sector is agro-industry, the omitted location is Tunis and the omitted categories in terms of employment and sales are the smaller size categories.

Table 3.10: *Weighted OLS estimation of difference-in-differences regressions with interactions using generated control firms.*

	Change in log exports (1)	Change in log exported products (2)	Change in log export destinations (3)
FAMEX	1.313***	0.169**	0.155***
	(0.405)	(0.071)	(0.057)
FAMEX × manufacturing	1.222***	0.196**	0.119**
	(0.432)	(0.080)	(0.054)
Number of observations	300	300	300

	Change in log exports (4)	Change in log exported products (5)	Change in log export destinations (6)
FAMEX × services	3.000***	0.330**	0.654***
	(1.070)	(0.132)	(0.222)
Number of observations	300	300	300

	Change in log exports (7)	Change in log exported products (8)	Change in log export destinations (9)
FAMEX × existing exporter	0.219	0.110	0.081
	(0.313)	(0.079)	(0.059)
FAMEX × new exporter	5.448***	0.548***	0.526***
	(1.016)	(0.135)	(0.123)
Number of observations	300	300	300

Note: robust standard errors in parentheses. *significant at 10%; **significant at 5%; ***significant at 1%. OLS estimation is weighted in that outcomes of untreated firms are weighted by their distance to the treated firm in the propensity score. The regressions include all the regressors included in the probit for the propensity to receive FAMEX assistance shown in Table 3.9. The sample includes only firms in the common support and excludes FAMEX drop-outs.

4

The Use of Experimental Designs in the Evaluation of Trade-Facilitation Programmes: An Example from Egypt

DAVID ATKIN AND AMIT KHANDELWAL

1 INTRODUCTION

Developing countries are becoming increasingly integrated into the world economy thanks to large declines in transportation costs and a reduction in trade barriers. Yet substantial constraints remain that prevent many firms in developing countries from successfully expanding and diversifying their exports. Such constraints include trade financing shortfalls, inadequate product quality, a lack of relevant certifications required for exporting, insufficient understanding of the preferences of foreign consumers and difficulties matching with appropriate foreign buyers. Overcoming these obstacles is the primary goal of many of the current 'aid-for-trade' initiatives.

Unfortunately, we know very little about the relative importance of, and interaction between, the various constraints holding back developing country enterprises. There has also been no rigorous evaluation of how effective existing aid-for-trade initiatives are at facilitating trade. This chapter lays out the various methodologies available for performing such an evaluation and discusses in detail one such evaluation focusing on handloom producers in Egypt.

1.1 Evaluating Aid for Trade

The aid-for-trade agenda recognises that, in addition to large-scale policy reforms, complementary policies are required to facilitate the participation of developing country firms in the global economy. The aid-for-trade agenda focuses on four key areas.

- *Trade policy and regulation:* this entails building capacity for developing countries to participate in formulating, negotiating, and implementing regional and multilateral trade and investment agreements.

- *Infrastructure:* this area recognises that improvements in infrastructure, such as road construction, and the reduction in costs imposed by road checkpoints, power generation and port development are necessary to reduce the costs of trade.
- *Productive capacity building:* this area aims to increase firms' capacity to export by bridging gaps in firms' technical knowledge about foreign transactions, which range from certifications for export markets and better quality control to leveraging global supply chains.
- *Adjustment assistance:* many countries rely on tariff revenues as an important source of government revenue. This fact makes it more difficult for countries to lower trade barriers. The aid-for-trade agenda aims to help countries adjust to liberalisations by improving customs and tax collections and helping with additional transaction costs.

A common theme in these initiatives is the focus on domestic constraints as barriers to international trade. The hope of aid-for-trade initiatives is that, by addressing these four areas, developing countries will be able to take advantage of the global reduction in external trade barriers and create successful exporting businesses. These successful exporters will boost employment, generate foreign exchange reserves and raise domestic growth rates.

But to our knowledge, no rigorous evaluation has yet been conducted on any of these initiatives. As a result, we are unsure how to design such programmes to guarantee the highest chance of success. For example, it may be critical for export facilitation initiatives to focus on increasing firms' understanding of consumer preferences and the market structure in the foreign market they are trying to penetrate, rather than only focusing on reducing blockages at ports. Additionally, we know very little about the cost–benefit analysis of such programmes and the potential spillovers to firms and industries not targeted in the initiative.

1.2   The Randomised Control Trial as a Tool for Programme Evaluation

One methodology, the randomised controlled trial (RCT), is well suited to addressing questions of this type. To understand the advantages of an RCT over *ex post* econometric analysis, consider how we might evaluate one particular programme that would fall under the aid-for-trade initiative: helping firms to obtain proper certifications to export their products. Organisations like the World Bank or international trade consulting agencies often provide this type of support to small and medium enterprises (SMEs) in developing countries.

These programmes generally work as follows.

- A consultant who understands a particular industry is assigned to help SMEs obtain the proper certification to export to foreign markets. Examples of certification might include regulatory forms required for exporting food products to the European Union.

- The consultant will identify a set of SMEs that are best positioned to begin exporting.
- The success of the programme is judged on the success of these SMEs in successfully exporting to foreign markets.

However, such a positive evaluation of the programme may be spurious. The SMEs that the consultant has chosen to work with were precisely the firms most likely to succeed as exporters. These firms will have been chosen for a multitude of reasons, such as:

- they manufacture higher quality products than other firms in the domestic market;
- they have demonstrated success in the domestic market;
- they presently export to foreign markets with similar characteristics to their home market (such as quality or design preferences);
- they have skilled and/or motivated managers and owners;
- they are located in or close to the city where the programme is based.

Even in the absence of an export certification programme, exports were likely to have grown over time for SMEs with these characteristics. The consultant chose these firms precisely for this reason: they were the best firms with the highest growth prospects. Since exports following the programme are larger than exports prior to the programme, the consultant may want to conclude that the programme was successful. However, without comparison with an adequate control group that did not receive assistance, it is impossible to know the effectiveness and impact of such a programme.

Three non-experimental options are available to evaluate the efficacy of the programme. All options involve collecting data on the SMEs' operations several periods before and after the programme implementation.

1. An evaluation could compare outcomes for similar firms based on observable characteristics, some of whom received the certification subsidy and some of whom did not. This approach is known as a matching estimator. In practice it is very difficult to locate these similar firms. In the extreme, if the consultant selects all the most promising firms to receive the certification programmes, then no similar firms would exist. Even if only a subset of promising firms are selected, the most promising firms may differ along a number of dimensions, many of which may be unobservable or immeasurable.

2. An evaluation could study the exports of many firms over time and hope to see a jump in exports for the firms receiving assistance at precisely the time that the certification programme went into operation. This approach is known as a regression discontinuity. In practice, even if a sharp jump in exports is observed, it is difficult to rule out other contemporaneous factors that may have boosted exports. For

example, other components of the facilitation programme are usually implemented simultaneously, or trade barriers may have been reduced around the time of the implementation.

3. An evaluation could combine the methodology of the discontinuity and matching approaches and look for differential changes in outcomes over time between certified and non-certified firms. With sufficient time-series data, it may be possible to show that exports grew at the same rate in both samples prior to the certification intervention and then diverged after it. This approach, known as a difference-in-difference estimator, requires many periods of data prior to the intervention in order to adequately control for trends in export growth between the two groups prior to the intervention. Additionally, several periods of data after the intervention are required in order to accurately measure a differential change in the growth of exports between the certified and uncertified firms.

The flaws and data needs associated with the three approaches above have meant that an increasing number of development practitioners are turning to RCTs in order to evaluate programmes that are amenable to such an approach. The methodology generates a set of ideal comparison firms, and is derived from the large literature in medical sciences that uses RCTs to evaluate the effectiveness of various drugs and medical procedures.

In the certification example, the firms for whom certification is expected to be most useful are placed in a sample. Half of these viable exporters are then randomly chosen through a lottery process and placed in the treatment sample. These treatment firms initially receive the certification programme. The other half of the firms is placed in the control sample.[1]

The key to the success of an RCT is that the control and treatment groups are likely to be identical along all characteristics prior to the intervention. Since the two groups were selected at random, a firm in the control group is just as likely to have a great manager or a low-quality design as a firm in the treatment group. Therefore, any average differences observed between the two groups after the certification programme can be attributed to the certification process itself.

1.3   An RCT Design to Evaluate a Trade-Facilitation Programme in Egypt

In this chapter, we discuss an ongoing project that uses an RCT methodology to study the impacts of a programme designed to link Egyptian microenterprises with global markets. The issues that arise in evaluating this interven-

---

[1]In the interest of fairness, the control group can be promised the certification programme a short time later. Such a strategy is particularly useful where there are budgetary or staffing issues that prevent all viable firms from simultaneously being offered the programme anyway. We discuss this issue further in Section 3.

tion are similar to the issues that would arise in many of the aid-for-trade initiatives listed earlier.

The motivations for such a study are twofold. First, the study can be viewed as an impact evaluation of a programme that matches microenterprises to US buyers, improves design quality and offers business training. The second motivation is to use an RCT in order to shed light on three central issues in international trade:

- does trade increase productivity?
- why do some firms become successful exporters, while others do not?
- how large are the spillovers from exporting?

The last two of these three issues are of special interest to policymakers. Understanding which firms are likely to become successful exporters allows trade-facilitation programmes to be better targeted. Meanwhile, if the programme successfully causes firms to start exporting and this phenomenon leads to substantial spillovers in the local economy, this suggests that there may be larger impacts per dollar spent than simply the direct effects observed among the treated firms.[2]

In Section 2 we focus on the 'ideal' experiment that can answer these or similar questions. The purpose is to think of the 'best-case' scenario where we have access to adequate resources, the randomisation is acceptable to the programme stakeholders and the local context poses no further difficulties. Section 3 will then discuss our attempt to implement this best-case scenario for a trade-facilitation experiment in Egypt. A number of problems have arisen while implementing the experiment, and we will describe how we have worked around these problems. We will also discuss the common questions and concerns brought up in our discussions with potential project partners and how we have tried to address these.

2 DESIGNING AN RCT THAT EVALUATES THE LINKING OF ENTERPRISES TO FOREIGN MARKETS

In this section, we discuss the key factors in designing an RCT to evaluate the efficacy of trade-facilitation interventions that aim to connect developing country enterprises with world markets. This intervention has the following three elements.

---

[2]Moreover, combined with identifying characteristics associated with firm success, assessing the magnitude and extent of spillovers will provide knowledge about how to target industrial development programmes spatially. We note that we may not observe spillovers if the programme offers similar information (market design help, etc) that the firms in an area have already received. Therefore, we expect to observe larger spillovers in instances where the programme successfully removes an exporting constraint.

1. *Provision of assistance to match local firms with foreign buyers.* This process involves putting buyers and sellers in contact with one another. Matches could be facilitated through trade show appearances, face-to-face meetings and other networking events, or by directly marketing the enterprises to foreign buyers.

2. *Provision of assistance to create designs that are appealing to foreign consumers.* This process involves contracting skilled local or foreign designers to work with the local firms in designing new products.

3. *Improvements to general business skills through local training programmes.*

In the experiment, firms are approached by our project partner to facilitate the sale of their products to retail clients in the US market. The basic experimental design is quite simple. Prior to the intervention, we draw up a list of viable firms in the region and industry and conduct a baseline survey of these firms. This set of firms is called the *sample*, and the total number of firms is the *sample size*. Every firm in this sample should be a firm that the project is willing to target, and so they are perceived as viable exporters if given the types of assistance listed above.

A random set of these sample firms are then approached by the project partner and provided with some combination of the three services listed above. This set of chosen firms is known as the *treatment* group. The firms from the sample who are not selected are labelled the *control* group. We then track the progress of all the sample firms over a one- or two-year period. Once the evaluation period is complete, the control firms can be provided with the assistance services if desired.

If the randomisation has been done correctly, evaluation is extremely simple. On average, the only difference between firms in the two groups is that the treatment group received trade assistance; therefore, any difference between firms in the treatment and control groups can be attributed to the trade assistance itself. In practice, this comparison is done by comparing means of various firm outcomes between the two groups. Depending on the particular intervention, the outcomes of interest could range from volume of exports, changes in the skill composition of firms' workforce, product quality, investments in new assets and the number of new clients.

Policymakers may often want to assess the relative efficacy of a set of feasible interventions and to determine whether certain interventions require other accompanying interventions in order to be successful. For example, it is possible that service 3 above—providing business training—has very low returns. Meanwhile, design help may be useless unless accompanied by assistance in finding new buyers. In order to answer these questions, different subsets of treatment firms can be randomly allocated different combinations of assistance services. If the sample size is large enough, treatment firms can

be allocated into seven groups, covering every possible combination of at least one service (1; 2; 3; 1 and 2; 1 and 3; 2 and 3; 1, 2 and 3).[3]

As before, by comparing means of various firm outcomes between groups, the evaluation can determine the relative effect of additional services as well as the impact of each service alone. This information can be combined with accounting estimates of the cost of delivering each service in order to carry out a cost–benefit analysis.

In order to improve the targeting of trade-facilitation programmes, it is important to know which types of firm benefit the most from such interventions, and which gain relatively little. A well-designed RCT can also reveal this information. If detailed data are collected in the baseline survey on the characteristics of each firm (owner's background, age of firm, number of employees, etc), then it is possible to compare the mean impact for the treatment and the control group separately for firms with each of these characteristics. For example, it may be the case that treatment and control firms are identical if the owner has no formal education, but the treatment firms grow much larger when provided with export assistance if the owner is educated.[4] Such an analysis will provide a list of the characteristics of the firms that are in a position to gain the most from export facilitation programmes.

Finally, in order to look for spillovers from firms that are offered export facilitation assistance to other firms, firms outside the sample that have close links with firms inside the sample can also be interviewed (*eg* they share suppliers, the two CEOs speak to each other regularly, they are located next to each other, etc). Spillovers are identified by comparing the responses of two types of firm outside the sample; those firms connected to treatment firms and those firms connected to control firms. Note that testing for spillovers requires additional firm surveys beyond the firms assigned to the treatment and control groups.

One of the most important decisions in designing the RCT is the size of the sample. In general, the larger the sample is, the more likely the RCT is to be informative. If the programme is successful, the RCT will only provide supportive evidence of that success if the standard errors are small enough that the experiment reveals statistically significant impacts. However, in practice large samples are expensive to survey. In the case of export facilitation programmes, the binding factor will often not be cost, but a shortage of viable firms to receive the export facilitation services in the first place. If there are only 50 firms in a particular country who are eligible for facilitation services, it

---

[3]An in-depth discussion of how to arrive at the appropriate sample size is beyond the scope of this chapter. We refer the readers to Duflo *et al* (2006), who discuss in greater detail the choice of appropriate sample sizes in experimental samples.

[4]In practice, rather than comparing means, a regression analysis is performed with a dummy treatment variable ('1' if offered assistance, '0' if not) interacted with various firm characteristics.

is unlikely an RCT will uncover positive treatment effects even for a successful programme.

Unfortunately, deciding on the minimum sample size is more of an art than a science. The JPAL website[5] contains various documents and computer programs to assist in these decisions, but all methods at some level rely on estimate of variances that are unknown unless there is substantial existing survey evidence on similar firms. Given the scarcity of trade-facilitation experiments, it may be difficult to obtain estimates of these variances, although non-experimental studies could serve as a rough guide.

In addition to designing the randomisation, the second critical feature of an RCT is a firm-level survey. Without well-measured firm-level outcomes, it is impossible to identify the changes in these outcomes caused by the trade-facilitation programme. While researchers in labour, health and development economics now have extensive experience in designing household surveys, the profession's experience in collecting firm-level information is relatively thin. Therefore, there are few existing surveys that can be used as templates. Given the paucity of experiments within the trade literature, a starting place for the survey design, questionnaire and implementation is the small literature that documents field experiments relating to small- and medium-scale enterprises. Good examples are Bloom *et al* (2011), del Mel *et al* (2008, 2009), Karlan and Valdivia (2011) and Kremer *et al* (2010). To our knowledge, these references are the best sources detailing efforts to collect sensitive information such as firm-level profits, inventories and assets.

All surveys should be piloted prior to the intervention in order to assess the time burden on responders, the specificity of key questions, potential recall errors and the proper units of measurement. For instance, in the handloom weaving sector in Akhmeem, Egypt, we initially piloted questionnaires that asked weavers about input and output usage in the past 30 days. We quickly learned that the weavers' production cycle is linked to raw material that was re-supplied every 45–60 days, and changed the recall period accordingly.

Some firm-level outcomes, for example, production techniques or investment, may change very rapidly. If understanding the precise timing of these decisions is important, monthly or quarterly surveys can be administered. Such frequent surveys also reduce problems with recall error. The trade-off, of course, is the additional costs of carrying out more surveys. The cost of these periodic surveys can be reduced if the information can be collected by mobile phone.[6]

---

[5]See http://www.jpal.org.

[6]In Table 4.1, we provide a rough budget that we proposed at the onset of the export facilitation programme in Egypt. The budget includes an approximate cost of hiring a part-time local research coordinator, a monitoring and evaluation team, data-entry costs (which can be offshored to firms in low-wage countries like India or Cambodia) and surveying equipment (*eg* GPS devices). Note that the budget can be reduced if the programme were to be implemented in a lower-waged country than Egypt.

Table 4.1: *Sample budget (in US dollars).*

	Pilot	Baseline	Follow-up
In-country project coordinator	1,200	4,800	4,800
Survey team	750	25,000	25,000
Gifts for survey team, SSEs in sample, exporters	—	2,000	2,000
Staff training	2,000	—	—
Travel and lodging for PIs	3,000	3,000	3,000
Miscellaneous expenditures	1,000	1,000	1,000
Data entry	—	2,500	2,500
Office space	250	250	250
Subtotal	8,200	38,550	38,550
Grand total	85,300		

3   A SPECIFIC EXAMPLE: HELPING EGYPTIAN HANDLOOM PRODUCERS ACCESS EXPORT MARKETS

The previous section describes the design for a general trade-facilitation experiment. We will now describe an evaluation in the handloom weaving sector in Egypt. This ongoing project closely follows the aforementioned RCT methodology. The purpose of the experiment is to evaluate firm-level responses to export market access. With the help of our project partner, the programme will establish links between weaving firms in Akhmeem, Upper Egypt, with retail clients in the USA. Our partner has considerable experience in providing export market access for craft industries in developing countries with overseas retailers who market these niche products as luxury items. Some of the handwoven products these firms manufacture include bedspreads, shawls, tablecloths and pillowcases. The experiment is simply to split the sample firms into a treatment and a control group, approach only the treatment firms with this opportunity to export and compare firm-level outcomes between the two groups.

The project partner provides various services to these craft producers. The first service is to work with design consultants to develop patterns that appeal to the preferences of US consumers. The second service is to market extensively these potential products to retailers, and the third major service is to provide general business training to the firms. The partner will work with these textile firms through an Egypt-based intermediary who has established connections with these weaving firms. The intermediary has agreed to initially source at least once from every firm in the treatment group. However, after this initial order, the intermediary is free to source only from the subset of treatment firms that were able to satisfactorily fulfil the initial order.

Soon after an order is placed by the retailer, we will administer our surveys to both the treatment and control groups. As mentioned in the previous section, well-designed surveys are essential in order to identify the precise

impacts of being provided with a match to a foreign buyer. The survey will ask questions ranging from inventory data, quality issues, employment changes but also includes impacts at the household level such as schooling decisions and expenditure information. Naturally, the researcher(s) must decide which responses they are most interested in, as the length of the survey increases response burden. Follow up surveys can be shorter and simply collect information on how key production variables such as inventory and sales have changed over time. Finally, we will re-survey the firms after approximately one year.

This experiment is a straightforward impact evaluation of a programme that matches Egyptian firms to US buyers, improves design quality and offers business training. More importantly, the RCT can shed light on three central issues in international trade listed in Section 1.3.

We chose the handloom sector in Egypt for three reasons.

(a) The market structure in the industry is such that there are many small firms producing viable export products; therefore, we are able to obtain a large sample of firms.

(b) The handloom technology is simple to understand and similar technologies are used by all firms; therefore, detailed surveys that capture all dimensions of production are feasible to write and later to administer. The more precisely we can measure firm outcomes, the more likely it is that the true impacts of the programme can be determined. Examples of such industries include garments, textiles, lightly processed food products and agriculture.

(c) Our project partner (PP) was implementing a new programme in Egypt that intended to incorporate the handloom weaving sector. By embarking on the evaluation at the programme planning stage, the sample can be randomised prior to the programme roll out.

In the process of finding a PP, and designing and implementing the experiment, we have encountered a number of problems. Here, we list and discuss the issues that have arisen and how we have attempted to resolve them.

3.1 Finding a Project Partner

Carrying out an RCT within the context of international trade depends crucially on finding a suitable PP. Unlike other contexts where the researcher(s) can carry out the projects themselves (such as providing cash grants), an experiment like the one described above relies critically on the PP providing trade-facilitation services. The PP must therefore agree to and comply with the study. With enough resources, it would be possible to link the weavers in Akhmeem ourselves by simply purchasing their products outright (and therefore retaining full control over the experiment). However, in this scenario, such conduct may be unethical if firms are expected to make investments

based on the expectation of future export orders that we know are not forth-coming. Additionally, conducting a survey within the set of standard market transactions increases the external validity of the experiment. In our context, the PP primarily serves the role of facilitator/matchmaker between retailers in the USA, intermediaries in the local market and the producers.

3.2 Convincing the PP that an RCT Is Both Desirable and Feasible

The most difficult aspect of conducting an RCT such as the one described above is convincing the PP that an RCT is both in their interest and feasible. After explaining the basics of an RCT (perhaps using the certification example in Section 1 as an illustration) and its advantages over their existing evaluation methods, PPs typically see the benefits of an RCT. However, our experience has been that, typically, the PP's immediate reaction is that it is impossible to carry out. They cite a number of reasons.

The RCT May Interfere with the PP's Targets and/or Funding Requirements

The PP is subject to meeting targets and deadlines set in the programme fund-ing guidelines. These targets often include explicit sales targets, which makes the PP uneasy about sourcing from an 'experimental group' that they believe may not be able to manufacture products up to the required standard. The RCT therefore threatens the commercial viability of their programme.

This concern is a valid one and should be carefully accounted for in the RCT set-up. One strategy that we proposed is to allow the PP to select who they think are the best set of initial firms that they can use to generate their initial targets. These firms must be excluded from the survey sample. The sample consists of a second tier of firms who are viable, but of higher risk than the first-tier firms. These initial firms can be used to pilot the surveys prior to the experiment, and so this set-up does provide some benefits. The following additional strategies could be tried to reduce the burden on the PP.

(i) If the researchers have enough resources, they may offset the additional costs of the experiment by paying for the PP's time devoted to designing the experiment and funding the evaluation team.

(ii) It is best to find a new programme that the PP is about to commence, in which they are able to incorporate specifically the RCT. Ideally, the RCT component should be included at the funding stage to ensure maxi-mum buy-in. This ensures that the researchers are involved at the earlier stages of the programme when decisions about the programme timeline are made.

(iii) It is important to stress that the RCT helps the PP in the long run because the evidence from the study assists the PP in deciding how to best tar-get their limited resources across multiple services and across different types of firm. Similarly, an RCT can establish the cost effectiveness of

the programme. Such information is becoming increasingly important, as donor agencies are requiring more rigorous evaluation methods for projects. Hence, positive evidence from the RCT can improve the chance of the PP receiving future project funding.

(iv) A successful RCT may not require that all firms in the treatment sample actually receive the full service under evaluation, only that treatment firms have a substantially higher chance of benefiting from the treatment. For example, if a firm is unable to meet production targets, the PP would be free to reallocate its orders towards another firm (within the treatment group).

The Control Group of Firms Is Not Fair or Ethical

There are different strategies to deal with this issue.[7] First, PPs usually do not 'ramp up' their projects immediately. A set of firms will therefore have to be excluded from the initial programme implementation and brought in at a later date. The RCT simply asks that, as the PP expands the number of firms that the programme reaches, this expansion be done in a randomised way. In this way all the firms eventually receive treatment, but the intervention is randomly staggered over time. In the eyes of the experiment's subjects, the random assignment could be viewed as unfair, thus threatening the presence of the PP in the locality. In practice, the randomisation could be done in a way that avoids this 'loss aversion' by announcing that this programme will be initially assigned based on the outcomes of a standard lottery. The treatment firms were simply lucky in having 'won' the right to receive subsidised export services immediately. This allocation method is often deemed fair by the local firms.

What Happens if other Firms Want to Participate? It Seems Unethical to Deny Them Access

If the firms that approach the PP are outside the control group, this is actually a positive outcome because it implies that information has spilled over to the firms outside of the sample. It is therefore appropriate, depending on the question, that these firms be allowed to participate, and importantly, the researchers collects information on these firms. If firms within the control group wish to participate, this represents a double-edged sword. On the one hand, the fact that firms ask to participate provides evidence that spillovers are occurring and can teach us about how knowledge diffuses among firms, and so this information should be recorded. However, if all control firms ask for and are allowed to receive the treatment, then the randomisation will fail

---

[7]We had one experience with an NGO that flat out refused our proposal because of the need for a control group. There is not much that can be done in such a case.

entirely, as both the treatment and control firms receive the same services. In general, contamination of the control group should be strongly discouraged.[8]

A Sample Frame Does Not Exist

In order to have secured funding, the PP is likely to have some estimate of the size of the industry under study. Depending on the hypothesis being tested, it is critical to know if these numbers reflect the number of workers or the number of managers. Certain industries may have a large number of small firms, or a small number of large firms. If the outcomes of interest depend on the actions of managers rather than workers, then that is the appropriate level for the randomisation (*eg* evaluating capital investment decisions as opposed to the effectiveness of employee contract types). It is important to know the industry structure prior to designing the experiment.

A typical programme such as the one outlined above may target initially one or two pre-selected firms. Then, as the programme expands, additional firms are targeted using the firm contacts provided by the initial firms. As a result, the PP may not have an initial sample frame. In practice, local business associations, local government offices and intermediaries operating in the area should be consulted in order to construct the most accurate sample frame possible, a crucial component in effectively implementing an RCT.

3.3   Ensuring Compliance with the RCT Once the Experiment Has Begun

Once the RCT begins, it is important to ensure that the various agents are complying with the experimental design. For instance, our design requires that intermediaries purchase from the randomly selected producers and sell to the US retailers. However, these intermediaries may not comply with the RCT and instead allocate orders to some favoured members of the control group.

Possible strategies to ensure compliance are given below.

(i) Intermediaries or other agents motivated by financial profit can be incentivised to comply with the experiment. These agents are in line to earn substantial profits if the firms they deal with generate successful exporting relationships. Incentive-compatible contracts should be designed to ensure compliance in exchange for the opportunity to expand their business.

---

[8]If any of the control group firms are provided with the treatment, an evaluation will be less likely to determine that the project was successful, as the contamination of the treatment and control groups will attenuate all the results towards finding no effect.

(ii) Effective monitoring guarantees that any non-compliance issues will be quickly spotted and can be dealt with appropriately. Close monitoring may also discourage some agents from violating the research design if they know they will be discovered. Researchers should collect reliable information about the operations of the producers. If export products are visibly different, one option is to do spot checks to ensure that control group firms are not actually producing export products. Another option is to compare the reported sales by the producers with the reported purchases by the retail client and check that these numbers match up.

(iii) Another potential source of non-compliance is resistance from the project coordinators, who are not used to having randomised components in their projects. It is important that the project coordinators see and understand the benefits of an effective RCT from the outset. Typically, the project under evaluation will benefit less from the evaluation than future projects of a similar nature, and so working with a coordinator who has vested interests beyond the success of the project under evaluation is important. This point is closely related to (ii) in Section 3.2.

4   CONCLUSION

The suggestions offered in this chapter provide a general framework for thinking about possible evaluations of programmes that fall under the aid-for-trade initiative. This framework will need to be refined for each specific programme, and these refinements will occur after extensive discussions between all agents involved in the evaluations: the researchers, the provider of the trade-facilitation services, the firms and any relevant governmental or nongovernmental organisations.

In all cases, it is essential that the researchers devising an evaluation experiment are familiar with the specifics of the programme that will be evaluated, as well as with the local context in which the programme will be implemented. The researchers should visit the potential target firms and meet with the local project coordinator. This knowledge and these interactions will allow researchers to address any constraints that the experiment must work around, as well as focus the experimental design such that it can answer the precise questions of practitioners and stakeholders.

Amit Khandelwal is the Assistant Professor of Finance and Economics at Columbia University.

David Atkin is Assistant Professor of Economics at Yale University.

REFERENCES

Bloom, N., B. Eifert, A. Mahajan, D. McKenzie, and J. Roberts (2011). Does management matter: evidence from India. NBER Working Paper 16658.

del Mel, S., D. McKenzie, and C. Woodruff (2008). Returns to capital in microenterprises: evidence from a field experiment. *Quarterly Journal of Economics* **123**, 1329-1372.

del Mel, S., D. McKenzie, and C. Woodruff (2009). Measuring microenterprises profits: must we ask how the sausage is made? *Journal of Development* **88**, 19–31.

Duflo, E., R. Glennerster, and M. Kremer (2006). Using randomization in development economics research: a toolkit. Mimeo, MIT, Cambridge, MA.

Karlan, D., and M. Valdivia (2011). Teaching entrepreneurship: impact of business training on microfinance clients and institutions. *Review of Economics and Statistics* **93**, 510–527.

Kremer, M., J. Lee, J. Robinson, and O. Rostapshova (2010). The return to capital for small retailers in Kenya: evidence from inventories. Mimeo, Harvard University.

Transport Costs and Firm Behaviour: Evidence from Mozambique and South Africa

SANDRA SEQUEIRA

1   THE IMPORTANCE OF TRANSPORT COSTS

An extensive trade literature has argued that trade costs are a significant impediment to trade and economic growth (Frankel and Romer 1999; Hummels 1999, 2008; Rodriguez and Rodrik 2000; Obstfeld and Rogoff 2000; Anderson and Wincoop 2004). Although there are many drivers of trade costs, the role played by direct and indirect costs associated with the transportation of goods has received renewed attention in recent years. Following the progressive dismantlement of tariffs in the 1990s and early 2000s, it is increasingly argued that transport costs have come to replace tariffs as the main barrier to trade in parts of the developing world. The most compelling evidence comes from sub-Saharan Africa (SSA) (Hummels 1999, 2007; Hoekman and Nicita 2008; Freund and Rocha 2010).[1]

As a result, a significant portion of aid efforts has in recent years been devoted to reducing trade costs and improving trade logistics.[2] The majority of these investments have targeted both physical transport infrastructure through the rebuilding of railways, roads and ports and the upgrading of

---

[1] In 2007, shipping a container from a firm located in the main city of the average country in sub-Saharan Africa was twice as expensive as shipping it from the US, Brazil or India (World Bank 2007). Shipping from SSA is also more time consuming. In 2007, it took an average of 35 days for a firm to get a standard 20-foot container from its warehouse through the closest port and on a ship. This was twice as long as in Brazil and six times longer than in the USA. Djankov *et al* (2010) suggest that each day the movement of cargo is delayed reduces a country's trade by 1% and distorts the ratio of trade in time-sensitive to time-insensitive goods by 6%. Similarly, Limão and Venables (2001) estimate that a 10% decline in transport costs for a cross-section of countries worldwide would increase trade by 25%.

[2] In 2008, approximately 20% of the World Bank's budget was spent on transport infrastructure projects in over 35 countries worldwide.

transport bureaucracies. These projects appeal to donors, since they translate into divisible and verifiable investments.

However, these projects are seldom guided by rigorous empirical evidence on how the associated investments can affect the real drivers of transport costs. In particular, the micro-level mechanisms through which transport can influence economic activity by affecting firms and households remain largely unexplored. Governments in the developing world often struggle with how to prioritise investments in different types of transport infrastructure, and with how to ensure that these investments are financially sustainable. Both tasks would be greatly assisted by a clear understanding of the impact of investments in different types of transport infrastructure on different economic agents, on different regions and, more broadly, on long-term development and growth.

The literature on the drivers of transport costs has developed around two main approaches. One strand in the literature focuses on the relative importance of hard transport infrastructure, such as the quality of roads, railways and ports, while a second strand focuses on soft transport infrastructure, such as the regulation of transport markets, port policies and rules regulating the movement of goods across borders. Limão and Venables (2001) argue that the quality of hard transport infrastructure accounts for most of SSA's poor trade performance, as measured by variables such as the density of a country's road and rail networks. The authors suggest that improving transport infrastructure from the bottom quarter of countries—primarily from SSA—to that of the median country in their sample could increase trade by 50%. Buys *et al* (2006) resort to spatial network analysis to simulate the effects of road upgrading and suggest that connecting all of SSA's capitals to population centres with more than 500,000 inhabitants would translate into a US$250 billion increase in trade volumes over 15 years.

However, decades of large investments in road infrastructure and port upgrading throughout Africa have yet to deliver a proportional decline in the cost of transport or the expected increase in trade volumes associated with a reduction in trade costs. As an example, even in a middle-income country like South Africa, which boasts the best road network in the region and some of the best equipped ports, expenditures on commercial transport and logistics are still equivalent to about 15–20% of GDP, which is double the figure for comparable countries like Brazil and India (CSIR 2005, 2006).

Motivated by this evidence, Raballand and Macchi (2008) argue instead that transport prices result less from poor transport infrastructure than from inefficiencies in the structure of transport markets. The authors use original trucking surveys to measure the costs, rates and performance of the trucking industry across Western and Central Africa. They conclude that transport prices charged to business are high, while the actual costs incurred by trucking companies in Western Africa to move cargo do not differ greatly from those

faced by a trucking company in the developed world.[3] Macchi and Sequeira (2009) provide a point of convergence in these two literatures. They argue that the poor functioning of the soft infrastructure of ports due to corruption directly affects demand for port services, which in turn has an impact on the returns to the investments made in the hard infrastructure of the port.

While these studies suggest important correlations, a clear identification of the determinants of transport costs and of how they affect firm behaviour remains elusive. Undertaking any type of evidence-based policymaking under these conditions becomes extremely difficult.

The recent surge of impact evaluations in social programmes is beginning to make inroads into the trade and industrial organisation literatures, with enormous potential to widen our understanding of how firms behave in response to policies aimed at reducing trade costs. The adoption of quasi-experimental and experimental methods for causal inference can set us on a promising path towards understanding these foundational mechanisms. While this optimism should surely extend to the study of transport policies, the nature of transport projects—in hard or soft form—poses an additional set of challenges that researchers will have to consider.

The first stumbling block is the absence of micro-data on transport costs and prices, before and after investments in transport infrastructure take place. With the exception of the trucking data generated in Raballand and Macchi (2008) and Macchi and Sequeira (2009), none of the major development microdata sets—namely the Living Standards Measurement Survey, the Doing Business and the World Bank's Enterprise Surveys—currently collect adequate data on transport costs or transport prices at the firm and household level. Data on railway and port tariffs is also proprietary and difficult to obtain, even for the main regional transport corridors in sub-Saharan Africa.

Lack of data creates enormous difficulties not only for researchers' understanding of the distribution of direct transport costs and how they affect firms and households, but also in identifying the indirect economy-wide effects, such as the effect of transport on the geography of business and on patterns of trade and specialisation. It also prevents a clear understanding of how the structure of transport markets may mediate the cost of transport between providers and end-users.

For example, the absence of micro-data prevents measurement of the extent to which cost savings induced by investments in the quality of roads are passed on to users of transport services. It also blocks an assessment of how benefits from investments in railways may have a differential effect on firms,

---

[3]This is primarily because the trucking industry in developing countries faces considerably lower fixed costs than their counterparts in the developed world, which results from a combination of low wages, low capital costs due to low barriers to entry of second-hand vehicles, and high vehicle usage. This is compensated by higher variable costs stemming from high fuel consumption due to age and vehicle fleet condition, as well as higher maintenance costs.

depending on the fare structures adopted and the allocation of freight to available slots in the railway network. Transport is a basic input for most economic activity, so the range of its impact on socioeconomic, environmental and political variables is broad, often requiring extensive and costly data-collection efforts.

A second stumbling block that emerges when estimating the magnitude and distribution of effects that can be attributed to investments in transport infrastructure relates to the endogeneity of project placement, and the absence of well-defined treatment and comparison regions. The placement of infrastructural projects is often determined by a complex interplay of political, economic, social and environmental constraints (Robinson and Torvik 2005), which is often inscrutable to the researcher, rendering standard assumptions of causal inference implausible. As a result, the unobservables that determine the placement of the infrastructure and affect outcomes are likely to differ between treatment and comparison regions. Most observational studies will therefore struggle with eliciting causation. The ideal experimental method that would randomly assign regions to treatment and comparison groups, and in the process ensure that pre-treatment characteristics are identical between them, is also unlikely to be feasible for large, often politically sensitive transport projects. These challenges extend to evaluations of soft infrastructure, given that rules, regulations and government agencies dealing with the movement of cargo across borders are often not amenable to random assignment at the micro level, or to the creation of comparison groups for the purposes of an impact evaluation.

A third stumbling block associated with the main identification problem identified above is the lack of a well-defined area of impact, and a clear unit of analysis when estimating the magnitude and distributional effects of transport infrastructure. The catchment area of different types of infrastructural projects is difficult to define due to spillover and network effects. It is also challenging to define the right unit of analysis for different types of investment: transport infrastructure can directly affect households and firms by determining mobility costs for both people and goods. But it can also affect broader geographical areas via an infusion of resources in the form of wages, sourcing of inputs, physical materials and technical capacity for both building and maintaining the infrastructure. The mechanisms through which transport projects create winners and losers are likely to be location, programme and context specific, depending on the underlying geography of communities and of economic activity.

A related concern is that we still lack a full understanding of the functional form of the effects caused by transport infrastructure projects. In most cases, the researcher will face a difficult trade-off between short windows of evaluation that fail to capture the full effects of the project and longer windows that increase the risk of contamination and other confounding factors affecting treated or comparison units. While this is a common concern for evaluators,

the transport case renders it more acute due to the difficulty in identifying the typology of direct and indirect effects of transport across time in a particular economic context.

The nature of these constraints and their importance to the evaluator will depend on the type of transport project under evaluation. An evaluation of rural roads may for instance be more amenable to quasi-experimental and experimental tools than an evaluation of the building of a port, or of an airport.[4]

In this chapter we discuss two research projects that attempt to estimate both the magnitude and distribution of the impact of hard and soft transport infrastructure on firm behaviour, by addressing some of the concerns discussed above. The first project analyses the impact of investments in hard infrastructure on firm behaviour by taking advantage of a quasi-experiment with the rebuilding of a railway in Southern Africa. The second project investigates how the soft infrastructure of transport can affect firm behaviour by looking at the specific context of how firms respond to corruption at ports.

2   RAILROAD TO GROWTH

2.1   *Impact of Railways on Firm Behaviour*

The debate over the importance of transport infrastructure for economic development and growth can be traced to Rostow's (1959) endorsement of the railways as a precondition for economic take-off. This assertion was challenged, with different degrees of scepticism, by Fogel (1964) and Fishlow (1965), through a careful analysis of the economic role played by the railways in North America's 19th-century economic growth. In a note of caution, both authors argued that the US experience was unlikely to be replicated elsewhere, particularly in the developing world due to poor institutions, fragmented markets and limited technical capacity. Some of these concerns were later illustrated in Hirschman's (1967) study of the Nigerian railways. With the exception of a recent study that estimates the aggregate impact of the Indian railways on inter-regional trade in imperial India (Donaldson 2010), the empirical literature has made limited progress since.

The goal of this project is threefold:

(a) to investigate how investments in railways affect transport costs for different types of firm and sectors;
(b) to estimate how firms respond to these changes;
(c) to identify spillover and network effects across rail and road transport.

---

[4]For a detailed discussion of evaluation techniques for rural roads, see Van de Walle (2002) and Galiani (2007). For a general discussion of impact evaluations in infrastructure, see Estache (2010).

We define the unit of analysis at the firm level and the area of impact as firms that are connected through the road network within a 50 km radius to a functioning station of the railway. We argue that the placement of the railway is exogenous to the location of business, and to a firm's decision of what product to ship, which can be more or less suitable for railway transport. We monitor spillover and network effects by collecting detailed baseline and endline firm performance data on a range of transport-intensive and non-transport-intensive firms, by monitoring changes in road traffic along the roads that run close to the railways and by monitoring changes in throughput at ports connected to the railway, before and after the investments in the rail network take place.

2.2 The Placement of the Railway and the
Geography of Business: A Natural Experiment

In this study we exploit a quasi-experiment with the rebuilding of a railway connecting the economic heartland of South Africa to the port of Maputo in neighbouring Mozambique. The deep sea water port of Maputo lies 92 km from the South African border by rail and has for over a century provided the nearest facilities for the importers and exporters of the booming South African provinces of Gauteng and Mpumalanga to reach the sea.[5] Given the layout of the road and rail networks (see Figure 5.1), the next-best alternative—the port of Durban—is at least 1.5 times farther than the port of Maputo for firms located in northeastern South Africa.

The colonial and civil wars that took place in the 1970s, 1980s and early 1990s brought all Mozambican corridors to a standstill, crippling the country's integration into the regional transport network.[6] As a result, the majority of South African firms diverted their shipments to the port of Durban in South Africa's province of KwaZulu Natal.

The end of Mozambique's civil war in the mid-1990s heightened awareness of the importance of transport, particularly as a tool to revive an economy battered by decades of conflict. At the same time, a decade-long privatisation process in the 1990s was giving birth to a new entrepreneurial class

---

[5]For several decades, the production areas of citrus, timber and sugar in landlocked Swaziland also relied on the Maputo corridor as its primary export corridor. In central and northern Mozambique (see Figure 5.2), the transport corridors of Beira and Nacala secured access to the sea for business in Zimbabwe and in Malawi (see Figure 5.2). In the early 20th century, 40% of the Mozambican national budget relied on these lucrative cross-border transport corridors.

[6]See Figure 5.3 for evidence of the sharp decline of transnational cargo going through the railway connecting South Africa to Maputo since the 1970s. Notice that the colonial war in Mozambique began in 1964, so the 1975 figure should already be considered as a sharp decline from the normal functioning of the railway prior to the disruption caused by the war. No further data from this period were available from the national railroad agency (Caminhos de Ferro de Moçambique, CFM).
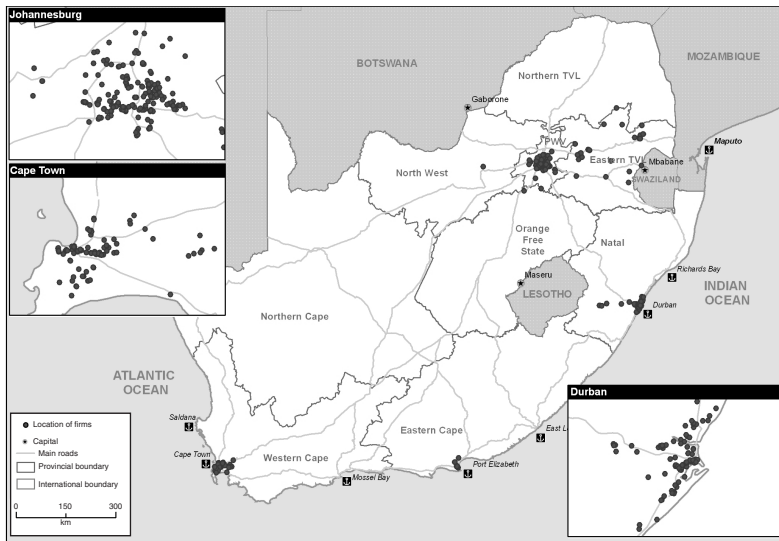
Figure 5.1: *Firms surveyed in South Africa and the choice of transport corridor between Maputo and Durban.*

closely connected to the party in government (Frelimo), which began to clamour for a viable transport system as the foundation of any process of economic growth. While Mozambican business exerted pressure for investments in a north–south connection that would integrate the Mozambican domestic market, scarce resources and the poor state of the country's infrastructure in the aftermath of two decades of war significantly constrained the government's set of feasible reforms.[7]

In the end, transport policies were primarily geared towards reviving the east–west oriented corridors from colonial times, connecting the regional international hinterland to Mozambican ports. The fall of apartheid in neighbouring South Africa led donors and multilaterals to focus more attentively on transnational economic initiatives that could contribute to regional peace through economic integration. For the government of Mozambique, the only way to attract concessional lending for transport infrastructure was to propose transnational investments that would promote regional integration.

---

[7]In the late 1990s, the Mozambican parastatal transport company CFM was plagued by an estimated capital deficit of over 80%; 75% of the nation's railway lines were operating at less than 15% of capacity and the estimated capital deficit across Mozambican ports ranged between 70 and 80% (interview with CFM 2006). Interviews with former Mozambican and South African politicians and transport experts suggested that the government of South Africa may have been directly implicated in the destruction of the Mozambican railway network as a means to force landlocked countries like Botswana or Zambia to use South African ports instead, in defiance of political sanctions in place against apartheid.
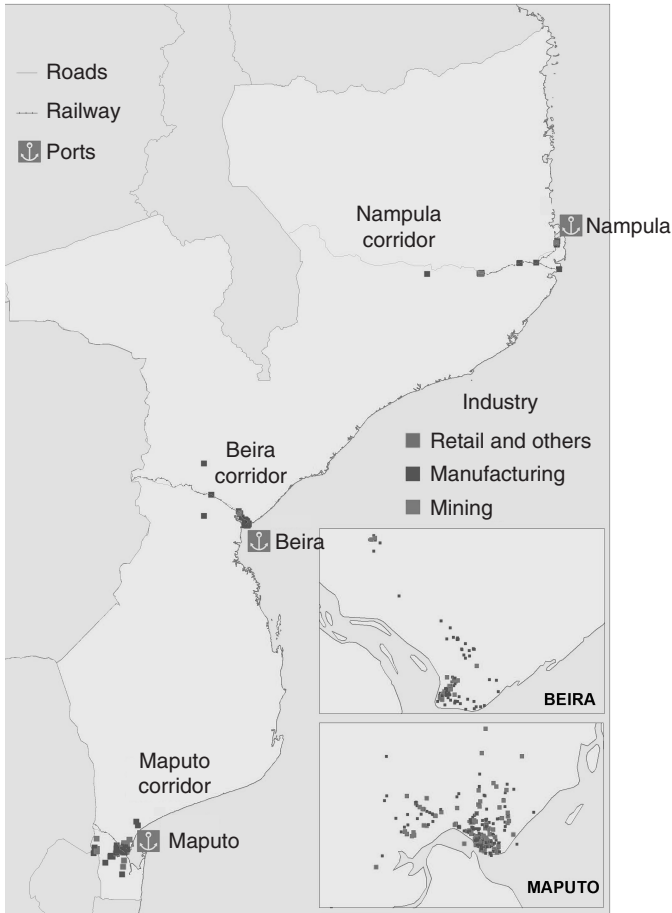
Figure 5.2: *Transport corridors in southern (Maputo), central (Beira) and northern (Nacala) Mozambique.*
Note: the dots correspond to the firms covered in the firm survey by industry type.

Constrained by the availability of funds, the decision was to rebuild the old railways in each transport corridor, following the layout determined in the 19th century to accommodate the transport needs of a very different set of economic players. The layout of the original railway had been determined primarily by the 19th-century geography of mining companies.

   In the early 2000s, three main transnational transport corridors crossing the southern, central and northern parts of Mozambique were slated for development (see Figure 5.2). Each would entail the rehabilitation and privatisation of a main deep-sea port (Maputo, Beira and Nacala), the rebuilding of a railway connecting each port to its transnational hinterland (South Africa, Zimbabwe
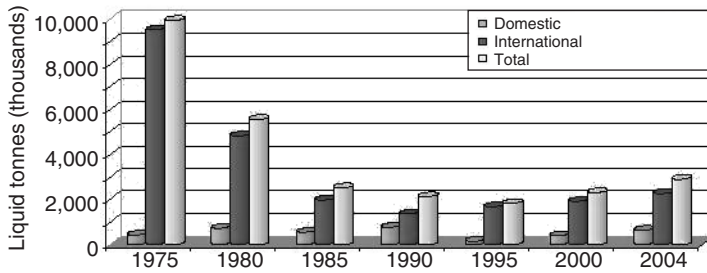
Figure 5.3: *Traffic volumes going through the Maputo Railway.*

Note: the colonial war began in Mozambique in 1964, so the 1975 base already reflects a significant drop in throughput from earlier times.

Figure 5.4: *Surveyed firms in South Africa.*

Note: Johannesburg and the area east of the city is served by the Maputo corridor. Durban in the southeastern seaboard of the country constitutes an alternative shipping route for these firms. Firms in the Western Cape (Cape Town) in western South Africa and in KwaZulu Natal (Durban) constitute the comparison regions for the impact of the Maputo railway.

and Malawi, respectively) and a main road. All three corridors would serve as a commitment device for regional integration and for the normalisation of economic and political relations among countries in the aftermath of a long period of political turmoil.

The main focus of this study is on the Maputo transport corridor connecting South Africa to the port of Maputo in Mozambique. This transport corridor links the most highly industrialised and productive regions of Southern Africa to the port of Maputo. In South African territory, the corridor crosses vast areas occupied by manufacturing, agro-processing, mining and smelting
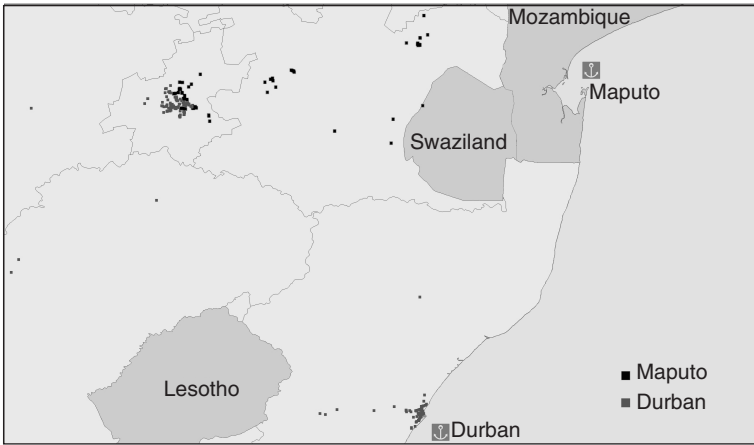
Figure 5.5: *Nearest port through the railway network for surveyed firms.*
Note: this map identifies the closest port through the railway network for all firms in the sample. Black squares denote firms that became closer to the port of Maputo when the railway was rebuilt in 2008.

industries. In Mozambique, the corridor serves areas of industrial and primary production containing steel mills, petro-chemicals, quarries, mines, smelters and plantations of forests, sugar cane, bananas and citrus.[8]

The port of Maputo was privatised, and the railway was rebuilt in 2008 under a public–private partnership. Though this railway had traditionally served bulk shipments since its inception, the spread of containerisation in the 1970s and 1980s significantly broadened the range of cargo that could travel by rail, while increasing the relative cost-effectiveness of this mode of transport due to economies of scale in shipments and significant fuel efficiency.[9] This is particularly important in the context of long distances between centres of production and consumption and ports, as is often the case in sub-Saharan Africa. The railway line that was originally designed to serve primarily mining companies now serves a diverse range of manufactur-

---

[8] See Figures 5.4 and 5.5 for a depiction of firms that were surveyed in this study in South Africa and the distribution of firms that, due to the railways, will become closer to the port of Maputo. The geographic market covered by the transport corridor today includes: Gauteng province, the services, industrial and financial hub of SSA, with a large concentration of manufacturing, agro-processing, retail, mining and smelting industries; Mpumalanga province, with a diversified economy in manufacturing, mining, tourism, chemicals, agriculture and forestry; Limpopo province, with the vast magnetite deposits of Phalaborwa; and Swaziland, with its exports of sugar, citrus and forest products and imports of cereals.

[9] See Figure 5.6 for evidence of the increase in container traffic reaching the port of Maputo in 2009 via rail. From January to December 2009 there was a 700% growth rate.
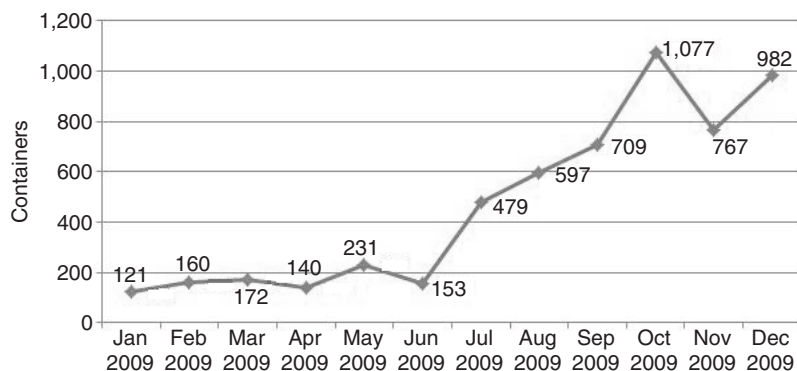
Figure 5.6: *Traffic volumes reaching the port of Maputo via rail in 2009.*
Note: between January and December 2009 there was a 700% increase in container traffic alone arriving by rail.

ing and retail firms, most of which emerged over the course of the century, before the railway was rebuilt.[10]

In addition, the depletion of mining reserves changed the geography of mining companies, moving them further away from the original layout of the railway. This creates plausibly exogenous variation in the 'placement' of the railway relative to the geography of business at the time of the rehabilitation of the line, which can be used to estimate a causal relationship between investments in the railway and their impact on firm performance.

To obtain an estimate of impact of the railway, we adopt an instrumental variables approach, in which exposure to treatment—defined as changes in transportation costs—is instrumented by the distance between a firm and a working station of the railway. To ensure the exogeneity of firm placement and of a firm's choice of sector—which determines whether its cargo is more or less suitable to rail transport—we restrict the estimation sample to firms that were established before 2002, prior to any investments in the Maputo transport corridor.

We further exploit the fact that the three Mozambican corridors developed at different speeds to conduct some important robustness checks to the main instrumental variable approach. Railways in both central and northern Mozambique were delayed by unexpected contractual disputes between the government and the parties to the public–private partnerships in charge of rehabilitating the lines, introducing an exogenous variation in the intensity and type of exposure to changes in transport costs experienced by different firms, depending on the corridor in which they were located. Contractual disagreements have also plagued the privatisation of the port of Nacala in

---

[10]See Figure 5.5 for evidence on the capacity of the port of Maputo to serve the South African market.
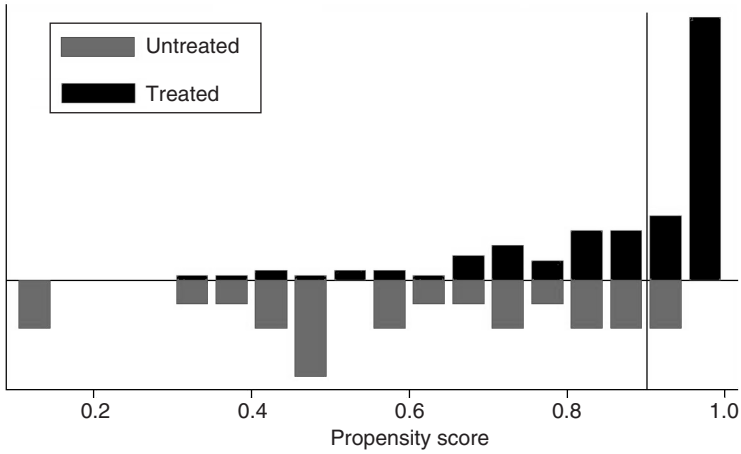
Figure 5.7: *Propensity scores for treated firms (Maputo region) and untreated firms (from Beira and Nacala) in Mozambique.*
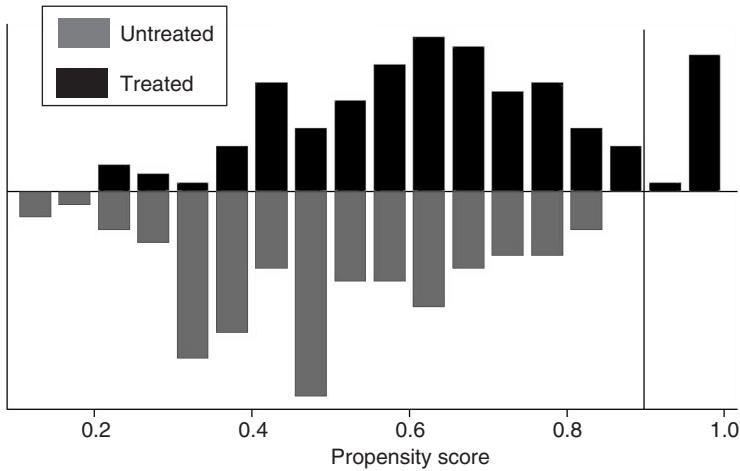
Figure 5.8: *Propensity scores for treated firms (Gauteng and Mpumalanga regions) and untreated firms (from Western Cape and KwaZulu Natal) in South Africa.*

northern Mozambique, delaying significantly the planned upgrading of port facilities.

This provides a unique opportunity to compare firms subject to similar institutional frameworks and general macroeconomic conditions, but with varying degrees of access to transport networks. Firms in the Beira corridor have access to a well-functioning and upgraded private port, but not to a
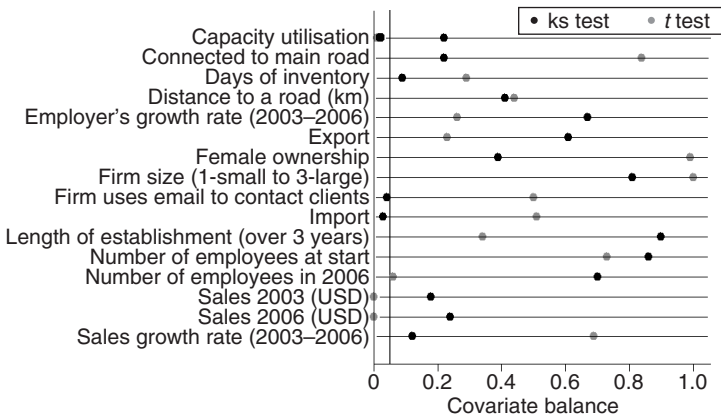
Figure 5.9: *Covariate balance for treated and untreated firms in Mozambique.*
Note: p-values for a two-sample t-test (with unequal variances) against the null of equal means, as well as a p-values from bootstrapped Kolmogorov–Smirnov tests against the null of equal distributions.

functioning railway. Firms in the northern Nacala corridor are yet to benefit from significant upgrades to either element of the planned transport corridor. Treated firms located in the catchment area of the Maputo corridor can therefore be matched to firms drawn from two distinct comparison groups located in the catchment areas of Beira (firms accessing a new port but no railway) and Nacala (firms without access to a new port or to a new railway). To isolate the impact of the upgrading of the Maputo railway, we apply a straightforward triple differences-in-differences estimation method to matched firms. This strategy relies on the assumption that the only factor that makes the trajectory of these three types of firm different during the period under analysis is that they were exposed to different transport choice sets.

Different techniques will allow us to test the robustness of these results. Following Crump *et al* (2009), propensities scores will be used to pre-screen firms with a probability of treatment between 10 and 90% to ensure that the regression analysis is conducted among firms with overlapping covariate distributions (Rosenbaum and Rubin 1983; Dehejia and Wahba 2002; Angrist and Hahn 2004).[11]

Figure 5.7 shows estimated propensity scores between treated units along the Maputo corridor and untreated units in the Beira and Nacala corridors for Mozambique. Figure 5.8 shows propensity scores between treated units in South Africa along the portion of the Maputo corridor connecting Mozambique to Johannesburg, with comparison units drawn from a sample of firms

---

[11]We follow Angrist and Hahn (2004) in considering that matching on propensity scores can increase the precision of the estimates in finite samples and make use of a wider set of covariate information, while reducing the dimensionality problem of simple matching.

Figure 5.10: *Covariate balance for treated and untreated firms in South Africa.*

Note: p-values for a two-sample t-test (with unequal variances) against the null of equal means, as well as a p-values from bootstrapped Kolmogorov–Smirnov tests against the null of equal distributions.

Figure 5.11: *Covariate balance for treated and untreated firms in South Africa, using KwaZulu Natal as the primary comparison group.*

Note: p-values for a two-sample t-test (with unequal variances) against the null of equal means, as well as a p-values from bootstrapped Kolmogorov–Smirnov tests against the null of equal distributions.
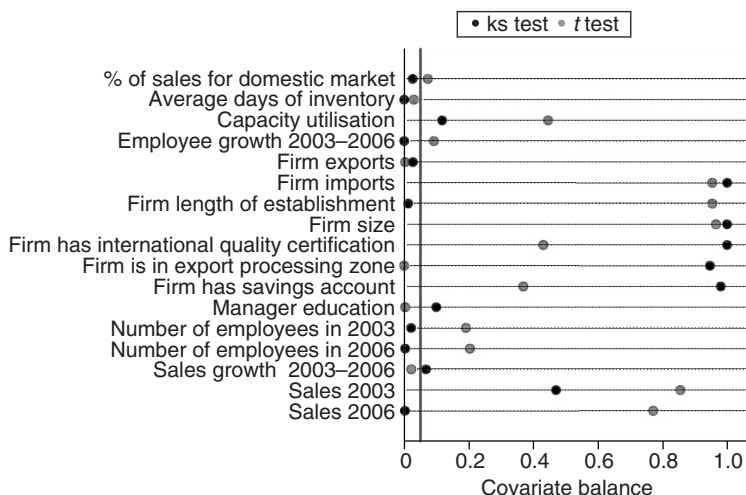
Figure 5.12: *Covariate balance for treated and untreated firms in South Africa, using the Western Cape as the primary comparison group.*
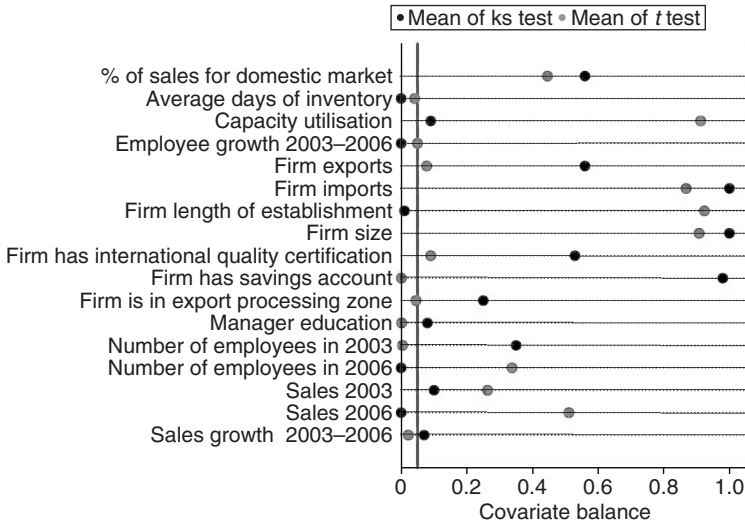
Note: p-values for a two-sample t-test (with unequal variances) against the null of equal means, as well as a p-values from bootstrapped Kolmogorov–Smirnov tests against the null of equal distributions.

in the Western Cape and in KwaZulu Natal provinces.[12] In both cases, the distribution of propensity scores ensures significant overlap in covariates of interest among treated and untreated units. In Figures 5.9–5.12 we report p-values for a two-sample t-test (with unequal variances) against the null of equal means, as well as a p-value from bootstrapped Kolmogorov–Smirnov tests against the null of equal distributions. The pre-treatment characteristics include past sales and sales growth, access to a main road, firm-level quality certifications, labour characteristics and export/import behaviour, among others. We obtain fairly good balance on observed characteristics as variable means are close for the most important variables, with the p-values indicating insignificant differences at conventional levels.[13]

---

[12]See Figure 5.4 for the distribution of firms in South Africa. We use a logit model to estimate the propensity scores with a few polynomials for continuous variables. Estimating regressions will also include the bias correction suggested by Abadie and Imbens (2006) to eliminate asymptotic bias from imperfect matches.

[13]Some time-invariant firm-level data were not adequately captured at baseline due to logistical constraints during survey implementation, so the sample currently used to estimate propensity scores and hypotheses testing is more restricted than the final sample that will be used to infer impact. Should the final sample not reveal satisfactory balance across the main covariates of interest, we will resort to synthetic matching following Abadie *et al* (2007).
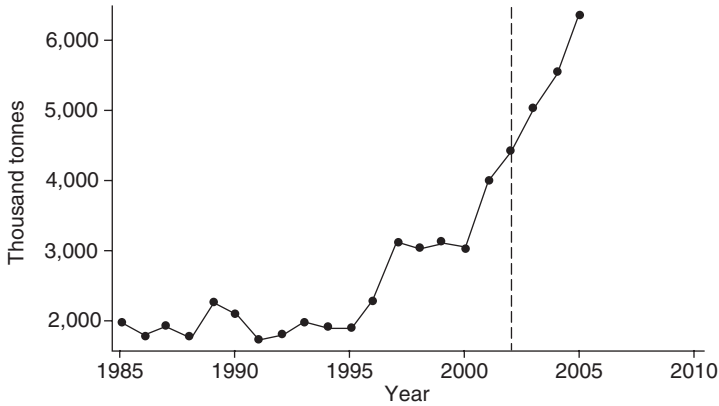
Figure 5.13: *Traffic volumes going through the Maputo port.*
Source: CFM. Note: there is some suspicion that the increasing trend in the early 2000s was artificially created to suggest to potential private investors that the port held great potential for growth.

The main outcomes of interest measured before and after the 2008 invest-ment in the Maputo railway are standard firm performance indicators such as investment levels, export behaviour, sales, product diversification and factor productivity. These indicators are being elicited through a survey of approx-imately 1,200 firms (600 in South Africa and 600 in Mozambique), located in treatment and comparison regions in both countries. The sample of firms was selected through stratified random sampling, with weights adjusted by the transport intensity of each sector.[14] The baseline survey was conducted in 2007–8 to obtain firm-level information pertaining to the financial year of 2006. The follow-up survey was scheduled for March 2011, and will obtain information for the fiscal year of 2010, two years after the reopening of the railway to international traffic. (See Figures 5.3 and 5.13 for an illustration of the immediate increase in rail throughput following the rehabilitation of the railway in 2008, and an increase in the number of international containers being exported through the port of Maputo.)

Another critical contribution of this study will be to attempt to identify interaction effects between transport modes. To assist in this task, we con-ducted a survey in 2007 of 220 trucking companies operating in South Africa and Mozambique, obtaining important information on vehicle operating costs, the contestability of road freight markets and the road freight rate spread for

---

[14]Transport intensity was calculated based on benchmark industry-level expenditures on transport costs in neighbouring countries, through data obtained from existing input–output tables.

different types of cargo.[15] A follow-up survey will be conducted in March 2011 to identify changes in trucking markets correlated with investments in the railway. This survey can provide further evidence on cross-price elasticities between road and rail transport. Unlike previous studies that extrapolate transport rates and costs based on small samples of shippers, this data set provides a more accurate and representative picture of the region's trucking industry.

To provide a fuller account of spillover and network effects, trucking data will be supplemented with road traffic data obtained from the South African and Mozambican road agencies. One important consideration is that the impact of improvements in railway access to the port can reach beyond the firms that use the railway due to network effects. The high volumes of cargo brought to the port by the railways can boost port throughput above the critical threshold that attracts ship calls and justifies regular dredging to accommodate larger and more frequent vessels. This can reduce overall shipping costs for any firm using the port, blurring the distinction between firms that benefit directly from the railways by using it, and firms that benefit from the railways due to spillover and network effects.

While this study will be unable to fully distinguish between these two effects, we hope to provide evidence of the relative impact of each. We will do so by observing firms' shipping decisions, their choice of transport mode and the transport costs they face, before and after the investments in the railway take place. To the best of our knowledge, this is the first study to exploit a quasi-experiment to estimate spillover and network effects across transport modes, using primary data collected at the level of the firm.

3   THE IMPACT OF SOFT TRANSPORT INFRASTRUCTURE ON FIRMS

Soft transport infrastructure can be defined as the network of transport bureaucracies and rules regulating the movement of goods within and across borders. One of the main challenges of conducting impact evaluations on soft transport infrastructure is the difficulty in identifying exogenous sources of variation in exposure to a given rule, regulation or transport bureaucracy.

A recent research project described in detail in Sequeira and Djankov (2010) attempts to assess the impact of the functioning of port bureaucracies on firms. The authors look at the specific context of corruption in ports and how it affects firm behaviour. The focus on corruption is motivated by the

---

[15]All interviews were conducted face to face, *in situ*; 80% of the respondents were selected by stratified random sampling of formal trucking companies, and 20% of the respondents were randomly sampled in the field, in areas where trucks tend to concentrate, such as the entrances of ports and lorry parks.

seemingly negative correlation between the capacity for a country to control corruption in public bureaucracies and the cost of trade, as illustrated by Figure 5.14.[16]

The setting for this project is identical to that described in Section 2. Prior to the rebuilding of the Maputo railway, the privatisation and rehabilitation of the port of Maputo in 2004 provided a set of South African firms with the choice between two equidistant ports: Maputo and Durban.[17] Since 2004, the barriers for freight transit along the transnational corridor connecting South Africa to the port of Maputo have been significantly reduced.[18]

Sequeira and Djankov (2010) began by generating a unique data set on directly observed bribe payments to port bureaucracies for a random sample of 1,300 shipments, and concluded that bribes were set according to the type of product being shipped. To assess the economic costs of corruption the authors then observed how bribe schedules at each port distorted two important margins of firm-level decision-making: the choice of port and the decision to source inputs domestically or internationally. The key identifying assumption to observe a causal effect between corruption and firm behaviour is that the location of firms and the type of product they ship are arguably exogenous to the level of corruption at the port of Maputo. The sample of firms under study is restricted to those that began operations in a given sector before 2002 to mitigate the problem of endogenous firm location and product choice, given that the port of Maputo only became a viable option for South African firms when it re-opened to international traffic in 2004. Figure 5.13

---

[16]Ports offer a rich setting for studying corruption, as they represent an administrative monopoly with limited institutional accountability. Clark *et al* (2004) rely on indirect measures of bribes to argue that corruption is an important source of port inefficiency, and Yang (2008a) provides evidence on how corruption in ports is hard to displace. According to the World Bank Enterprise Survey of 2008, in Mozambique alone, over 50% of firms reported having to pay bribes to transport bureaucracies at ports in 2007.

[17]Given its strategic location, the port of Maputo has historically been considered a critical part of South Africa's transport network. The port suffered great losses during the Mozambican civil war in the 1980s and 1990s, and reopened to international traffic in 2004 under private management. Today, and together with Durban, the port of Maputo serves as the primary transportation route to the sea for the booming South African provinces of Mpumalanga, Gauteng and Kwazulu Natal. There is a third port in the region, the port of Richards Bay, which is located approximately halfway between Durban and Maputo along South Africa's eastern seaboard. This port was developed in the late 1970s to serve a select group of private shareholders and is primarily used by large mining conglomerates to ship bulk cargo. Given the restricted access to this port, it is not considered a substitute for either Durban or Maputo for the type of firms covered in the study.

[18]For example, there are no visa requirements for truck drivers from either country to operate along the transnational Maputo corridor.
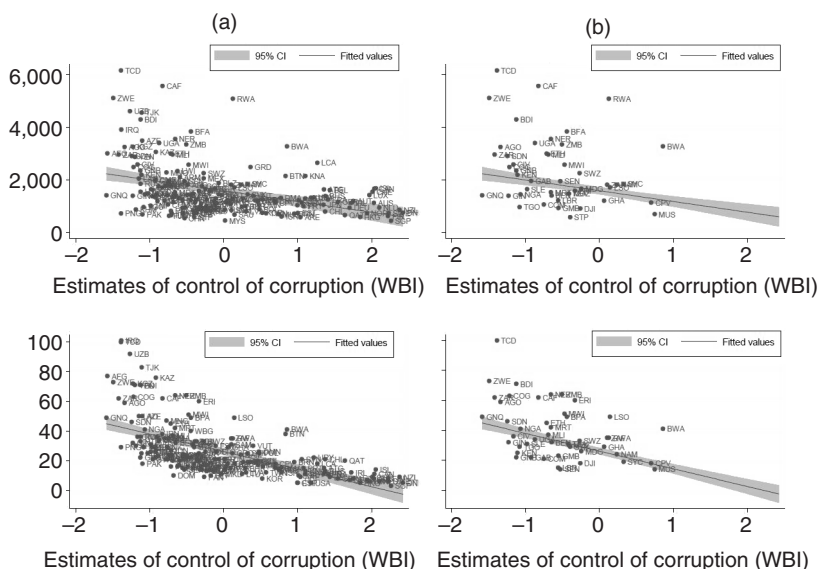
Figure 5.14: *Trade costs and the capacity for a country to control corruption. (a) Cost of imports (US dollars) (top) and days to import (bottom). (b) Cost of imports (US dollars) SSA (top) and days to import (bottom).*
Source: World Bank Governance Indicators. Note: the two graphs that make up part (b) represent data for sub-Saharan Africa (SSA) alone.

illustrates the rapid increase in volumes (+200%) going through the port of Maputo following privatisation in 2004.[19]

3.1 Sources of Bureaucratic Variation and Corruption

Sequeira and Djankov (2010) identify exogenous sources of bureaucratic variation in an attempt to understand how bureaucratic structure can affect the level and type of bribes observed in equilibrium. The study begins by identifying two types of officials who differ in their authority and in their discretion to stop cargo and create opportunities for bribe payments at each port: customs officials and port operators. In principle, customs officials have greater discretionary power to extract bribes than regular port operators, given their broader bureaucratic mandate and because they can access full information

---

[19]Interviews with the Mozambican transport parastatal (CFM) suggested that the increasing trend in port volumes noticeable prior to 2004 was an artefact of the political negotiations surrounding the public–private partnership. The port was portrayed as already being on a path of expansion, but it is unclear that this was indeed the case.

on each shipment and each shipper at all times.[20] Regular port operators, on the other hand, have a narrower mandate to move or protect cargo on the docks, and they lack access to the shipment's documentation specifying the value of the cargo, the client firm and its origin/destination, among other valuable information.[21]

The port bureaucracies of Maputo and Durban differed in two important organisational features that determined which of the two types of port officials described above had more opportunities for bribe extraction: the high extractive types—customs agents—or the low extractive types—port operators. These differences were driven by the degree of interaction between clearing agents and customs officials, and by the type of management overseeing port operations. In Durban, the level of direct interaction between clearing agents and customs' agents was kept to a minimum since all clearance documentation was processed online. In contrast, the level of interaction in Maputo was high since all clearance documentation was submitted in-person by the clearing agent.[22] The close interaction between clearing agents and customs officials created more opportunities for corrupt behaviour to emerge in customs in Maputo relative to Durban.

In Maputo, port operators were privately managed, but in Durban this was only the case for bulk cargo terminals, since container terminals were still under public control. Private management in Maputo and in the bulk terminals in Durban was associated with fewer opportunities for bribe payments due to better monitoring strategies and stricter punishment practices.[23] These organisational features determined that the high extractive types in customs had more opportunities to extract bribes in Maputo, while the low extractive types in port operations had more opportunities to extract bribes in Durban.

---

[20]Customs officials possess discretionary power to single-handedly decide which cargo to stop and whether to reassess the classification of goods for tariff purposes or to validate prices. They can also threaten to conduct a physical inspection of the shipment, which can delay clearance for up to four days, or arbitrarily request additional documentation from the shipper.

[21]In the sample of 1,300 shipments observed in this study, bribes were paid to different types of port officials: agents in charge of adjusting reefer temperatures for refrigerated cargo stationed at the port; port gate officials who determine the acceptance of late cargo arrivals; stevedores who auction off forklifts and equipment on the docks; document clerks who stamp import, export and transit documentation for submission to customs; port security who oversee high-value cargo vulnerable to theft; shipping planners who auction off priority slots in shipping vessels; and scanner agents who move cargo through non-intrusive scanning technology (Sequeira and Djankov 2010).

[22]The level of red tape was, however, similar in both countries. South Africa and Mozambique require the same number of documents to process the clearing of goods through their ports (World Bank 2007).

[23]Interviews with SAPO, Portnet, SARS, Customs and the Maputo Port Development Corporation (2006, 2007).

A second important difference between the two port bureaucracies was that officials with opportunities to extract bribes differed in their time horizons. As part of a comprehensive reform programme triggered by the merging of the Customs Agency and the Tax Authority in 2006, customs in Maputo adopted a policy of frequently rotating agents across different terminals and ports. While customs officials in Maputo could be in a post for as little as six weeks, port operators in Durban had extended time horizons given the stable support received from dock workers' unions.[24] Sequeira and Djankov (2010) predict that the higher extractive types with the shortest time horizons—customs' officials in Maputo—would extract higher and more frequent bribes (Campante *et al* 2009).

An important identifying assumption is that differences in the organisational structure of each port bureaucracy—private versus public management and the level of technological investment—were not determined by the level of corruption at each port. In South Africa, dock workers' unions spearheaded a long and successful fight against the privatisation of port operations, particularly for container terminals. The political strength of the organisation is deeply rooted in the historical role played in the struggle against apartheid, which culminated in the active participation of labour unions in the tripartite political alliance that gave birth to the first post-apartheid government in South Africa.[25]

Bulk terminals in Durban are owned and managed by large mining conglomerates that successfully negotiated control over their own transport chains with government. The economic strength of mining groups goes back to the 1950s and 1960s, when the mass export of minerals funded South Africa's Import-Substitution Industrialization (ISI) model of development. In the 1980s and 1990s, as South Africa struggled under the weight of economic sanctions, the export of coal and iron ore became once more the primary sources of foreign exchange, and the largest contributors to GDP. As a result of their economic importance across time, private groups were able to develop and manage all bulk terminals in South Africa's ports to this day (Feinstein 2005).

As the African National Congress (ANC) secured its grip on power in the late 1990s, the party began to drift away from its earlier links to the labour movement. South Africa's new leadership revealed a more technocratic bent, becoming increasingly concerned with the macro-fundamentals of the economy and sparring frequently with trade unions over economic policy. New business groups clamoured for the privatisation of port operations, with the

---

[24]Information obtained through interviews with the Customs Agency in Maputo, the management of port operations in Durban (SAPO), and South African Transport and Allied Workers Union (SATAWU), the transport union in Durban.

[25]SATAWU enlists 82,000 members and is affiliated with the Congress of South African Trade Unions (COSATU). COSATU is an active member in the tripartite political alliance with the ANC and the Communist Party currently in power. It derives its power from the capacity to mobilise large masses for electoral turnout (Feinstein 2005).

dual goal of increasing the productivity and reducing the cost of port services. There was a clear understanding that the transport needs of the emerging manufacturing sector diverged from those of mining and farming, which had dictated past developments in transport policy.

But the privatisation of port operations presented the executive with several political and financial challenges. For one, ports represented the most profitable branch of the transport parastatal's business. In addition, revenue from port activities in South Africa was locked into a complex cross-subsidisation scheme to support costly railway operations and a large pension scheme for its workers, which was inherited from the apartheid days.[26] The privatisation of ports would challenge this fine balance, while creating unsustainable uproar in one of the ANC's most powerful bases of popular support. Finally, the port of Durban is nestled in KwaZulu Natal (KZN) province, long a bastion of the opposition party (Inkatha Freedom, IFP) and increasingly a swing province in post-2000s election cycles. The privatisation of the port of Durban would therefore have carried high political risk of costly conflict with labour, just when the ANC sought to gain political control over the region (Gumede 2005).

In Mozambique the capital requirements to reverse the derelict state of the country's port infrastructure forced the government to resort to concessional lending and to agree to the privatisation of port services. The 1980s and 1990s brought fast-paced technological change driven by containerisation, which significantly altered the production function of ports. Stevedores, who in the past were called upon to coordinate complex processes of loading and offloading on the docks, became increasingly less important for harbour operations. New capital-intensive investments required instead a flexible and small labour force. While in Durban the strength of dock worker unions had avoided retrenchments and privatisation, in Maputo there was no tradition of unionised dock workers.[27] As a result, the port of Durban has kept container terminals under public management but bulk terminals under private management, while the port of Maputo has been entirely under private management.

3.2   Primary Data Collection

This section describes in more detail the empirical setting and the nature of the shipment process in South Africa and in Mozambique. By law, no firm in either country is allowed to interact directly with customs or port operators. Instead, firms have to resort to clearing agents, who specialise in clearing

---

[26]Information obtained through interviews with Portnet and Transnet, South Africa.

[27]In the late 1990s, an attempt was made to privatise the Mozambican customs' agency in partnership with the British Crown Agents. The partnership ended in 1998, and consisted mostly of a series of nominal reforms and technological upgrades for data management (Interview with Alfandegas de Mocambique 2007).

cargo through the port or border post, mostly through ad hoc, shipment-based contracts. The market for clearing agents is moderately competitive following the deregulation of the trade in the 1980s in South Africa and in the 1990s in Mozambique. In the sample tracked by Sequeira and Djankov (2010), 80% of firms engaged in direct contracts with clearing agents, 65% of which were for a one-time shipment. Clearing agents play a pivotal role, as they make all bribe payments to port officials on behalf of client firms.

Sequeira and Djankov (2010) relied on three main sources of primary data. The first source of data was a tracking study designed and implemented by the International Finance Corporation (IFC) in the ports of Maputo and Durban, and in the border post between South Africa and Mozambique. The IFC hired well-established clearing agents to track all bribe payments to officials in a random sample of 1,300 shipments between March 2007 and July 2008.[28] Clearing agents agreed to record detailed information on the date, time of arrival and clearance of each shipment as well as on expected storage costs at the port, the size of the client firm and a wide range of cargo characteristics (size, value and product type). They also noted the primary recipients of bribes, the bribe amounts requested and the reason for a bribe, ranging from the need to jump a long queue of trucks to get into the port to evading tariffs or missing important clearance documentation. For a random subset of shipments, the IFC hired local observers who accompanied clearing agents throughout the clearing process to verify the accuracy of the data. These observers began shadowing clearing agents several weeks before the tracking study took place in order to become familiar with all clearing procedures. To avoid any suspicion, the observers were similar in age and appearance to any other clerk who normally assists clearing agents in their interactions with port officials. There was no significant difference between the data reported with and without the observer present. Data from this tracking study provided a clear measurement of expected bribes at each port for different types of shippers and different types of shipments.

The second source of data was an enterprise survey conducted by the authors in 2007 covering 250 firms located in the overlapping hinterland of the ports of Durban and Maputo that elicited firms' responses to bribe schedules at each port. The survey also collected information on firms' perceptions of the quality of each port, their shipping strategies, and on the characteristics of their average shipments, such as frequency, size and degree of urgency proxied by measures of deviation from sector-level firm inventories. The sample was stratified by firm size and industry, covering a range of both transport-intensive and non-transport-intensive firms. The data were then used to identify firms' choices of transport corridor and port, given their

---

[28]The sample size was restricted to eight clearing agents, given the illicit nature of the bribe payments and the IFC's concern with ensuring discretion in the data collection to maximise its accuracy. More details are available in Sequeira and Djankov (2010).

location, the urgency of their shipments and the characteristics of their cargo that would make them more or less vulnerable to corruption.

The third source of data was a trucking survey, covering a random sample of 220 trucking companies operating in both the Maputo and Durban corridors, that the authors conducted in order to accurately measure overland transport costs in the region. As described in Section 2, this survey elicited detailed information on vehicle operating costs, including maintenance and fuel costs, average transit times on each corridor and transport rates charged to firms.[29] The study also collected a range of fees associated with the movement of goods to enable an accurate calculation of road transport costs in different transport corridors. These data enabled a more precise calculation of transport costs at the firm level.

3.3   Efficiency Costs of Corruption

Sequeira and Djankov (2010) argue that the efficiency costs of corruption will depend on the structure of the market for bribes, the type of corruption officials engage in and whether and how public officials price discriminate when setting bribes. Motivated by the literature on market structure and corruption, Sequeira and Djankov (2010) predict that the costs corruption imposes on users of public services depend on the type of competition port bureaucracies engage in, and on how front-line officials set bribes. For one, corruption in port bureaucracies can have limited efficiency costs when bureaucrats engage in perfect competition or perfect collusion (Shleifer and Vishny 1993). Whether the conditions for perfect competition or perfect collusion hold depends on the way bureaucracies are organised.

As shown in Table 5.1, the bribe data collected in Sequeira and Djankov (2010) did not support the perfect-competition hypothesis, in which bribes were competed to zero across ports, or the hypothesis that port bureaucracies were able to collude when setting bribes. This non-cooperative outcome in bribe-setting across bureaucracies is likely to be the result of high coordination and communication costs between different levels of bureaucrats in different countries, and the fact that price-cutting and any deviation from 'joint monopolist' prices would not be credible in the face of capacity constraints that would prevent a single port from serving the entire market.

More importantly, due to the way in which each port bureaucracy was organised, public officials involved in corruption at each port differed in their discount rates. Customs officials in Maputo had high discount rates, while port operators in Durban had low discount rates, implying that deviations from the 'joint monopolist' bribe level would not be internalised in the same way by the different bureaucrats. Bribe levels in each port were instead determined by

---

[29]These data were validated through 'mystery client' exercises through which over 75 transport firms were contacted with a request for specific rates for a standard shipment of goods to and from each port.

Table 5.1: *Comparing the ports of Durban and Maputo.*

Port characteristics	Maputo	Durban
Average quay length (m)	238.4	225.9
Average alongside depth (m)	10.8	10.54
Minimum alongside depth (m)	9.5	6.1
Berth occupancy rates (%)	30	100
Crane movements per hour (TEU)	15	15
Days of free storage	21	3
Average number of days to clear customs (median of the distribution)	4	4
Longest number of days to clear customs (median of the distribution)	7	8
Average distance to Johannesburg (km)	586	578
Technology in customs	In-person submission	Online submission
Port performance ranking (out of 5)	3.4	3.7
Security	ISPS certified	ISPS certified
Document submission	In-person	Online
Management of terminals	Private	Public

Source: Sequeira and Djankov (2010). *Notes*: The port performance ranking was obtained through the IFC's survey of 250 firms in South Africa and corresponds to an unweighted average of the score assigned to each port in a scale of 1 (very poor) to 5 (very good), along the following dimensions: (a) facilities for large and abnormal cargo and flexibility in meeting special handling requirements; (b) frequency of cargo loss and damage; (c) convenient pick-up and delivery times; (d) availability of information concerning shipments and port facilities; (e) speed of on the dock handling of containers; (f) availability of intermodal arrangements (rail, road and port); (g) port cost. ISPS stands for the International Ship and Port Facility Security Code. All countries that are members of the SOLAS convention are required to be ISPS certified. SOLAS (the International Convention for the Safety of Life at Sea) is the most important of all international treaties concerning the safety of merchant ships. TEU ('twenty-foot equivalent unit') is a unit of cargo capacity often used to describe the capacity of container ships and container terminals, based on the volume of a 20-foot container.

the extractive capacity of the different bureaucrats who were able to engage in corruption. Bureaucrats acted as independent monopolists when setting bribes, maximising their own individual bribe revenue as opposed to that of the bureaucracy they belonged to. This uncoordinated bribe setting significantly increased the efficiency costs of corruption, leading to a distortionary tax on firms.

Sequeira and Djankov (2010) identify another important margin of variation with the potential to affect the cost that corruption imposes on users of port services. 'Collusive' corruption emerged when port officials and private agents colluded to share rents generated by the illicit transaction. A clear example of this was when private agents colluded with customs officials to evade tariffs. 'Coercive' corruption took place when a public bureaucrat coerced a private agent into paying a fee just to gain access to the public service. In this case, the private agent did not benefit from any rent from the illicit transaction, as the bribe was mostly extortionary. The distinction between cost-reducing (collusive) or cost-increasing (coercive) corruption highlights

how corruption can bring advantages or disadvantages to firms in different environments.

Sequeira and Djankov (2010) build on Shleifer and Vishny (1993) and Olken and Barron (2007) to propose that the costs of corruption will also depend on the price discrimination strategy public officials engage in when setting bribes. For example, the efficiency costs of corruption would be low if bureaucrats did not price discriminate or if they price discriminated efficiently, but would be high otherwise. The no price discrimination case would correspond to a lump-sum bribe payment over each shipment, which would be equivalent to a non-distortionary tax on accessing port services. If bureaucrats price discriminated efficiently, bribes would also not distort firms' decisions.

Examples of efficient price discrimination would be setting bribes according to the time preferences of users, according to their ability to pay, or according to the distance each firm needs to travel to reach the port. While still costly to firms, corruption in soft transport infrastructure with this type of price discrimination would just represent a transfer from private agents to bureaucrats that would not distort allocative efficiency (Leff 1964; Huntington 1968; Lui 1985).

Bribes at the ports of Maputo and Durban were determined primarily by product characteristics and were high in magnitude, as illustrated in Table 5.1. In Maputo, the median bribe represented a 129% increase in total port costs for a standard 20-foot container, and was equivalent to a 14% increase in total shipping costs—including overland transport, port clearance costs and sea shipping—for the container to be shipped between Eastern Africa or the Far East and the hub of economic activity in South Africa.

In Durban the incidence of bribe payments was lower, but the median bribe was still equivalent to a 32% increase in total port costs for a standard 20-foot container. It was also equivalent to a 4% increase in total shipping costs for a container on the same routes from Eastern Africa or the Far East.[30] In Maputo, bribes were paid primarily to customs by shippers of high-tariff goods. These represented 'collusive' arrangements with Mozambican firms intending to evade tariffs and a 'coercive' tax for South African cargo in transit through the port. While domestic cargo could pay a bribe to evade tariffs, transit cargo en route to or from South Africa had to pay a bribe just to avoid an arbitrary increase in the transit bond, since no tariff evasion is possible for

---

[30]Bribes were also high and significant when measured as a percentage of each bureaucrat's salary. The median bribe in Maputo was equivalent to approximately 24% of the monthly salary of a customs official, while in Durban the median bribe was equivalent to 4% of the monthly salary of a regular port operator (CPI adjusted). Sequeira and Djankov (2010) conducted a back-of-the-envelope calculation suggesting that the Mozambican customs' official's monthly salary could grow by more than 600% due to corruption. Assuming higher volumes for a regular port operator in Durban, the salary increase due to corruption would be in the order of 144% per month.

these shipments in Maputo.[31] The transit bond is a common precautionary measure taken by transit countries to ensure that duties are paid were the cargo to be diverted and remain in transit countries.[32] The amount of this transit bond is in principle determined by the tariff amount the cargo would have to pay according to the Mozambican tariff code.

While the transit bond is refunded once the cargo enters South Africa, and should in principle not represent a real cost, firms would often complain that the arbitrary fluctuations in the magnitude of the transit bond due to corruption created significant logistical costs. The transit bond is important in the empirical set-up insofar as it creates an exogenous variation in South African firms' exposure to coercive corruption at the port of Maputo, and is therefore critical for the identification of the distortionary effect of this type of corruption.[33]

In Durban, bribes were paid to document clerks, cargo handlers and port security, all of which had low extractive power due to limited access to information on the shipment, and limited authority to stop and delay cargo. Bribes were set according to the storage costs the cargo would have to pay were it to be moved from the general docks into private depots.

Just like in the case of tariffs, associating the level of the bribe to potential storage costs combined the desirable features of reducing both the informational cost of bargaining over bribes, and the risk associated with the illicit transaction. Storage costs are easy to calculate based on the volume of the shipment and on the type of product being stored.

Tariff levels are also directly observed by both parties. Port operators and customs officials assume that firms will always want to avoid these costs. For storage costs, the timing of when the cargo had to move to the depot also depended on port congestion levels—a variable not directly observed by the shippers—allowing a port operator to exploit an important informational asymmetry to extract a higher bribe with low probability of detection. These payments fell under the category of 'coercive' corruption, since they

---

[31] South African firms pay tariffs only when the cargo enters South Africa, irrespective of whether the point of entry of the shipments is the port of Maputo or the port of Durban. While South African cargo is travelling approximately 83 km in Mozambique before entering South African territory, firms will have to pay a transit bond to Mozambican customs, which is refunded once the cargo crosses into South Africa.

[32] All the clearing agents who participated in the Sequeira and Djankov (2010) study confirmed that, while transit bond procedures were in principle straightforward and easy to implement, customs in Maputo would often seek to re-classify shipments or change shipment values in order to negotiate a bribe against the threat of an arbitrary increase in the amount of the transit bond.

[33] By restricting the analysis to firms that chose their geographic location and the products they ship before the port of Maputo reopened to international traffic in 2004, the study ensures that a firm's choice of product is orthogonal to the type of corruption observed later at the port of Maputo.

Table 5.2: *Summary statistics of bribes at each port.*

Variable	Maputo	Durban
Probability of paying a bribe (%)	52.75	36.09
Mean bribe amount (US dollar)	275.3	95
Mean bribe as % of port costs	129	32
Mean bribe as % of overland costs	25	9
Mean bribe as % of ocean shipping to East Africa	37	13
Mean bribe as % of ocean shipping to Far East	46	37
Mean bribe as % of total shipping costs	14	4
(overland, port and ocean shipping)		
Median bribe (US dollar) if firm >500 km from port	192	35
Median bribe (US dollar) if firm <5 km from port	190	32
Monthly salary increase of port official (%)	600	144
Real monthly wage of port official in US dollars (CPI adjusted)	692	699

Source: Sequeira and Djankov (2010).

ultimately represented a cost above and beyond what shippers would normally have to pay in the absence of corruption. In most instances, the payment of a bribe took place before the cargo had remained in the general docks for the full three days for which it was entitled to stay for free (Sequeira and Djankov 2010).

The study concludes that the efficiency costs of corruption were high because bureaucracies were organised in a way that prevented perfect collusion or perfect competition in bribe setting, and because public officials did not price discriminate efficiently to maximise joint welfare with the shipper, but were instead responding to their own strategic concerns about how to minimise both the informational costs of bargaining over bribes and the probability of detection of the illicit transaction.

3.4   Corruption in Ports and Firm Behaviour

To measure the costs corruption in ports imposed on firms, Sequeira and Djankov (2010) begin by observing how firms choose which port to use, given their location and the type of product they ship. The type of shipment determines a firm's vulnerability to corruption at each port. The analysis was restricted to firms that were equidistant from the ports of Maputo and Durban (or closer to the port of Maputo), facing similar cargo-handling technologies and similar logistics services for standard cargo, but with different levels of expected corruption.[34] An important feature of this empirical set-up is that neither port dominates the other in terms of overall speed and quality of

---

[34]The choice of port is not trivial in this setting, since cargo travels long distances—an average of 588 km—between centres of production or consumption and ports, at very high costs of overland transport.
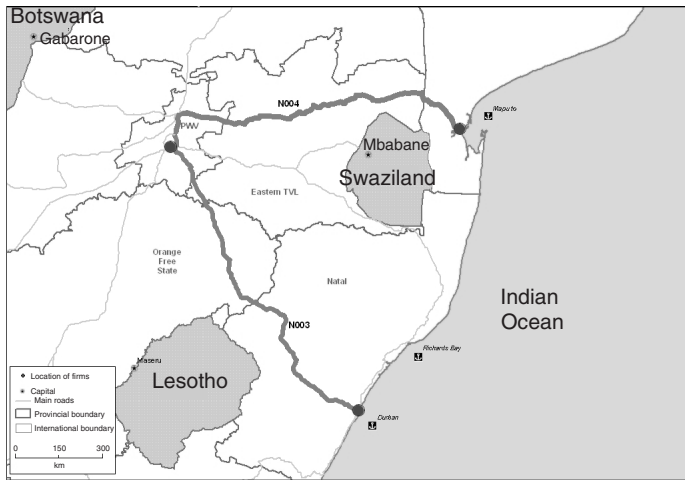
Figure 5.15: *Road network in South Africa.*

cargo handling. (See Table 5.2 for a summary of the main characteristics of each port.[35])

One assumption in the analysis of firms' choices is that firms are able to switch between ports at low cost. The survey data supported this assumption, as over 65% of surveyed firms outsourced transport services to freight forwarders and clearing agents, primarily through spot contracts with high turnover rates. Less than 4% of the firms surveyed in the study had ever made a long-term investment in either port. When asked about an alternative transport route, more than 50% of firms using either corridor identified Maputo or Durban as a real alternative.

To measure the impact of corruption on firms' choice of port, Sequeira and Djankov (2010) restrict their analysis to a sample of South African firms established prior to 2002 (two years before the port of Maputo was privatised and rebuilt).[36] The underlying assumption is that, in the absence of corruption, firms would minimise overall transport costs, which are a linear function of

---

[35]Sequeira and Djankov (2010) gathered both administrative and survey data to assess the trade-offs associated with the ports of Maputo and Durban. For instance, in addition to the actual cost of shipping and handling, a firm's shipping choice may also be influenced by the time it takes to clear cargo at each port. Survey data revealed that the median of the distribution of the average number of days reported by firms to clear customs was similar for both ports (four days) and the median of the distribution of the longest number of days reported to clear customs was only slightly higher in Durban (eight days) than in Maputo (seven days).

[36]Given the layout of the road network, it is not viable for any Mozambican firm to use the port of Durban as its main port of entry or exit. See Figure 5.15 for evidence.
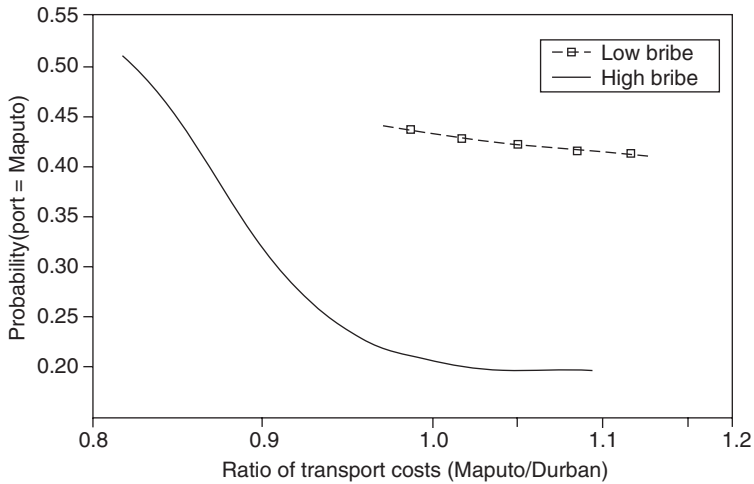
Figure 5.16: *Non-parametric regression of the probability of choosing the port of Maputo relative to the transport cost of reaching the alternative port of Durban.*

Source: Sequeira and Djankov (2010). Note: 'high bribes' represents firms shipping cargo that falls under a high-tariff category in Mozambique and that are therefore more vulnerable to corruption, and 'low bribes' represents firms that ship cargo subject to a low tariff in Mozambique.

geographic distance. With corruption, firms would minimise both transport costs and expected bribes when deciding which port to use.

Sequeira and Djankov (2010) then regress parametrically and non-parametrically the probability of a firm choosing the port of Maputo, given its location (measured implicitly as the transport cost of reaching the alternative port), the urgency of its shipments proxied by the deviation from the average level of inventories for each firm by size and sector, and the attributes of the goods it ships, which would determine the schedule of bribes faced at each port.[37]

The main findings were that a firm that shipped goods subject to a high-tariff classification in Mozambique would have a 22–23% lower probability of choosing the port of Maputo. Another way of illustrating this distortion is to assume that, in the absence of corruption or if bribes were set according to the geographic location of the firm, the indifferent firm would be located at the point that equated transport costs to either port. As seen in Figure 5.16, at the point of transport cost equivalence around 1, South African goods that are less vulnerable to corruption have a higher probability of being shipped

---

[37]Transport cost indicators include all costs of road transport, border fees, tolls and port handling costs. The specification also includes indicators of the perishability of the cargo, its value and several other shipment and shipper level characteristics, like the size of the firm, and whether it is import or export cargo, with multiple interactions.

through the port of Maputo. In this figure, low and high bribe firms are those that fall under low- and high-tariff categories in Mozambique, respectively. The only channel through which Mozambican tariffs should affect the choice of port by South African firms is through its effect on the transit bond. In the absence of corruption in customs in Maputo, the transit bond would be predetermined by the Mozambican tariff code and swiftly paid and refunded the same day, once the cargo reached South Africa.

However, South African firms reported that there was significant uncertainty about the level of transit bonds they would have to pay, since this value was not revealed by Mozambican customs until the cargo reached the port. Given that corrupt officials at the port of Maputo targeted high-tariff goods to attempt to extract a bribe, regardless of whether cargo was in transit or not, South African firms shipping goods that happened to fall under a high-tariff classification in Mozambique would avoid the coercive corruption faced in Mozambique and would opt to ship through Durban instead.[38]

The main results suggested that even when accounting for distance, perishability and the urgency of the shipment, expected bribes were a strong predictor of the choice of port. For example, 46% of South African firms in the sample located in regions in which overland costs to the port of Maputo were 57% lower still followed the long way around to Durban in order to avoid higher bribe payments. Of these, 75% were shipping perishable cargo and 74% were shipping urgent cargo.[39] For firms rerouting to the least corrupt port, additional costs in the form of road transport could add up to an 8% overall increase in yearly transport costs relative to a firm that shipped cargo that was less vulnerable to corruption from the same location, which could still use the port of Maputo.[40]

---

[38]Despite the fact that local and transit cargo have very different elasticities of demand for port services in Maputo, Sequeira and Djankov (2010) discuss two possible reasons for why customs officials in Mozambique were not discriminating between transit and local cargo when setting bribes. First, customs officials in Mozambique had very short time horizons and consequently high discount rates, given that the probability of staying in the same post for longer than six weeks was fairly low. The cost of requesting high bribes from firms with a high elasticity of demand would be felt primarily in the future so officials did not fully internalise it in the present. Furthermore, at the time the study was conducted, transit cargo only represented about 15% of the total number of shipments moving through Maputo. Adopting more sophisticated bribe-setting strategies that discriminate between transit and local cargo could potentially increase the probability of detection of the illegal transaction due to the perceived unfairness of charging different bribes to South African and Mozambican shippers (Sequeira and Djankov 2010).

[39]See Figure 5.18 for an illustration of the weak correlation between the distance between each South African firm and each port and their choice of port for imports.

[40]This calculation is based on the average number of shipments a firm in this region ships a year, the average size of the shipments and self-reported data on the total transportation costs of the firm for 2007, which were obtained through the enterprise survey.
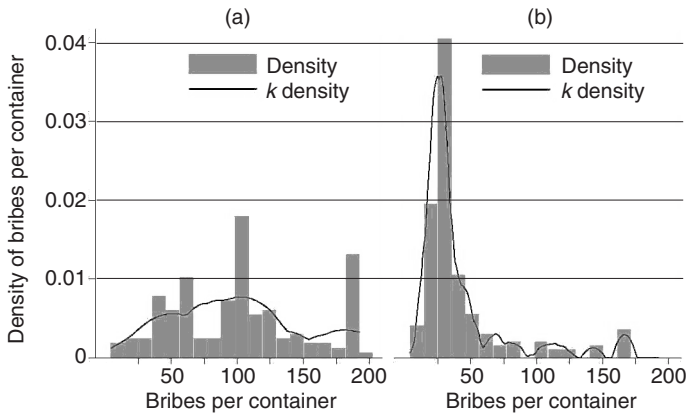
Figure 5.17: *Distribution of bribes per container at the ports of Durban and Maputo. (a) Maputo; (b) Durban.*
Source: Sequeira and Djankov (2010). Note: coefficient of variation in Maputo was higher (2) than in Durban (1.5).

Figure 5.18: *South African firms' choice of port for imports.*

The 'diversion costs' of corruption for each individual firm were on average four times higher than the actual bribe collected by the customs' official in Maputo. This also suggested that firms exhibited considerable aversion to the uncertainty associated with bribe payments in Maputo.[41] This aversion was reported in survey data as well. Sequeira and Djankov (2010) argue that,

---

[41] As observed in Figure 5.17, the coefficient of variation of the distribution of bribes was much higher in Maputo (2) than in Durban (1.5).

in an environment of higher and more unpredictable bribes, the asymmetry of information between firms, clearing agents and port officials about bribe payments became more salient, making firms less able, and less willing, to guarantee the required liquidity of clearing agents (Shleifer and Vishny 1993; Bardhan 1997; Campos *et al* 1999; Hallward-Driemeier *et al* 2010).

This study also suggests that the full set of costs imposed by corruption in the soft infrastructure of transport networks goes well beyond the rerouting costs of a given firm. Every time a firm rerouted its imports away from the most corrupt port, it imposed a negative externality on other firms. On the one hand it added to congestion in the least corrupt port; on the other hand, it contributed to fewer and more imbalanced cargo flows to the more corrupt one, resulting in higher overall transport costs through both corridors. On the Maputo corridor imports were more vulnerable to paying high bribes than were exports, resulting in more outbound than inbound cargo.

In the trucking survey, Sequeira and Djankov (2010) observed that even though the actual costs of operating in either corridor were almost identical for all trucking companies, the absence of a regular flow of outbound cargo along the Maputo corridor resulted in a 70% increase in transport rates charged to firms on that route. Though this difference cannot be solely attributed to the 'diversion' effects of corruption proposed in Sequeira and Djankov (2010), the pattern of bribe payments in Maputo and its effect on South African firms' demand for the port and corridor is likely to play an important role in this result. These effects can be magnified across the region given that the South African and Mozambican transport networks also serve six landlocked and neighbouring countries in Southern Africa: Malawi, Lesotho, Swaziland, Botswana, Zambia and Zimbabwe.

The study identifies a second type of distortion introduced by corruption in ports on firm behaviour: the decision to source inputs domestically or internationally. Decisions on the sourcing of inputs matter not only because they affect the productivity of the firm itself, but also because they have significant spillover effects in the whole economy by affecting the nature and the extent of backward and forward linkages between firms. The intuition is straightforward: all else being equal, if corruption increases the cost of using a port, firms should have an incentive to decrease demand for imported inputs, while the opposite should happen when corruption decreases port costs.

The study exploits the fact that, on average, South African firms going through the port of Maputo or Durban only face coercive corruption, while Mozambican firms only face collusive corruption.[42] Controlling for several firm- and product-level characteristics, the study finds that coercive and col-

---

[42]In Durban, South African firms pay coercive bribes to prevent cargo from being moved around the docks, while in Maputo they pay coercive bribes to avoid the arbitrary increase in the transit bond. Mozambican firms, on the other hand, face collusive corruption in Maputo that allows them to evade tariffs.

lusive corruption have the opposite effect on firms' sourcing decisions. A one-standard-deviation increase in expected bribes is associated with a 3% (0.089 standard deviations) increase in the proportion of inputs sourced domestically for firms facing coercive corruption and a 7% (0.173 standard deviations) decline in the proportion of inputs sourced domestically for firms facing collusive corruption. Collusive corruption reduces the relative cost of imported inputs, while coercive corruption increases it. Thus, firms in Mozambique that can pay bribes to evade tariffs are less likely to source their inputs domestically, whereas in South Africa, where bribes are paid coercively, providing no rent to the shipper, firms are more likely to source domestically.

The empirical set-up prevents the identification of a causal effect between corruption at ports and firms' sourcing decisions, but these results are suggestive of how firms may respond to different types of corruption by organising production in ways that increase or decrease demand for the public service.

A natural extension to this research agenda would discuss the effectiveness of different anti-corruption policies in reducing different types of corruption. With some notable exceptions, empirical evidence on the relative success of different anti-corruption policies is scarce. Both top-down (Olken 2007) and third-party (Yang 2008) monitoring have been suggested as effective tools to reduce corruption. The effectiveness of other tools discussed in the theoretical literature—such as increased wages (Becker and Stigler 1974), technology, and changes in rules and regulations—has so far received more limited empirical validation. In recent years, a growing literature has suggested that the simplification of tariff codes and the introduction of flat tariff rates could reduce opportunities for corruption in customs (Fisman and Wei 2007).

However, the net welfare effects from such policies are unclear, as customs officials may respond by extracting bribes through other means. In an extension of the project described above, Sequeira (2011) tests for this possibility by exploiting a quasi-experiment taking place in Mozambique to estimate the impact of a reduction in tariffs on corruption.

Mozambique is one of the founding members of the Southern African Development Community (SADC), and in 1980 it agreed to a gradual phase-out of all tariffs with neighbouring South Africa. This meant that, in the years in which a phase-out was scheduled (2008 and 2010), tariffs decreased for a predetermined set of products but not for others. This created a natural variation in product tariff level that is independent of the level of bribes being paid for each product.

Using the same methodology as in Sequeira and Djankov (2010), we collect data on bribe payments before and after each change in the tariff code took place, and then apply a straightforward triple difference technique to estimate the impact of changes in tariff levels on the level and distribution of bribes paid by firms. Since we observe the entire chain of clearance at the port of Maputo, we can test if corruption is displaced onto other phases of the shipping process. This research project is currently in the field.

4 CONCLUSIONS

In this chapter we have discussed two research projects that attempt to identify the impact of different types of transport costs on firms. The first project exploits a natural experiment with the rebuilding of a railway in Southern Africa to estimate the impact of a reduction of transport costs on firm performance. The set-up allows for the identification and monitoring of important spillover and network effects between rail and road transport. To the best of our knowledge, this represents the first study of the impact of railways on firms in a contemporary setting, using primary firm-level data from before and after the investments take place.

The second project focuses on the impact of the soft infrastructure of transport on firms by looking at the specific context of corruption in ports. This study collected original data on bribe payments by infiltrating observers into two competing ports, traced differences in bribe schedules to the organisational structure of each port and measured the impact of corruption at ports on firm behaviour by observing how firms adapt their shipping and sourcing decisions to the type of corruption they face at each port. Both shipping and sourcing decisions are likely to be directly affected by the overall cost of using each port.

The preferred estimate of the distortion imposed by corruption is what the study labels the 'diversion' effect of corruption, when firms travel on average an additional 322 km—more than doubling their transport costs—just to avoid 'coercive' corruption at a port. This effect is only observed for firms facing a higher probability of being coerced into a bribe due to the type of product they ship. Interestingly, the cost for a firm to reroute can be up to four times higher than the cost of the actual bribe requested. Survey data revealed that firms were willing to incur the higher costs to avoid corruption because of an extreme aversion to the uncertainty surrounding bribe payments at the most corrupt port. The uncertainty in Maputo appears to be linked to the short time horizons of customs officials because of high job turnovers. This result warrants further investigation and perhaps even explicit testing through experimental methods.

Given that corruption at ports directly affects the relative cost of imports by increasing the cost of using the port, firms also seem to respond to different types of corruption by adjusting their decisions on the sourcing of inputs. The study finds a correlation between cost-reducing 'collusive' corruption and increased usage of the port to import inputs—and between cost-increasing 'coercive' corruption and a higher proportion of domestic sourcing. While the study falls short of identifying a causal relationship, these correlations suggest how firms may respond to different types of corruption by organising production in ways that increase or decrease demand for the public service.

By observing the entire chain between competing port bureaucracies, front-line bureaucrats setting bribes and users making shipping and sourcing decisions, the study provides a full account of the determinants and consequences of corruption in ports and the many indirect and direct ways in which corruption can affect the economy. Depending on the type of corruption bureaucrats engage in, bribes can generate more or less deadweight loss and reduce tariff revenue for the government, but also increase or decrease demand for the public service, with important implications for economic activity.

In the Southern African corridors, corruption seems to be significantly reducing demand for the Maputo port, thereby stifling the returns to the massive investments in the hard infrastructure of the corridor that have taken place in recent years (Macchi and Sequeira 2009). The studies also suggest important policy changes to the organisation of ports and to the nature of the interaction between shippers and port officials that hold potential to reduce corruption, such as reducing the discretion of port officials in the clearance process, and eliminating face-to-face interactions between clearing agents and port officials.

These studies also suggest that more empirical research is needed to understand the impact of other types of transport infrastructure—soft, hard and the interaction between them—on firm behaviour. Longitudinal studies could also help clarify the dynamic effects of transport on important dimensions like firm size, specialisation, location and trade, among others, and their implications for long-run economic growth. This research would not only provide fundamental insights into the role of transport in determining the micro-foundations of growth, but also provide invaluable data for effective policymaking, suggesting how investments in different types of transport infrastructure, soft and hard, could be prioritised to provide maximum returns for economic growth.

Sandra Sequeira is a Lecturer in Development Economics at the London School of Economics.

REFERENCES

Abadie, A., and G. Imbens (2006), Large sample properties of matching estimators for average treatment effects. *Econometrica* **74**, 235–267.

Abadie, A., A. Diamond, and J. Hainmueller (2007). Synthetic control methods for comparative case studies: estimating the effect of California's tobacco control program. NBER Working Paper 12831.

Anderson, J., and E. Wincoop (2004). Trade costs. *Journal of Economic Literature* **42**, 691–751.

Angrist, J., and J. Hahn (2004). When to control for covariates? Panel asymptotics for estimates of treatment effects. *Review of Economics and Statistics* **86**(1), 58–72.

Bardhan, P. (1997). Corruption and development: a review of issues. *Journal of Economic Literature* **35**, 1320–1321.

Becker, G., and G. Stigler (1974). Law enforcement, malfeasance and compensation of enforcers. *Journal of Legal Studies* **3**, 1–18.

Bertrand, M., S. Djankov, R. Hanna, and S. Mullainathan (2007). Obtaining a driver's license in India: an experimental approach to studying corruption. *Quarterly Journal of Economics* **122**, 1639–1676.

Buys, P., U. Deichmann, and D. Wheeler (2006). Road network upgrading and overland trade expansion in sub-Saharan Africa. World Bank Policy Research Working Paper 4097.

Campante, F., D. Chor, and Q. Do (2009). Instability and the incentives for corruption. *Economics and Politics* **21**, 42–92.

Campos, J. Edgardo, D. Lien, and S. Pradhan (1999). The impact of corruption on investment: predictability matters. *World Development* **27**, 1059–1067.

Clark, X., D. Dollar, and A. Micco (2004). Port efficiency, maritime transport costs and bilateral trade. NBER Working Paper W10353.

Crump, R., V. Hotz, G. Imbens, and O. Mitnik (2009). Dealing with limited overlap in estimation of average treatment effects. *Biometrika* **96**(1), 187–199.

CSIR (2005). Study of logistics and transport costs in South Africa. http://www.csir.co.za/sol.

CSIR (2006). Study of logistics and transport costs in South Africa. http://www.csir.co.za/sol.

Dehejia, R., and S. Wahba (2002). Propensity score-matching methods for nonexperimental causal studies. *Review of Economics and Statistics* **84**(1), 151–161.

Djankov, S., C. Freund, and C. S. Pham (2010). Trading on time. *The Review of Economics and Statistics* **92**, 166–173.

Donaldson, D. (2010). Railroads of the Raj: estimating the impact of transportation infrastructure. Mimeo, MIT, Cambridge, MA.

Estache, A. (2010). A survey of impact evaluation of infrastructure projects, programs and policies. Ecore Discussion Paper.

Feinstein, C. (2005). *An Economic History of South Africa: Conquest, Discrimination and Development.* Cambridge University Press.

Fishlow, A. (1965). *American Railroads and the Transformation of the Antebellum Economy.* Harvard University Press.

Fisman, R., and S. Wei (2007). Tax rates and tax evasion: evidence from missing imports in China. *Journal of Political Economy* **112**, 471–500.

Fisman, R., and J. Svensson (2007). Are corruption and taxation really harmful to growth? Firm-level evidence. *Journal of Development Economics* **83**, 63–75.

Fogel, R. (1964). *Railroads and American Economic Growth.* Baltimore, MA: Johns Hopkins Press.

Frankel, J., and D. Romer (1999). Does trade cause growth? *The American Economic Review* **89**, 379–399.

Freund, C., and N. Rocha (2010). What constrains Africa's exports? World Bank Policy Research Working Paper 5184.

Galiani, S. (2007). Notes for the impact evaluation of the construction and rehabilitation of rural roads. World Bank.

Gumede, W. (2005). *Thabo Mbeki and the Battle for the Soul of the ANC.* Cape Town: Zebra Press.

Hallward-Driemeier, M., G. Khun Jush, and L. Pritchett (2010). Deals versus rules: policy implementation uncertainty and why firms hate it. NBER Working Paper 16001.

Hirschman, A. (1967). *Development Projects Observed.* Washington, DC: Brookings Institution Press.

Hoekman, B., and A. Nicita (2008). Trade policy, trade costs and developing country trade. World Bank Policy Research Working Paper 4797.

Hummels, D. (1999). Toward a geography of trade costs. Mimeo, Purdue University.

Hummels, D. (2007). Transportation costs and international trade over time. *Journal of Economic Perspectives* **21**, 131–154.

Huntington, S. P. (1968). *Modernization and Corruption: Political Order in Changing Societies*, pp 59–71. New Haven, CT: Yale University Press.

Krueger, A. O. (1974). The political economy of the rent-seeking society. *American Economic Review* **64**, 291–303.

Leff, N. H. (1964). Economic development through bureaucratic corruption. *American Behavioral Scientist* **8**(3), 8–14.

Limão, N., and A. Venables (2001). Infrastructure, geographical disadvantage, transport costs, and trade. *World Bank Economic Review* **15**(3), 451–79.

Lui, F. (1985). An equilibrium queuing model of bribery. *Journal of Political Economy* **93**, 760–781.

Macchi, P., and S. Sequeira (2009). L'infrastructure de transport 'immaterielle' des ports de Maputo et Durban. *Afrique Contemporaine* **230**, 69–82.

Obstfeld, M., and K. Rogoff (2000). The six major puzzles in international macroeconomics: is there a common cause? *NBER Macroeconomics Annual* **15**, 339–390.

Olken, B. A. (2007). Monitoring corruption: evidence from a field experiment in Indonesia. *Journal of Political Economy* **115**, 200–249.

Olken, B. A. (2009). Corruption perceptions vs corruption reality. *Journal of Public Economics* **93**, 950–964.

Olken, B. A., and P. Barron (2007). The simple economics of extortion: evidence from trucking in Aceh. *Journal of Political Economy* **117**, 417–452.

Raballand, G. and P. Macchi (2008). Transport prices and costs: the need to revisit donors' policies in transport in Africa. Bureau for Research & Economic Analysis of Development Working Paper 190.

Rauch, J. (1999). Networks versus markets in international trade. *Journal of International Economics* **48**, 7–35.

Robinson, J., and R. Torvik (2005). White elephants. *Journal of Public Economics* **89**, 197–210.

Rodriguez, F., and D. Rodrik (2000). Trade policy and economic growth: a skeptic's guide to the cross-national evidence. *NBER Macroeconomics Annual* **15**, 261–325.

Rosenbaum, P. R., and D. Rubin (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika* **70**, 41–55.

Rostow, W. W. (1959). The stages of economic growth. *The Economic History Review* **12**, 1–16.

Sequeira, S. (2011). Tariffs and collusive corruption. Mimeo, London School of Economics.

Sequeira, S., and S. Djankov (2010). An empirical study of corruption at ports. Mimeo, London School of Economics.

Schelling, T. (1956). An essay on bargaining. *The American Economic Review* **46**, 281–306.

Shleifer, A., and R. W. Vishny (1992). Pervasive shortages under socialism. *RAND Journal of Economics* **23**, 237–246.

Shleifer, A., and R. W. Vishny (1993). Corruption. *Quarterly Journal of Economics* **108**, 599–617.

Stigler, G. (1974). A theory of oligopoly. *Journal of Political Economy* **72**, 44–61.

Svensson, J. (2003). Who must pay bribes and how much? Evidence from a cross-section of Firms. *Quarterly Journal of Economics* **118**, 207–230.

Van de Walle, D. (2002). Choosing rural roads investments to help reduce poverty. *World Development* **30**, 575–589.

World Bank (2007). Doing business: how to reform. Report.

Yang, D. (2008). Integrity for hire: an analysis of a widespread customs reform. *Journal of Law and Economics* **51**, 25–57.

6

Half-Baked Interventions: Staggered Pre-Shipment Inspections in the Philippines and Colombia

MOHINI DATT AND DEAN YANG[1]

1 INTRODUCTION

Corruption is an endemic problem that is especially prevalent in developing countries. The impact of corruption on welfare has been a widely debated topic among economists: much has been said on the second-best option that bribery may present in the face of a rigid and inefficient bureaucracy. Notwithstanding the arguments on the allocative efficiency that corruption might unwittingly lead to, citizens of developing countries who live with the realities of corruption would undoubtedly concur that corruption is a nuisance that needs to be rooted out.[2]

Systematic empirical evidence evaluating the effectiveness of policies designed to combat corruption is scarce. The seminal theoretical work of Becker and Stigler (1974) identified a pair of generic remedies for bureaucratic corruption in government: increased monitoring and higher wages.[3]

But for many reasons anti-corruption reforms may fail. For example, a reform that increases monitoring of potentially corrupt officials might fail if the monitors are corrupt and thus provide inaccurate information to higher authorities. In turn, the higher-level officials may also be corrupt and not put the information gathered to good use; or the monitoring programme may simply be implemented to demonstrate the government's anti-corruption credentials What is more, even if enforcers are honest, corrupt officials may find

---

[1]For econometric and analytical details of the study, the reader is referred to Yang (2008a) and the working paper version, Yang (2005).

[2]See Huntington (1968) and Myrdal (1968). For overviews of the relationship between corruption and development, and the impact of corruption on welfare, see, for example, Bardhan (1997), Rose-Ackerman (2004) and Sequeira and Djankov (2010).

[3]Other contributions include Besley and McLaren (1993), Mookherjee and Png (1992, 1995) and Polinsky and Shavell (2001).

alternative methods of continuing their corrupt dealings. Therefore, empirical work—particularly at the micro level—is necessary to determine the effectiveness of any given anti-corruption effort.

A recent empirical literature sheds light on the impact of monitoring on bureaucratic corruption. In a variety of contexts these studies show that increased monitoring—and transparency—plays a positive role in reducing corruption or cheating.[4] Di Tella and Schargrodsky (2003) examine the impact of increased enforcement on corruption in hospital procurement in Argentina. In Uganda, Reinikka and Svensson (2004) find that diversion of government funds intended for education is reduced when intended funding levels are publicised in newspapers. Nagin *et al* (2002) use a field experiment to document the impact of increased monitoring on the opportunistic behaviour of telephone call centre employees in the USA.

Examining corruption in trade is important, given trade's potential to act as an engine of growth. Smuggling[5] and customs corruption are rife in many countries, and trade taxes have traditionally been an important source of government revenue. In 1995, import duties accounted for an average of 23% of central government tax revenue among non-OECD developing countries, but the fraction was as high as 31% in the Philippines, 32% in Côte d'Ivoire and 37% in the Dominican Republic (World Bank 2002). Since import tariffs have usually not been totally dismantled in developing countries, customs corruption in these countries often takes the following form: an importer under-reports the value of an incoming shipment in order to pay lower import duties (a practice known as under-invoicing), and customs officials may agree to overlook the fraudulent declaration in return for bribes.

Several papers have examined the avoidance of taxes on international trade. Pritchett and Sethi (1994) find that collected import duties as a share of import value rise less than one-for-one with the tariff rate, and interpret this as evidence of tax evasion or avoidance. Fisman and Wei (2004) show that the extent of import under-invoicing rises as the tariff rate rises for Chinese imports from Hong Kong. Bernard and Weiner (1990), Hines and Rice (1994) and Clausing (2003) examine tax-induced transfer pricing within multinational firms.

The Philippine and Colombian customs reforms examined in Yang (2006, 2008a,b), and reviewed in this chapter, reveal are of general interest as examples of a common approach to fighting smuggling and customs corruption: hiring private firms to conduct pre-shipment inspection (PSI) of imports. Many developing country governments use PSI as a solution to fight corruption in customs. When a government implements a PSI programme, foreign inspectors verify the tariff classification and value of individual incoming shipments

---

[4]These results are aligned with those of Becker and Stigler (1974).

[5]We use the term 'smuggling' here to refer to any illegal means by which an importer reduces duties paid on a shipment, whether the shipment passes through formal customs channels or not.

before they leave their country of origin. PSI simply improves the information available to higher-level enforcers, who can use the reports to hold individual customs officials accountable for collecting the correct amount of duties on shipments. PSI reports may improve the bargaining power of importers against customs officials seeking bribes, which could facilitate trade and raise the total amount of taxable import activity.

While the aforementioned research finds evidence of tax avoidance and evasion in the context of international trade, few studies examine the impact of enforcement on these activities. In nearly all implemented PSI programmes, the responsibility for collecting customs duties remains in the hands of the importing country's customs officials, who may choose to ignore PSI reports on specific shipments. PSI programmes can fail to reduce import duty evasion if enforcers fail to use the new information available to them in the PSI reports, or if importers can find alternative methods to avoid duties (that may involve avoiding PSI, or bringing shipments into the country outside of formal customs channels so that shipments never appear before a customs officer).

Sequeira and Djankov (2010) lament the current lack of research on the mechanics of bribe payments and how it may affect firm behaviour. If smuggling is displaced, where does it go? Empirical analysis on the impact of enforcement addresses displacement as an aside, at most examining the existence or amount of displacement.[6] In this chapter we review evidence on the alternative routes to duty avoidance taken by importers in the Philippines and in Colombia, when enforcement of customs corruption increases.

Studying the impact of enforcement in customs is doubly important for the public finances of nearly all developing countries: on the one hand smuggling drains customs revenue; on the other hand pre-shipment inspections can be expensive. Anson *et al* (2006) point out that at 1% of the value of inspected shipments, and given that customs duty is usually below 10% of import value, PSI fees could come to more than 10% of customs revenue. Ramírez Acuna (1992) and Byrne (1995) evaluate the arguments justifying the use of PSIs, and Goorman and De Wulf (2003) discuss the difficulties encountered by the WTO Agreement on Customs Valuation in developing countries.

Yang (2008b) examines the aggregate, country-level impact of pre-shipment inspection services worldwide, finding that countries implementing PSI programmes subsequently experience large increases in the growth rate of import duties. The study uses panel data at the country level to examine the relationship between the implementation of PSI programmes and import duty collections between 1980 and 2000. After the implementation of PSI programmes, import duties increase on average by 15–30 percentage points. Additional evidence suggests that reductions in corruption are the cause of the import duty improvements: PSI programmes are accompanied by declines in under-invoicing and in misreporting of goods classifications in customs.

---

[6]See Hesseling (1994) for a review.

From a cross-country perspective, pre-shipment inspections appear to be cost-effective: improvements in import duties in the first five years after programme implementation were two-to-three times larger than programme costs.

However, microeconomic studies of the effects of PSI suggest an ambiguous result. Anson *et al* (2006) examine how the introduction of PSI affects incentives for under-invoicing and collusive behaviour between importers and customs. They find that, with the introduction of PSI, fraud increased in Argentina and Indonesia but decreased, albeit not significantly, in the Philippines. The effect of PSI was found to be ambiguous, if not perverse, since the results from the PSI were not corroborated by the government against what customs reported, and the implementation of the PSI in the first place de-motivated the customs administration.

Success with PSI is therefore not guaranteed. But examining situations where PSI programmes failed to produce the desired results can shed light on the conditions under which such programmes are likely to succeed. In this chapter we review Yang (2008a), which was a microeconomic empirical study of PSI programmes in the Philippines, as well as Yang (2006), which conducted a similar analysis for the PSI programme in Colombia. We focus here on the way that the customs reforms in both countries lent themselves to analysis as natural experiments.

Experimental and quasi-experimental methodologies have rarely been used to evaluate policy interventions to combat corruption. Olken (2007) implemented a field experiment examining how different types of monitoring affect corruption in Indonesian road projects, finding that 'top-down' was more effective than 'bottom-up' monitoring. Bertrand *et al* (2007) carried out a randomised controlled trial on the allocation of driver's licences in India and the likelihood of resorting to corruption to obtain the licence.

Sequeira and Djankov (2010) exploited a natural experiment within the wider context of their study of corruption in ports in Southern Africa—specifically, the port of Maputo in Mozambique and the port of Durban in South Africa, both of which lead, via two competing transport corridors, to the industrial, agricultural and mining heartland of South Africa. They modelled their difference-in-difference estimation around the event of a 20-percentage-point reduction in tariffs in Mozambique on a selected category of goods. This change affected cargo entering through Maputo in transit to South Africa because it affected the size of the refundable transit bond that South African firms were forced to pay as insurance against the possibility of cargo being diverted and remaining in Mozambique. The valuation of the transit bond was said to be exploited by customs in Maputo seeking to extract bribes. The impact evaluation of Sequeira and Djankov estimated the probability of bribe-paying on goods that received the tariff reduction and those that did not, before and after the reduction took place. The study found that high-tariff

goods are more likely to lead to a 'collusive' form of corruption in the port of Maputo.

For the purposes of impact evaluation, a standard problem when assessing the impact of increased enforcement is establishing a credible counterfactual, whereby the causal impact of an enforcement policy can be separated from time trends in crime rates, from the impact of other simultaneous enforcement efforts, or from reverse causation (higher crime rates leading to more enforcement). To establish the causal impact of increased enforcement, Yang (2008a) takes advantage of the fact that, for the 28-month period of interest, the Philippines' PSI programme covered imports from only a subset of countries. Then, during the course of the study period, the Philippine government increased enforcement against duty avoidance for imports from PSI-covered countries by expanding inspections to a previously exempt shipment category (shipments below a certain dollar value).

The key element of the identification strategy is the existence of a plausible control category—imports from countries not covered by the Philippine PSI programme—against which changes in imports from PSI countries can be compared. Because changes over time in a treatment group are compared with corresponding changes in a control group, this difference-in-difference identification strategy purges the empirical estimates of time trends in the use of various duty-avoidance methods, as well as general changes in customs enforcement that should affect imports from all countries equally (such as changes in legal sanctions against smugglers, salaries of customs officials, etc).

Yang (2008a) finds that, when the Philippine government increased enforcement by expanding inspections to low-value shipments, imports from treatment countries shifted differentially to an alternative duty-avoidance method: shipping via duty-exempt EPZs. Thus, increased enforcement reduced the targeted method of duty avoidance but led to substantial displacement to an alternative duty-avoidance method, amounting to 2.7% of total imports from treatment countries. The hypothesis that the reform led to zero change in total duty avoidance cannot be rejected.

Also, displacement was greater for products with higher tariff rates and import volumes, consistent with the existence of fixed costs of switching to alternative duty-avoidance methods. The revealed preferences of importers in the Philippines had shown that, as long as the minimum value threshold was US$5,000, they continued to prefer using the regular import channels rather than straightaway routing shipments via EPZs when the government first introduced PSI. Once the minimum value threshold loophole was closed, they were compelled to ship via EPZs and incur the fixed costs associated with that switch. Thus, society as a whole was worse off: government revenue did not increase and the importers suffered higher costs.

The analysis finds that when the increase in enforcement is only partial— in that it only addresses a subset of potential methods of avoiding import

duties—there can then be substantial displacement to alternative methods of avoiding import duties. The PSI experience in the Philippines and in Colombia suggest that, to be successful, anti-corruption reforms should be 'broad' in the sense of encompassing a wide range of possible alternative methods of committing the illegal activity of interest; otherwise, displacement to alternative methods can negate the original goals of the reform. Moreover, the displacement bears out the second-best theorem, in that addressing only one distortion of two arguably makes things worse in the continued presence of the second distortion.

The remainder of this chapter is organised as follows. In Section 2 we describe the pre-shipment inspection reform in the Philippines studied in Yang (2008a) and presents an analytical model. In Section 3 we discuss the data, estimation issues and main findings, and go on to examine the Colombian pre-shipment inspection reform. We give our conclusions in Section 4.

2   THE PRE-SHIPMENT INSPECTION REFORM IN THE PHILIPPINES

The government of the Philippines maintained a PSI programme from 1987 to 2000. Over the 28-month period that is the focus of Yang's (2008a) study (November 1989 to February 1992), the Philippines government required PSI on shipments from only a subset of countries rather than from all its trading partners. PSI was required on imports from nine countries in Asia: Brunei, Hong Kong, Japan, Indonesia, Malaysia, Singapore, South Korea, Taiwan and Thailand. After February 1992, the government decided to extend PSI cover to shipments from all its import partners. The PSI programme was initially implemented in a staggered fashion. In the beginning, PSI would only be required if shipments from these countries had a declared FOB[7] value greater than US$5,000. It is often the case that governments require PSI on the goods of only certain trade partners, or of only certain value, or on only a certain categorisation of goods, given that if the costs of implementing a PSI programme are greater than anticipated benefits, PSI is not cost effective.

However, the fallout from such a partial or staggered PSI programme is that tax evaders could exploit the situation. In the case of the Philippines, importers did seek to take advantage of the US$5,000 threshold to avoid PSI by simply declaring shipment values below that threshold. Anecdotal evidence showed that what importers typically did was to split shipments into batches that could reasonably be declared as valued below US$5,000. In addition, single shipments that were above the threshold in actual value were simply under-invoiced to fall below the threshold. Throughout the year 1990, the government took steps to remove this exemption. On 2 May 1990, it lowered

---

[7] The FOB or 'free-on-board' value of a shipment is its value excluding the costs of freight and insurance.

the minimum value threshold to US$2,500. Six months later, on 1 November 1990, it lowered the threshold still further, to US$500.

Essentially, the Philippine government was targeting duty avoidance. The Yang (2008a) study examined the likelihood of importers finding alternative methods of duty avoidance. Importers had two options to avoid duty once the government started to mandate PSI.

- To value shipments below the minimum value threshold and so be exempt from PSI.
- To route imports via an export processing zone where PSI was not required. EPZs are especially designed geographic zones where no trade barriers exist, *ie* goods are not subject to import tariffs or duties. As the EPZ is designed to be free of all trade barriers, customs inspections are often perfunctory. Manufacturing firms obtain government licences to relocate to EPZs, where they can freely import raw materials and export finished goods. As the purpose of EPZs is to encourage foreign investment, imports required by enterprises in the zones are exempted from PSI.

2.1 Analytical Model

Given the staggered fashion in which the Philippine government deployed its PSI programme, the policy intervention lent itself to an impact-evaluation analysis. The reform constituted a quasi-experiment because the increased enforcement applied only to shipments from a subset of countries (the nine mentioned earlier), which form the 'treatment group', with shipments from all other trading partners serving as a 'control group'. November 1989 is taken as the start of the study period, because six of the treatment countries were newly added to the PSI programme in the previous month.[8] The end of the study period was February 1992, as the PSI programme was extended to cover all countries in March 1992, thus doing away with a control group.[9] The Philippine policy intervention to lower the minimum value threshold further lent itself to difference-in-difference quasi-experimental analysis with the earlier minimum value threshold for PSI inspection of US$5,000 constituting a 'before' period (November 1989 to April 1990), and the lowering of the

---

[8]As Yang (2008a) details, the countries added to the programme in October 1989 were Brunei, Indonesia, Malaysia, Singapore, South Korea and Thailand.

[9]Table 1 in the Appendix of Yang (2005)—the working paper version of Yang (2008a)—provides a detailed description of changes in country coverage and minimum value thresholds for the Philippine PSI programme.

threshold to US$500 (December 1990–February 1992) constituting an 'after' period.[10]

Though the government phased down the minimum value threshold gradually over a six month period from May to November 1990, first lowering the threshold to US$2500 and then to US$500, for the purpose of an impact-evaluation study, the focus is on just two time periods—when the threshold was US$5,000 and when it was lowered to US$500. Intuitively this makes sense. The two-time-period focus is aligned with the purpose of the reform, which was to clamp down on the loophole that allowed importers to evade inspection by splitting shipments into smaller value chunks, and to examine what happened when there was an easy loophole (the 'before' period), and then when there was no longer an easy loophole (the 'after' period). The point of interest here is the change in import behaviour once the minimum value threshold is lowered.

One possible concern addressed at the outset is selection bias: were the goods of these nine countries chosen to be subjected to PSIs because it was *expected* that imports from these countries would exhibit the observed changes in duty-avoidance methods that we end up seeing? In other words, would those changes have occurred anyway, even without the increased enforcement? Would those imports have in any case been rerouted via EPZs? This possibility is difficult to rule out directly; but an indirect test suggests that this may not be the case. There is no evidence of differential trends in the use of the different duty-avoidance methods immediately prior to the policy change in May 1990 when the government started to lower the minimum value threshold.[11]

The data on individual shipments and their characteristics used in Yang (2008a) were obtained from the National Statistics Office of the Philippines. The data set contains information on each shipment that entered the country via formal customs procedures, and had hitherto only been used to calculate aggregate Philippine import statistics.

This study's purpose was not simply to look at changes in the aggregate sum of collected import duties, for which data are freely available in the World Bank's annually collected *World Development Indicators*. Instead, it intended to examine changes in the behaviour of importers in order to give more

---

[10]Changes in PSI rules applied for shipments whose letters of credit were opened after the rule change, regardless of whether the arrival date of that shipment in the Philippines was some weeks later. Therefore, shipments arriving during the month when the threshold was first lowered to US$500 may not yet have been subject to the new PSI rules. November 1990 is therefore excluded from the 'after' period, even though that was when the threshold was lowered to US$500.

[11]Conducting a 'false experiment', where the 'before' period is November 1988–April 1989 and the 'after' period is November 1989–April 1990, yields a coefficient of interest that is small and not statistically significant. It appears that there is no shift in imports to alternative methods of duty avoidance. See Yang (2008a) for details.

nuanced policy recommendations: not just to state that the policy failed, but to examine what piece of the puzzle made it fail. The study aimed to understand the 'mechanics' of paying bribes and how that affects firm behaviour.[12]

The empirical analysis therefore gauges the extent to which shipments from PSI countries shift to alternative methods of duty avoidance in response to the reduction in the minimum value threshold. It tracks changes in the frequency of the following shipment types from PSI (treatment group) and non-PSI (control group) countries:

1. shipments valued between US$5,000 and US$500;
2. shipments valued under US$500;
3. shipments destined for export processing zones.

The study used the reported declarations of the FOB value of each shipment (reported in nominal US dollars) to determine whether a shipment fell into type 1 or 2 listed above. The data set also indicates whether a shipment was destined for an export processing zone, which is then our type 3.[13] The dependent variable in the main regressions is the main outcome variable of interest defined as the probability of shipments falling under one of the three types: shipments valued $US\$500 < j < US\$5,000$, shipments valued less than US$500, shipments destined for EPZs.[14]

Yang (2008a) outlines a stylised model of the impact of enforcement on criminal activity such as smuggling, when enforcement only targets a subset of methods by which a particular crime can be carried out. According to that model, when alternative 'lawbreaking' methods (which in the Philippines PSI example would be the action of relocating to EPZs)[15] involve fixed costs of entry, crime displacement should respond positively to the size of illicit profits threatened by enforcement, with the extent of displacement depending on the relative fixed and variable costs of the different methods.

---

[12]See Sequeira and Djankov (2010) for a detailed study on the mechanics of paying bribes.

[13]The analysis tracks displacement from the first shipment type, valued between US$5,000 and $500, to the other two types on the list, valued under US$500 or destined for EPZs. It is important to avoid double-counting of shipments: valued as type 1 or type 2 and destined for EPZs (type 3). Luckily, very few shipments to export processing zones also fall into one of the type 1 or 2 categories (0.02% of imports by value). Hence, we allocate these overlapping shipments to type 1 (valued between US$5,000 and US$500) or type 2 (valued under US$500). The results are not sensitive to allocating these shipments to the export processing zone type instead.

[14]If other unobservable methods of avoiding duty are important, that would understate the measure of total displacement.

[15]While, technically, relocating to EPZs is legal, here we refer to it as an 'alternative lawbreaking' method to the extent that it moves from illegal duty avoidance via undervaluation to a kind of legal tax avoidance in moving to EPZs, *ie* importers do not simply start to report shipment values honestly and then pay the required duty; they are still resorting to avoidance of the standard route.

Undervaluation as a means to avoid PSI on imports involved low (though not zero) fixed costs but significant variable costs. Importers would have had to persuade their suppliers overseas to split shipments into lower-valued chunks, and might even have had to source out suppliers willing to engage in this subterfuge, *ie* these would be constraints with some but not substantial fixed costs to the importers. But the many different lots of lower-value shipments would have increased the costs of per-unit packaging and documentation. Also, importers would have had to plan for, and likely even face, the possibility of a corrupt and alert customs official noticing the increased number of shipments and hence the undervaluation.

In contrast, the relocation to an EPZ as a route for avoiding PSI on imports is likely to have had high fixed costs and low variable costs. The high fixed costs would have been in obtaining a government licence, the physical relocation costs and the maintenance of a continued presence in the EPZ, either via their own outlets or via firms willing to import on their behalf. But once this system was in place, variable costs would have been negligible: shipments could enter in any size and valuation, duty-free and largely free of harassment by customs officials who would have had fewer tools to leverage in a government-mandated free trade zone.

As predicted by this model, displacement in both the Philippines and Colombia would be greater when the size of profits from crime displacement were higher. In the case of the Philippines, the tariff rate and import volume at the product group level proxy for the size of profits from crime displacement.

3   ESTIMATION AND REGRESSION RESULTS

The Philippines study estimates a separate linear probability model for each of the three shipment types outlined above, and also does so for a fourth type: *any of the three shipment types that allows avoidance of pre-shipment inspection.* To reiterate, the 'before' period is November 1989 to April 1990, when the minimum value threshold for PSIs was US$5,000; and the 'after' period is December 1990 to February 1992, when the threshold was US$500.

The model examines the probability that imports from country g and product group h enter the Philippines as shipment type j during period p, denoted P_ghp^j. T_{gh} is the treatment group and P_p is the treatment period (post reform). The simplest difference-in-difference estimate of the impact of the treatment (the reduction of the minimum value threshold) on the probability that a shipment enters as shipment type j is estimated using the following regression equation:

$$P_{ghp}^j = \beta_0^j + \beta_1^j T_{gh} P_p + \beta_2^j T_{gh} + \beta_3^j P_p + u_{ghp}^j. \qquad (6.1)$$

In this difference-in-difference regression, the coefficient on the interaction term between the treatment group and the treatment period (β_1^j) is the esti-

mate of the treatment effect. It captures for shipments from treatment countries the differential change in the likelihood of observing shipment type j in response to a change in enforcement (the reduction of the minimum value threshold). The identification assumption is that, without the minimum value threshold reduction, changes in the observed frequency of particular shipment types over time would have been the same for imports from treatment and control countries.

Equation (6.1) is estimated using a weighted least squares linear probability model, with the dependent variable being the fraction of the total value of imports entering as the particular shipment type. The unit of observation is imports of a given product group by a given country in a given month; product groups follow the classification in the 21 Harmonized System sections (Harmonized Commodity Description and Coding System, version 1988). The sample consists of 11,303 observations, where each observation stands for a shipment. Each observation is weighted by mean monthly imports (by value) in the 'before' period (November 1989–April 1990) for that particular country/product group, allowing for the coefficients to be interpreted as reflecting the probability that a particular dollar of imports falls into a shipment type. It is possible that error terms in Equation (6.1) may be serially correlated among observations from the same country; therefore, robust standard errors clustered by country of origin of imports are used.

The prediction is that the reduction in the minimum value threshold from US$5,000 to US$500 makes import duty evasion on shipments valued above that threshold harder. We would therefore expect that when the shipment type in question is 'shipments valued between US$5,000 and US$500', $\beta_1^j < 0$, which would show that shipment type had become differentially less common. If displacement occurs to alternative duty-avoidance methods associated with the other shipment types, we should expect that $\beta_1^j > 0$ for the other shipment types, which should become differentially more common. These predictions are indeed borne out in Yang's (2008a) results.

The results show that the main coefficient of interest, β_1^j, on the interaction term $T_{gh} \times P_p$ is negative and highly statistically significant, for shipment type 'valued between US$5,000 and US$500'. That is, between the before period and the after period it became differentially less likely (in fact less likely by 1.7 percentage points) that one dollar of imports from a treatment country entered into the Philippines in a shipment valued in this category. This result strongly suggests that increasing inspection requirements deterred importers from valuing their shipments in this category.

So what did importers in the Philippines do instead? The same main coefficient of interest, β_1^j, on the interaction term $T_{gh} \times P_p$, is positive for shipment types 'valued less than US$500' and 'destined for EPZs', suggesting that importers switched to these alternative methods when inspection was ramped up. The coefficient on shipment type 'destined for EPZs' is significantly different from zero, and an interpretation shows that one dollar of imports from a

treatment country became 2.7 percentage points differentially more likely to be in a shipment destined for an export processing zone between the before and after periods. The coefficient on the interaction term for the variable that combines the three different shipment types ('valued between US$5,000 and US$500' plus 'valued under $500' plus 'destined for EPZs')is positive but not statistically significant. The conclusion is that we cannot reject the null hypothesis that the combined use of all three methods of avoiding import duties was unchanged in response to the minimum value threshold reduction.[16]

3.1   Tariffs, Import Volumes and the Magnitude of Displacement

The model proposed by Yang (2008a) to examine the impact of enforcement on criminal activity postulates that when alternative lawbreaking methods involve high fixed costs to entry, crime displacement responds positively to the size of illicit profits threatened by enforcement. In this section we discuss the magnitude of displacement to alternative methods of importing, depending on the size of profits, here proxied by tariff rates and import volumes. Testing whether the magnitude of displacement responds to tariff rates will also help to show whether the shifts to alternative importing methods are motivated by a desire to avoid import duties or by a desire to avoid the pre-shipment inspection itself.

To perform this test, Yang (2008a) introduces tariff rate and import volume variables for each product group, respectively τ_h and $\ln M_h$, into Equation (6.1) through a series of interactions as shown in the equation below:

$$\begin{aligned}
P_{ghp}^j = \alpha_0^j &+ \alpha_1^j T_{gh} P_p + \alpha_2^j \tau_h T_{gh} P_p + \alpha_3^j (T_{gh} P_p \ln M_h) \\
&+ \alpha_4^j (\tau_h T_{gh} P_p \ln M_h) + \alpha_5^j (\tau_h P_p) + \alpha_6^j (P_p \ln M_h) \\
&+ \alpha_7^j (\tau_h P_p \ln M_h) + \gamma_{gh} + \theta_p + u_{ghp}^j.
\end{aligned} \tag{6.2}$$

The coefficients of interest are α_2^j, α_3^j and α_4^j, which show how the magnitude of displacement across shipment types varies with tariff rates and import volumes. Theoretically, we would expect more displacement for products with higher tariffs and higher import volumes, and the impact of tariffs (import volume) on displacement should be lower when import volume (tariffs) is already high. That is exactly what the estimates of Equation (6.2) show.[17]

An example illustrates the results obtained for the type 'shipments destined for EPZs' for which the estimated α_2^j and α_3^j are positive (as predicted

---

[16]Detailed results from the regressions reported here can be found in Table 2 in Yang (2008a, p 10). That paper discusses the many possible objections that could occur with setting up the regression like this, including using a linear probability model and not a multinomial logit, and adding fixed effects for country/product group and month. Adding such fixed effects is shown to have no impact on the results.

[17]The estimates are presented in Table 3 in Yang (2008a, p 11).

by the theoretical model): for a product group such as HS10, 'Pulp, paper, paperboard and articles thereof', whose $\ln M_h$ is 16.74, near the mean across product groups, a tariff rate increase of 10 percentage points from its previous level would have led to an increase in displacement to EPZs amounting to 1.6 percentage points of total imports. The coefficients obtained for the shipment type 'valued between US$5,000 and US$500' are of the opposite sign relative to those in the shipment type 'destined for EPZs', and show that displacement is taking place out of shipments of that type and into shipments to EPZs.

The results show that, before the reduction in the minimum value threshold, the products most likely to be valued below US$5,000 to avoid PSIs were those with high tariffs or high import volumes (but that the impact of either of these variables singly was lower the higher the level of the other variable). Once the minimum value threshold was lowered, importers shifted the products in question away from valuations between US$5,000 and US$500 into shipments going to EPZs. Thus, higher tariffs and higher import volumes make it more attractive to displace shipments to alternative methods of duty avoidance.

3.2 Did the Total Value of Imports and Total Government Revenue Change?

In anticipation of possible objections, the Philippine study also looks at the total value of imports originating in the treatment countries and in the control countries. It could be the case that increased enforcement would lead importers to switch trade partners from treatment to control countries, or indeed to misreport the origin of their shipments—said to come from control countries, when in fact they come from treatment countries. Both these actions would see imports from treatment countries decline and imports from control countries go up. However, estimating a regression analogous to Equation (6.1), but where the dependent variable is the natural log of the total value of imports, shows no such effect. Imports do not decline differentially across treatment and control countries.[18]

Even by conservative estimates of tariff revenue gains and losses (net of PSI fees), the minimum value threshold reductions proved to be unprofitable propositions since they led to significant losses in net revenue for the Philippines government.[19]

On the one hand there were revenue gains from two sources. First, because importers were no longer able to avoid the PSI requirement by valuing shipments between US$5,000 and US$500, import duty collections should have increased on shipments that would not have been inspected before. Second,

---

[18] These results are presented in Table 3 of the appendix of Yang (2005).

[19] See Yang (2008a, Table 4 and Appendix) for details of the calculations showing those losses.

shipments were not subject to PSI if they were shifted to valuation under US$500 or to EPZs, and would then have saved the government the cost of the inspection fees. The estimated total revenue gains from these two sources amounted to roughly US$24.6 million.

On the other hand, the costs to the Philippines government far outweigh these modest revenue gains. First, additional inspection fees of shipments valued between US$500 and US$5,000 would have amounted to US$28 million. Second, losses in import duties due to shifts to valuation under US$500 and shipments routed via EPZs would have totalled US$33.3 million. Therefore, the minimum value threshold reductions led to a net loss of US$36.8 million for the Philippine government. Given the magnitude of the estimated gross losses relative to the gross gains, the overall conclusion should be robust to relatively large changes in the assumptions used for the calculation.

It is possible that the government suspected at the outset of the programme that this would transpire, but was constrained by a lack of computerised tallies of shipments. In addition, it was unclear beforehand what fraction of shipments under US$5,000 was declared as being in that value range purely to avoid the PSI requirement. However, the large displacement to EPZs was probably unanticipated.

3.3 PSI Reform in Colombia

Yang (2006) conducts another micro-level empirical analysis of the impact of a PSI reform in Colombia. The implementation of the Colombia pre-shipment inspections was also done in a way that made analysis amenable to difference-in-difference estimation. The years 1993–1994 constitute the 'before' period, prior to the PSI programme. The Colombian government started its PSI programme in August 1995, taking until March 1996 to finalise the list of products for which PSI was required. These years are excluded from the analysis, as the programme design was still changing. The years 1997–1998 constitute the 'after' period, when the PSI was fully operational on a subset of products. Due to broader changes within the Colombian public administration system, the PSI programme was cancelled altogether in July 1999. The treatment group was a subset of products on which PSI was required, and the control group comprised products for which PSI was not required.

The study takes SITC Rev. 3 product codes[20] disaggregated at the four- or five-digit level level as the unit of observation and covers a total of 2,427 products and 9,314 units of observation, where an observation is a product at that disaggregated level (explained below) in a given year. Each observation is weighted by the product's mean annual dollar imports in 1993–1994.

Creating the right treatment and control group in this study was not straightforward. The treatment group, called 'PSI products', comprised an

---

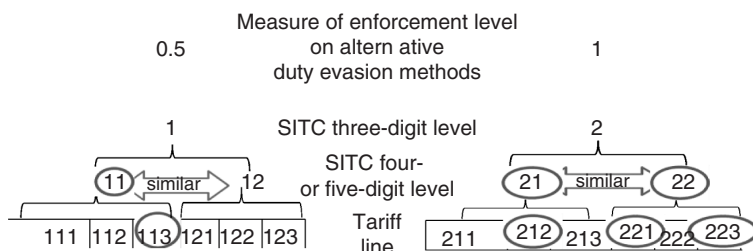[20]Standard International Trade Classification, Revision 3.

Figure 6.1: *Identifying a control group.*

SITC Rev. 3 product at the four- or five-digit level if PSIs were required on any of the HS 1996 tariff lines within that four- or five-digit-level product. Figure 6.1 is an example of this hierarchy. It shows that for the SITC four- or five-digit-level product coded as 11, only one of the subgroups faces tariffs: 113 (circled). This is enough for the product at the four- or five-digit level to be classified as a 'PSI product' and to fall into the treatment group. Product 12 (at the SITC four- or five-digit level) faces no tariffs in any of its subgroups, but it does not classify as the control group. This, as the study explains, occurs because these 'similar' products (11 and 12) would make it too easy for importers to misclassify PSI products as non-PSI products, leaving us with a control group that is biased. In the case of products 21 and 22 at the SITC four- or five-digit level, both face tariffs in their subgroups, thus making it harder for importers to misclassify product 21 as product 22 or vice versa.

The true control group comprises products that are neither 'PSI products' nor 'similar' products. Such products are not shown in Figure 6.1 but we could imagine that they would be product 3 at the SITC three-digit level, which faces no tariffs at the most disaggregated level.

For the difference-in-difference estimation, the identification assumption is that, were the PSI programme not in place, changes in the import capture ratio for PSI products (treatment group) would have been the same as those for non-PSI products (control group), *ie* products not receiving PSIs and not in the same aggregated three-digit-level product group as PSI products (our imagined product 3 in Figure 6.1).

In the Colombia analysis the impact of the PSI programme on duty avoidance is estimated via the import capture ratio,[21] which is the ratio of Colombia's reported imports of a given product divided by its trade partners' reported exports of that same product:

$$\frac{\text{home reported } M \text{ of product } h}{\text{rest-of-world reported } X \text{ of product } h}.$$

---

[21] Technically, the log of the import capture ratio is used due to wide variation in this ratio across different products.

When the import capture ratio equals 1, there would appear to be no misre-porting or undervaluation of imports at Colombian customs (with the assump-tion that exports are not being misreported by the trading partner).[22] When the import capture ratio is less than 1, it would appear that importers are either undervaluing shipments to pay less import duty or misclassifying prod-ucts into categories where PSIs are not needed—or practising outright smug-gling. Undervaluation is harder when PSI is mandated on a given product, but the other two methods of duty avoidance can still be used. We are therefore unclear *ex ante* on how PSI affects the import capture ratio: whether it should increase or decrease. When it increases, PSI is working and imports are being more accurately reported.

The difference-in-difference regression equation captures the treatment effect for PSI products; but it also includes a variable to control for the 'sim-ilar' category described above: a product not subject to PSIs but in the same aggregated product group at the SITC three-digit level as PSI products, which is then interacted with the post-treatment period.[23] The regression includes product and year fixed effects.

The Colombia study follows the same analytical model for smuggling used in the Philippines study: in a situation where enforcement targets a subset (here, products subjected to PSIs), crime displacement to alternative meth-ods of 'lawbreaking' depends on the size of illicit profits threatened by the enforcement. Hence, the study seeks to understand whether PSIs raise the import capture ratio of the treatment group *more* when enforcement is higher on alternative methods of duty avoidance, and *less* when there is more at stake in terms of illicit profits[24] that importers stand to lose due to the PSI. The mea-sure of 'enforcement levels' on alternative methods of duty avoidance is the mean PSI coverage at the three-digit-level aggregated product group: in Fig-ure 6.1 for product 1 at the three-digit level, only one of two sub-products faces PSIs (11); therefore the measure of enforcement levels is 0.5.

For product 2 at the three-digit level, both sub-products 21 and 22 face PSIs and hence the measure of enforcement levels is 1. In other words, 'enforce-ment levels' captures whether another product at the aggregated three-digit level also is subject to PSI, the idea being that misclassification is anticipated, and hence protected against.

---

[22]The working paper version where this study is first presented (Yang 2006) acknow-ledges this assumption, and also discusses the fact that imports are valued cost, insur-ance and freight (CIF) but exports are valued FOB, meaning that reported import value will never be exactly equal to the reported export value.

[23]The original paper discusses the issue of negative bias generated here.

[24]Proxied once again by the tariff rate. See Yang (2006) for detailed discussion on the tariff rate.

The empirical estimates bear out the analytical model. The import capture ratio for Colombian products declines when import tariffs that an importer might need to pay because of the PSI are higher. In other words, when they stand to lose a larger share of profits, importers are more likely to resort to other means of duty avoidance, either misclassification or smuggling.

In addition, duty avoidance on products subject to PSI shifts from under-valuation to misclassification when it is easy to misclassify products into categories that are similar to but not quite the same as the PSI products. In Figure 6.1, product 11 is subject to PSI, and thus tends to be misclassified as product 12. The study gives the example of 'new pneumatic car tyres', classified SITC Rev. 3 code 6251, which can be more easily misclassified as code SITC Rev. 3 62593, which stands for the similar product 'used pneumatic car tyres'.

But if similar products face higher levels of PSI coverage (in our figure, products 221 and 223), the chances of misclassification (in our figure, from product 21 to 22) being noticed by customs authorities are higher, and so the import capture ratio goes up significantly. The latter possibility is amply shown in the model.

4 CONCLUSION

The findings of Yang (2008a, 2006) are intuitive and yet surprising. The studies show that if a country is facing two distortions, remedying one does not necessarily make society better off and in fact might even make it worse off. This is in line with the principle of the second best, according to which correcting one market failure in the presence of a second one does not necessarily raise welfare.

In the Philippine customs example, lowering the minimum value threshold did not lead to greater import duty collections, because importers could avail themselves of a second distortion: the export processing zones. In fact, as the fraction of imports destined for EPZs increased, the government would have collected even less revenue than it did before. Therefore, closing just one loophole—the particularly high minimum value threshold—arguably made things worse in the presence of another loophole: the EPZ. By revealed preference—the fact that importers preferred using the minimum value threshold loophole by slicing their shipments when they could also avail themselves of the EPZ, we can conclude that the latter was costlier than the former.

The EPZ channel was likely to have been costlier because it required physical relocation there or the establishment of relationships with cooperative importers in the EPZs, involving sunk and fixed costs. Thus, forcing importers, so to speak, to use the EPZ made importing more costly to them. This would not necessarily have been detrimental overall if it had brought more revenues to the customs authorities. However, this was not the case. So the government was not better off and the importers were worse off.

In the long run, we might expect that market arrangements, such as support services firms, would emerge that would reduce the costs associated with EPZ relocation, thus reducing the dissipative nature of the EPZ channel. But the benefits of the partial customs reform could only be marginal at best, which only a simultaneous study of the two loopholes could show.

Similarly in the case of Colombia, if 'product 11' was mandated to undergo a PSI, but a similar product, 'product 12', was not, importers could resort to the option of misclassification. Thus, the Colombian government's reform to implement a PSI on a certain subset of products and therefore tackle under-valuation left open the loophole of misclassification to a product in the same aggregate group that did not have to undergo PSI; so, the reform did not raise import duty. The import capture ratio went up significantly only when all products in the same aggregate group had to undergo PSI. In other words, it is only addressing both distortions simultaneously that led to positive results.

In a wide variety of contexts, governments seeking to discourage an unde-sirable activity face the possibility that increased enforcement could simply push the activity to alternative channels. In this chapter we have reviewed the evidence that, in the context of corruption in customs, enforcement-induced crime displacement responds to the size of illicit profits threatened by enforcement. In addition, we have documented that crime displacement can be very large, leading the amount of crime to be essentially unchanged after the increase in enforcement. Displacement in the case of the Philippines PSI reform was greatest for product groups with higher tariffs and higher import volumes.

Mohini Datt is a Consultant in the International Trade Department, Poverty Reduction and Economic Management Network at the World Bank.

Dean Yang is an Associate Professor at the Gerald R. Ford School of Public Policy and Department of Economics, University of Michigan.

REFERENCES

Anson, J., O. Cadot, and M. Olarreaga (2006). Tariff evasion and customs corruption: does pre-shipment inspection help?, *Contributions to Economic Analysis & Policy* **5**(1), article 33 http://works.bepress.com/jose_anson/1.

Bardhan, P. (1997). Corruption and development: a review of issues. *Journal of Economic Literature* **35**, 1320–1346.

Becker, G., and G. Stigler (1974). Law enforcement, malfeasance, and compensation of enforcers. *Journal of Legal Studies* **3**, 1–18.

Bernard, J.-T., and R. J. Weiner (1990). Multinational corporations, transfer prices, and taxes: evidence from the us petroleum industry. In A. Razin and J. Slemrod (eds), *Taxation in the Global Economy*, pp 123–154. University of Chicago Press.

Bertrand, M., S. Djankov, R. Hanna, and S. Mullainathan (2007). Obtaining a driver's license in India: an experimental approach to studying corruption. *Quarterly Journal of Economics* **122**, 1639–1676.

Besley, T., and J. McLaren (1993). Taxes and bribery: the role of wage incentives. *The Economic Journal* **103**, 119–141.

Byrne, P. (1995). An overview of privatization in the area of tax administration. *Bulletin for International Fiscal Documentation* **49**, 10–16.

Clausing, K. (2003). The impact of transfer pricing on intrafirm trade. In J. R. Hines Jr (ed), *International Taxation and Multinational Activity*, pp 173–199. University of Chicago Press.

Di Tella, R., and E. Schargrodsky (2003). The role of wages and auditing during a crackdown on corruption in the city of Buenos Aires. *Journal of Law and Economics* **46**, 269.

Fisman, R., and S.-J. Wei (2004). Tax rates and tax evasion: evidence from 'missing imports' in China. *Journal of Political Economy* **112**, 471–496.

Goorman, A., and L. De Wulf (2003). Customs valuations under the new WTO rules: problems and possible measures with particular attention for developing countries. Mimeo, World Bank.

Hesseling, R. (1994). Displacement: a review of the literature. In R. V. Clarke (ed), *Crime Prevention Studies, Volume 3*. Monsey, NY: Criminal Justice Press.

Hines Jr, J. R., and E. Rice (1994). Fiscal paradise: foreign tax havens and American business. *Quarterly Journal of Economics* **109**, 149–182.

Huntington, S. P. (1968). *Modernization and Corruption: Political Order in Changing Societies*, New Haven, CT: Yale University Press.

Myrdal, G. (1968). *Asian Drama: An Enquiry in the Poverty of Nations*, Volume II. New York, NY: The Twentieth Century Fund.

Mookherjee, D., and I. P. L. Png (1992). Monitoring vis-à-vis Investigation in enforcement of law. *American Economic Review* **82**, 556–565.

Mookherjee, D., and I. P. L. Png (1995). Corruptible law enforcers: how should they be compensated? *The Economic Journal* **105**, 145–159.

Nagin, D., J. Rebitzer, S. Sanders, and L. Taylor (2002). Monitoring, motivation, and management: the determinants of opportunistic behavior in a field experiment. *American Economic Review* **92**, 850–873.

Olken, B. A. (2007). Monitoring corruption: evidence from a field experiment in Indonesia. *Journal of Political Economy* **115**, 200–249.

Polinsky, A. M., and S. Shavell (2001). Corruption and optimal law enforcement. *Journal of Public Economics* **81**, 1–24.

Pritchett, L., and G. Sethi (1994). Tariff rates, tariff revenue, and tariff reform: some new facts. *World Bank Economic Review* **8**, 1–16.

Ramírez Acuna, L. (1992). Privatization of tax administration. In R. M. Bird and M. Casanegra de Jantscher (eds), *Improving Tax Administration in Developing Countries*. Washington, DC: International Monetary Fund.

Reinikka, R., and J. Svensson (2004). Local capture: evidence from a central government transfer program in Uganda. *Quarterly Journal of Economics* **119**, 679–706.

Rose-Ackerman, S. (2004). the challenge of poor governance and corruption. Mimeo, Copenhagen Consensus, Frederiksberg.

Sequeira, S., and S. Djankov (2010). On the waterfront: an empirical study of corruption in ports. Mimeo, London School of Economics.

World Bank (2002). World Development Indicators 2002. Report. The International Bank for Reconstruction and Development/The World Bank, Washington, DC.

Yang, D. (2005). Can enforcement backfire? Crime displacement in the context of customs reform in the Philippines. Gerald R. Ford School of Public Policy Working Paper Series 02-010, University of Michigan.

Yang, D. (2006). The economics of anti-corruption: lessons from a widespread customs reform. in S. Rose-Ackerman (ed), *International Handbook of the Economics of Corruption*. Northampton, MA: Edward Elgar.

Yang, D. (2008a). Can enforcement backfire? Crime displacement in the context of customs reform in the Philippines. *Review of Economics and Statistics* **90**, 1–14.

Yang, D. (2008b). Integrity for hire: an analysis of a widespread customs reform. *Journal of Law and Economics* **51**, 25–57.

7

Reforming Customs by Measuring Performance: A Cameroon Case Study

THOMAS CANTENS, GAËL RABALLAND, SAMSON BILANGNA
AND MARCELLIN DJEUWO[1]

Des Lupeaulx: After all, though statistics are the childish foible of modern statesmen, who think that figures are estimates, we must cipher to estimate. Figures are, moreover, the convincing argument of societies based on self-interest and money, and that is the sort of society the Charter has given us, in my opinion, at any rate. Nothing convinces the 'intelligent masses' as much as a row of figures. All things in the long run, say the statesmen of the Left, resolve themselves into figures. Well then, let us measure.

H. de Balzac (1844, *Les Employés* ['Bureaucracy'], Chapter 9)

1 INTRODUCTION

In 2007 Cameroon Customs launched a reform and modernisation initiative meant to reduce corruption. Fraud and corruption had long stained the administration's reputation and hindered fulfillment of its mandates. The reform began with the installation of ASYCUDA,[2] a customs clearance system that allowed the administration not only to track the processing of each consignment, but also to measure a substantial number of criteria relevant to the reform, such as compliance with the deadline for recording the manifest by consignees.

---

[2] Automated SYstem for CUstoms DAta.

For almost two years, upper management and front-line officers in Cameroon Customs shared the same reality thanks to 'figures' (performance indicators) that measured how the reforms initiated by the former were applied by the latter. But, while the initial quantification phase bore fruit, its impact later stalled. A possible solution was adopted when, beginning in 2010, Cameroon Customs introduced a system of individual performance contracts to measure the actions and behaviours of customs officers operating at two of the seven Douala port bureaus, using indicators extracted from AsycudA.

The outcomes are encouraging. After more than 10 months of implementation, the Cameroon Customs bureaus in the experimental group have shown better results than the control group on indicators related to the reduction of corruption, revenue collection, and trade facilitation.

In this chapter we trace the Cameroon reform from the introduction of the performance indicators to the measured results of performance contracts. In Section 2 we provide an overview of measurement theory and the Cameroon Customs reform. In Section 3 we recount the events and decisions that over the last four years that gradually led Cameroon Customs to introduce quantification of their actions. In Section 4 we discuss the principles of the performance contracts and describe the indicators and in Section 5 we present the results. In Section 6 we then describe the difficulties raised by the introduction of performance measurement. Finally, we offer concluding assessments in Section 7.

2   NEW PUBLIC MANAGEMENT AND PERFORMANCE CONTRACTS

In Honoré de Balzac's novel *Les Employés* about public servants in the 19th century, one of the characters, the secretary general of finance, Monsieur Des Lupeaulx, admits that figures are important to reform. However, in Balzac's time, when reform was applied to a ministry, it generally meant that a number of bureaucrats were themselves *réformés*, *ie* 'reformed' in the sense of 'fired' (Ymbert 1825).

Over a century later in the 1970s, the *new public management* (NPM) approach introduced private sector management techniques into the public sector. NPM encompasses a heterogeneous range of instruments and ideas that share the neoliberal critique of the welfare state. Starting as a simple desire to reduce the power of the state—mainly by privatising its actions in the social sectors—NPM evolved into a softer version with the introduction of the 'public entrepreneur' figure: a public servant who, in order to achieve global objectives, enjoys a degree of flexibility in organising their resources.[3]

---

[3]Some preliminary conclusions on the effects of NPM for the USA, the UK, New Zealand and Australia are provided by Mascarenhas (1993), Considine and Lewis (2003) and Julnes *et al* (2001).

In February 2010 Cameroon Customs launched an experiment that shares a number of *mots d'ordre* with NPM, namely, 'autonomy', 'performance' and 'quantification'. This pilot resulted in the implementation of individual performance contracts signed between the director general of customs and two key customs offices in the port of Douala. These contracts represented a new step in the process of improving performance that had begun three years earlier.

Performance contracts are based on the objective measurement of the actions of public servants and financial incentives or career advancement policies. In Cameroon, performance contracts aim to encourage customs officials to adopt good professional practices. Indeed, fighting bad practices is a key element in improving customs revenues: corruption has a direct negative impact on customs revenues and on competition in the private sector.

The history of Cameroon Customs shows that their public sector has the same capacity to absorb private sector techniques as countries such as the USA and the UK, where NPM developed: a pre-shipment inspection company (in place since 1987) was responsible for evaluating imported goods and some exports; public servants, including customs officials, were dismissed and public sector salaries were cut in the wake of the structural adjustment plans that Cameroon underwent during the 1990s.[4]

In Cameroon, external constraints such as the structural adjustment have often been responsible for imposing techniques from the private sector. This is similar to many countries that have seen an evaluation culture imposed from outside. The 1990s therefore saw their first *deflates* in Cameroon customs, a French term coined to describe lay-offs, including voluntary departures, from the public service (Mbonji 1999), which echoes Balzac's *réformés*.

Still, the flexibility of Cameroon Customs' contracts differs from that of the NPM's 'public entrepreneur'. Because of the climate of corruption, the director general knows less than his subordinates do about how or to what extent agents are applying the reforms adopted. This applies to a large number of administrations in sub-Saharan Africa (Mbembe 1999; Raffinot 2001; World Bank 2005).

The contracts also aim to strengthen the hierarchy to achieve reform. The Cameroon experiment does not therefore have the same vision as the NPM policies, which question the Weber model of administration (Rouban 1998; Spanou 2003) and result from agency theories based on principal-agent models.

Despite its importance in debates about state reform, performance measurement in practice is not widespread enough to provide much literature on the subject (Julnes and Holzer 2001). The impact of incentives policies is rarely measured rigorously, and studies evaluating reforms often focus on

---

[4]For the development on NPM in the USA and the UK see Merrien (1999) and Considine and Lewis (2003).

public services in social areas, such as health or education, and on front-line agents (Considine and Lewis 2003). To our knowledge, no project evaluations of fiscal administrations have been carried out *anywhere* in the world, much less in sub-Saharan Africa. In addition, although there are often claims of resistance to policies that aim to quantify public actions, there is little research measuring their implementation (Wholey and Hatry 1992). This chapter will present the effects of the performance measurement policy in Cameroon Customs by correlating the history of the reform and the figures that measure its results.

3   HISTORY OF THE REFORM

3.1   Asycuda *Implementation*

On 1 January 2007 the head of the IT Division of Cameroon Customs in Douala became, in the eyes of his colleagues and of freight forwarders, the agent of a mini-revolution. By disconnecting the Pagode[5] computerised customs clearance system, he put an end to a beleaguered 20-year history of a software system that processed 90% of customs revenue. The following day he launched customs activities on Asycuda, a system developed by UNCTAD.[6] This seemingly independent IT switchover was in reality the culmination of an eight-month process to reform customs procedures. From the start, this act and Asycuda were challenged, since customs revenues represented 27% of national revenues in Cameroon, the public sector was the country's largest employer and most consumer products were imported.[7]

The automation process first abolished the 'release note', a process that had forced customs brokers or importers to return to the customs inspector after paying their customs debt in order to obtain the 'release'. Automation also allowed customs officials to concentrate more fully on their core activities. The common customs clearance halls were closed; customs officials no longer jointly managed customs warehouses and areas in the port together with their private owners; and they no longer managed the connections to the customs network. This combination of automation and employee empowerment consolidated real change. By the end of 2007 all customs officials were aware of the system's potential with regard to internal auditing of the service, having themselves been victims of or having exploited it. The cross-checking of data now regulated the operational managers' reports to the director general,

---

[5]*Procédures Automatisées de Gestion des Opérations de la Douane et du commerce Extérieur* (computerised management procedures for customs and external trade operations).

[6]United Nations Conference for Trade and Development.

[7]The public sector in Cameroon employs 170,000 public servants and military servicemen.

and a drop in revenue could no longer be explained by a decline in economic activities.

3.2   The Launch of a Measurement Policy: A Key Pillar of the Cameroon Customs Reform

In January 2008 the director general decided to make performance indicators a pillar of his reform policy and set up a team of computer experts and customs officials. However, by his not being in the field, his flow of information from operational staff was not complete—especially within a climate of corruption. This situation was made worse by the relocation of the directorate general to Yaoundé, more than three hours away from the Douala port by road. The director general's 'co-management' of customs operations, as described by some senior officers who denounced it in 2006, had at least allowed operators to complain quickly to the superior authority. By relocating, the director general cut himself off from direct sources of information.

Starting in February 2008, the Director General's team established 25 indicators for the 11 customs offices in Douala. These indicators measured economic activity from a customs viewpoint: the times taken by customs officials and brokers to process files, the effectiveness of controls and sensitive procedures and compliance with the customs channels. The full list of indicators is presented in Table 7.1.

The indicators are based on the notion that fraud and corruption are necessarily linked (Libom *et al* 2009). After seeing the gains made from automating procedures, the next step in revenue capture was to attack corruption. Evaluation became institutionalised, making it more effective than sporadic controls, and thus became a 'social process' (Varone and Jacob 2004).

By introducing indicators, the director general chose a radically different course from that of his predecessors who managed PAGODE. Because it is automated, no customs clearance system, be it PAGODE or ASYCUDA, can compete with the vivid imagination and determination of fraudsters. Under PAGODE, the development policy sought to strengthen the system permanently by detecting frauds on a case-by-case basis. But ultimately, the system's complexity created a strong dependence on computer experts. Under ASYCUDA, this policy was reversed. The system continues to offer computer security designed to identify the actors and their acts. But it is also judged on its capacity to provide a realistic image of customs clearance in the field.

In 2008 and 2009 the director general and operational managers met in Douala to examine the monthly report on indicators. This report was distributed to them before the meeting, with each noting their own results and the results of their colleagues as well as each unit and each inspector. These meetings gave managers the opportunity to better manage their subordinates.

Table 7.1: *Customs indicators.*

Activity Indicators	Performance Indicators	Control Indicators	Risk Management Indicators
Number of manifests recorded	Compliance with the deadline for recording the manifest by the consignees	Adjustment manifests	Declarations in the red channel with removal report but not seen by scanner
Number of containers recorded	Assessment period	Amendments before assessment and offsetting entries	Declarations rerouted to other channels
Number and amounts of declarations recorded	Payment period	Value added of the amendments and offsetting entries by inspector	Value-added of rerouting declarations
Removal notes validated	Average time between assessment and issue of removal note by freight forwarder	Declarations assessed but not paid	Transit documents cancelled
	Number of declarations not assessed	Exemptions	Monitoring volumes
	Removal notes not found	Reassignments of declarations	Number of operations carried out by physical persons
	Reporting time by freight forwarder	Workload by inspector	
		Declarations cancelled	
		Declarations paid without removal note	

3.3 The Impact of the Measurement Policy

The results from the performance indicators policy were real.[8] The tax yield of a declaration increased consistently: an additional 21% between the first quarter of 2007 and the second quarter of 2009 despite the tax exemption measures for staple foods adopted in March 2008. For containerised imported goods for domestic use, the average yield of a declaration increased by 10% in 2007.[9] Disputed claims increased without any additional pressure on operators: the share of duties and taxes collected following controls rose from 0.75% to 1.02% of revenue. The duties and taxes collected in this way increased by 56%, while the number of disputed claims increased by only 12%.

In terms of trade facilitation, 75% of maritime manifests were recorded in the system 24 hours before the arrival of the vessel, allowing 18% of declarations to be submitted before unloading the goods. In the port of Douala between 2007 and 2009, the average assessment time of all customs offices collectively was reduced from 1.2 days to 0.8 days.[10] The total processing time between 2007 and 2009 was reduced from 6.5 days to 5 days.[11] Efforts were also made by the freight forwarders and managers of customs warehouses and customs clearance areas. This significant improvement in processing time for the maritime professions shows the indirect impacts of the performance indicators policy.

Backed by performance indicators, customs officials strengthened their capacity for dialogue with the private sector and weakened the intermediation of the shipping agents. By disseminating some indicators to importers and exporters, the director general was able to demonstrate the responsibility of all actors in the customs clearance process. As a result, some freight forwarders processing large volumes of goods reduced their intervention times by half between the beginning of 2008 and the end of 2009.

Other non-quantifiable results pointed to a gradual acceptance of the constraints linked to performance measurement and its integration into hierarchical reports. At meetings, some operational managers came with their

---

[8]All figures are calculated for the import declarations cleared for home use, which represent 80% of declarations in terms of amounts and in numbers at the port of Douala. The calculation thus avoids taking account of the different offices for which the remit may have been amended or of changes in rules affecting special procedures or procedures specific to public contracts.

[9]In all offices, the number of items per declaration has not varied significantly; the indicator of average assessment remains relevant over the period.

[10]The assessment time is the period between the submission of the declaration by the freight forwarder and assessment by the customs service.

[11]The total processing time is the total of three periods: assessment period, payment period (period between the assessment by the customs officer and payment by the customs broker) and removal period (period between payment and obtaining the removal note issued by the manager of customs clearance warehouses).

laptops, allowing them to view the indicator report directly. Others submitted a monthly report matching the indicators with their own data, which they had been obliged to collect.

Some indicators even raised questions that forced the heads of operational services to conduct their own internal investigations in order to justify results. In one case, where the interpretation of the results continued to oppose that of the directorate general and operational services, cross-checks revealed the failings of the pre-shipment inspection provider.

After two years of operation, the directorate general of customs decided that the operational services in Douala had had enough time to adopt the performance measurement techniques, while at the same time a winding down of this policy threatened to combine with the impact of the global financial crisis of 2008–9. On the one hand, economic activity linked to external trade was dwindling: exports dropped by 40% between 2008 and 2009, and imports concomitantly declined. On the other hand, revenue targets continued to grow (by 6% in 2010), while there was a gradual fall in values declared from the end of 2009 onwards. Meanwhile, the level of enforcement in the field by customs officials seemed to have evened off. While this has little impact on revenue, it remains an indicator of relations between customs officials and users.

4   PERFORMANCE CONTRACTS: A NEW STEP

Over time, the impact of the performance indicators policy seemed to fade. One proposed solution was to move from a purely descriptive performance measurement to a prescriptive measurement. At the end of 2009, Cameroon Customs obtained funding and technical assistance from the World Bank and the World Customs Organization to support the implementation of individual performance contracts in the port of Douala over a ten-month experiment period.

4.1   Principles

During the pilot stage, the performance contracts were launched in two of the seven offices in the port of Douala that collect 76% of the port's revenue. Office DP I handles imports of goods in containers for clearance for home use, with the exception of vehicles, has 10 or 11 inspectors and collects 60% of revenue. Office DP V handles imports of vehicles, including in containers, has between five and seven inspectors and collects 16% of revenue.

Like any other contract, the performance contracts formalise an agreement between two parties, specifying mutual obligations regarding results. The contracts go beyond revenue targets, which are fixed annually for the government by Customs. Still, these revenue targets, as well as the distribution of products of disputed claims and of protocols, already formed a 'numbers system'

(Ogien 2010). This situation is common to all fiscal administrations and offers good potential for measuring performance.

The Cameroon contracts incorporate two specific features that address corruption. First, they are signed between the director general and, individually, the sector head, the two office heads and the customs inspectors. Each commits themselves directly to the director general and not to a direct hierarchical superior. Second, the global objectives inherent in every customs administration (trade facilitation and enforcement) are complemented by objectives to abolish bad practices. In very simple terms, the performance contracts aim to urge customs inspectors to clear declarations faster, detect more fraud and give up bad practices.

Unlike NPM, which seeks to optimise an administration/structure in relation to its objectives (Wholey and Hatry 1992; Strathern 2001), the Cameroon Customs contracts are not based on a relationship between resources allocated and services provided. Instead, the goal is for individuals to comply with the formal structure, *ie* the match between their actions and the rules of the structure is measured, which reinforces the formal framework and does not call into question Weber's model of organisation of the administration. The contracts evaluate the adherence of individuals to the organisation's rules. However, the principle of rational choice, which assumes that individual behaviour is guided by individual profit, and which characterises NPM (Mascarenhas 1993), is pushed to its extreme in Cameroon, by the individualisation of the contracts.[12] The system of incentives and sanctions is therefore at the heart of the reform.

A clear distinction needs to be made between sanctions and incentives. Cameroon Customs has long since adopted a policy of financial incentives via the distribution of the yield from fines and various memoranda of understanding with its partner professions (Bilangna 2009; Cantens 2009). During meetings with the inspectors and their superiors to draft the contracts, exchanges showed the importance of non-financial incentives. Indeed, given the method of distributing the yield from fines, which legally guaranteed each inspector 10% of the fine imposed with no upper limit, inspectors showed no interest in additional financial incentives. At any rate, they did not believe that the administration could better this 10% any more than it could compensate, in the case of those inspectors who were corrupt, for the profits lost due to ethical behaviour.

A number of incentives have been introduced: congratulatory letters; recording the congratulations in agents' personnel files; easier access to the director general through regular meetings; reviewing the professional aspirations of successful agents; and training courses.

---

[12]The idea of team performance is also present to the extent that indicators targeting the operation of the team have been introduced for heads of the Customs office and of the sector.

Regarding sanctions, the contracts include a process of interviews and warnings. For agents, the main sanction remains eviction from offices with strong fiscal potential and where the possibilities of earning money legally through disputed claims are high. From this point of view, the threat of sanction by transfer to an office with little earning potential would have a greater impact on personal behaviour than the hope of an incentive.

This concurs with the observations of Besley and Ghatak (2005), who found no evidence that incentives mattered in organisations structured around the notion of mission rather than of profit. The Cameroon experience rests more on contractual governance of deviant behaviour (Crawford 2003): five of the eight contract indicators relate explicitly to bad practices that must be curtailed, even though the eight indicators are equally split between four trade-facilitation indicators and four enforcement indicators (following the two cardinal missions of every customs administration). These indicators are described in the following sections.

4.2   The Experimental Protocol and the Measurement of the Impact of the Pilot

For each objective, a comprehensive review was conducted to determine which parameters would be taken into account. Once these parameters were defined, the performance contract set a minimum or maximum threshold. This threshold is a median calculated on the basis of the declarations processed by the offices over the previous three years, 2007–2009. The sample covered 74,591 declarations for office DP I and 63,761 for office DP V.

Preparations for the deployment of the performance contracts lasted from September 2009 to February 2010. The inspectors and their managers were involved at all stages of preparations—from the writing up of the contracts to the choice of indicators and periodic performance reviews. Before the launch, stakeholders were brought together in a contract design workshop. This prompted the creation of a unit specifically in charge of the programme, comprising customs officials and computer staff. During the preparation stage, the contracts did not provoke much debate, given their newness and, above all, the distrust of the operational actors. While this eased the signing of the contracts, the first regular meetings (held every ten days) were tense, many inspectors having not foreseen the consequences of the contracts. The continuing discord led to a number of amendments in the course of the first quarterly evaluation.

Indicators were computed every ten days and presented to the inspectors and to the heads of office by the team in charge of the project.[13] Presenting results to the front-line inspectors was one thing, but monitoring and evaluating the impact of performance contracts on customs efficiency was

---

[13]This calculation is performed via an IT application run on the ASYCUDA database.

another. On the one hand, since revenue collection is the main institutional objective of customs, impact monitoring had to detect any negative impact on revenues. On the other hand, in case of positive impact, evaluation had to provide evidence to convince the political authority to continue supporting such an original experiment.

One gold standard of impact evaluation of public aid is the use of an RCT, whose importance has been increasing substantially since the 1960s in the social, behavioural and mostly educational fields (Duflo and Kremer 2003; Turner *et al* 2003). The very principle of a random allocation raises questions about the Cameroon Customs performance contracts. In this context, it would mean that some front-line inspectors would have been under performance contracts, while others were not, each group being discriminated in a randomised way.

An RCT was not implemented for the following technical reasons. First, the sample is too small. At Douala Port, the seven customs bureaus are specialised: oil imports; special customs regimes related to public trends; transit; exports; bulk cargo; and the two bureaus under contracts. Types of fraud and, above all, customs practices are so different that it is usually impossible to compare one bureau to another. In addition, the two bureaus under contracts are the main ones, both in terms of revenue collection and staff. But the staff comprises at most 10 civil servants; therefore, it is impossible to split one bureau between treated and controlled groups of inspectors.

Second, time was limited. In order to get over the small sample size issue, one solution could have been to organise a turnover within each bureau, which would have artificially increased the number of controlled and treated officers. Despite the fact that 'blinding' would not have been possible, this turnover would have raised some political issues. Indeed, what does 'treated' mean? As already mentioned, contract incentives are not financial, and rewarding good performers with non-financial incentives requires time. A head of customs cannot deliver congratulatory letters on a monthly basis nor appoint good inspectors to better positions every six months. Still, there is no stable core of front-line inspectors: from 2006 to 2010, there were three appointment cycles, whose unpredictability had a major impact on the working environment and practices (Cantens 2011).

In addition to these technical issues, the experiment did not have to identify the best mechanism because there are not many possible mechanisms in this field. Moreover, implementing individual contracts corresponded to the performance measurement policy undertaken in 2008.

The ethical concerns raised by randomised experiments are crucial when experiments deal with illegal practices (Farrington 2003): how can the government and importers accept that bad practices are only partially addressed? In addition, when one officer is under contract and supposed to be more efficient, while another in the same bureau is less efficient, it is difficult for importers to predict the quality of the service they will receive.

Finally, evaluation has to unveil an impact on some outcome, but in this case, the 'bad practices' are difficult to identify. Impact evaluation must take into account that any pre-defined quantitative approach cannot guess what kind of new bad practices could be invented on the field. This is very similar to what Munane *et al* (2007, p 311) described in educational research: the 'non-routine aspects of… practice are especially important to high performance'. Due to corruption, *ex ante* design cannot take into account all illegal or infra-legal possible behaviours. This specific environment—and the fact that reducing corruption is one major objective of the experiment—decreased the possible scope for using an RCT to make causality explicit.

Impact evaluation of the Cameroon pilot performance contracts has thus been conducted in three stages. First, an extensive analysis of the situation before the experiment was carried out, based on a descriptive statistical analysis of the inspectors' behaviours and qualitative observations. Second, monthly quantitative monitoring through comparisons between similar periods before and after the launch of the experiment. This monitoring also included comparisons with the bureau dedicated to bulk cargo imports. The comparison with the evolution of this counterfactual bureau is only feasible for measuring the impact on clearance time, since import procedures are similar for the three bureaus. Third, the quantitative appraisal was compared with a qualitative one through interviews and observations (during evaluation meetings). Moreover, individual interviews of inspectors under contracts were organised with international experts some months after the end of the experiment.

4.3   The Trade-Facilitation Indicators

Two opposing objectives have been integrated into the performance contracts: releasing the goods more quickly and increasing the numbers of disputed claims. It is the balance of the two objectives that limits the harmful effects. Trade facilitation alone would not regulate corruption issues, while enforcement alone could legitimise corruption.

The first global objective of trade facilitation is the time measured between the declaration's entry by the customs broker and its assessment by the customs office inspector.[14] A study of the sample from the last three years shows that a large majority of declarations were assessed on the same day or the next day in both offices. The difference lay at the level of same-day assessment, with DP V processing over 80% of declarations the same day, compared with over 64% in DP I. But, by the end of the next day, both offices had assessed over 90% of declarations entered in the system. A second, smaller tranche was cleared by the maximum three-day deadline, by which time 96–97% of declarations

---

[14]The times are calculated in terms of full days, excluding only those public holidays falling on a Saturday or Sunday.

had been assessed. After this, progress was slower. Thus, the global objective of trade facilitation was defined with two measurable indicators: *a minimum threshold of declarations assessed within one day, and a maximum threshold of declarations assessed within five or more days.*

Two potential problems made it necessary to integrate two more indicators into the performance contracts. The first was the non-assessment of declarations. Not assessing problematic declarations and leaving them on hold in the system artificially reduced the assessment time. The contract has therefore set a *maximum threshold for non-assessed declarations.*

The second more complex problem was the speed of assessment and the offsetting entry. Inspectors chose to offset after assessment rather than amending the declaration prior to the assessment. This practice concerned 80% of adjustments, regardless of control channel, be it physical or documentary. It reduced the assessment time, which therefore no longer showed the time actually taken by the inspector to carry out the control. This choice was a bad practice that may be interpreted in two ways, each compatible with the other:

- once offsetting entries becomes routine, users were at permanent risk of readjustment by the inspector who assessed their declaration;
- by systematically offsetting their adjustments, inspectors ensured that they had a maximum number of declarations to process.

Asycuda automatically assigns declarations to inspectors on the basis of their workload, calculated on the basis of the numbers of declarations already assigned to them waiting assessment. By carrying out the assessment rapidly, inspectors kept their workload at a low level but were obliged to adjust declarations via offsetting entries. Apart from the negative impact on the relevance of the assessment time, this practice resulted in competition between the inspectors. On average, in office DP I, the fastest inspector managed to process up to six times more declarations a day than the slowest. This competition adversely affected the equal treatment of users and could at the same time induce corruption.

The contracts therefore laid down two indicators. The first, in the performance contracts for heads of customs offices, required a *maximum deviation of 1.5 between inspectors' processing speeds*; the head of a customs office could suspend inspectors who processed declarations too 'rapidly'. The second indicator, in the performance contracts for inspectors, set a *maximum threshold for offsetting entries* by inspectors who had assessed the declarations redirected from the physical inspection channel to the documentary control channel.

4.4   The Fight against Fraud, and Bad Practice Indicators

The second global objective of performance contracts is enforcement. The amount of duties and taxes raised depends on both on the number of decla-

rations collected and the additional amounts collected following the controls. The green channel is not currently activated, and so all declarations are subject either to documentary controls or to a physical inspection.

The contracts set a global objective for inspectors: a *minimum percentage of amounts of duties and taxes collected following adjustments compared to the amounts of duties and taxes assessed*. However, this objective may be affected by two biases. The first is the size of the disputed claim. To achieve a minimum amount of adjustments, inspectors could increase the number of small disputed claims. Prior analysis had confirmed the average size of disputed claims over recent years: in office DP V, 65% of adjustments were below €300 (compared with an average assessment of over €3,000). In office DP I, the red channel (physical inspection) showed a paradox: in numerical terms, low-level adjustments predominated over high-level adjustments in the case of high-risk declarations. Forty per cent of disputed claims yielded between 1% and 5% of the amounts assessed. To remedy this tendency, the contracts set a maximum threshold in numerical terms for small disputed claims and a minimum threshold for the highest disputed claims in the red channel.

The second bias that may affect this objective was linked to the rerouting of declarations.[15] Rerouting by inspectors is a legitimate action. However, rerouting through the red channel may also be a means of pressuring the user. Thus, it was necessary to monitor the adjustments carried out on those declarations rerouted to the red channel. The contracts have not set a limit on the number and proportion of declarations rerouted; moreover, the operational services already had enough constraints, obliging them to rapidly process low-risk declarations. The only measurable indicator consisted of setting a *rate of adjustment for declarations rerouted to the red channel greater than that for those declarations that were not rerouted*.

5   THE IMPACT OF PERFORMANCE CONTRACTS

5.1   Quantifiable Results

After only 13 weeks of implementation, halfway through the performance contracts experiment, there were positive results. This testified as much to the effectiveness of the contracts as to the willingness and growth potential of the customs agents. The measurement of the effects is based on a comparison of various performance indicators before and after the implementation of the performance contracts experiment in the two treated customs bureaus and in some cases is compared with the performance indicators in the counterfactual bureau, as mentioned in Section 4.2.[16]

---

[15]'Rerouting' means the redirection of the declaration to a processing channel other than the original channel.

[16]In each customs bureau, indicators are averaged across inspectors.

The impact on revenues is measured relative to trends in economic activity (for which it is difficult to obtain reliable data in real time). Hence, several variables, such as the numbers of imported containers, values declared and the numbers of items or total tonnage, which can easily be extracted from the Asycuda system, were selected. None of this data gives irrefutable evidence for economic activity, but rather it points to trends that may help to interpret revenue developments.

In office DP I, the duties and taxes assessed over the period increased by 6.2% in 2010 relative to the contract period in 2009, while the number of imported containers fell by 3%. The tax yield of the declarations in office DP I rose by 3% over the contract period in 2010 compared with the same period in 2009. In office DP V, it rose by 23%.[17]

We then computed the estimated revenue added during the pilot (all other things being equal), which is equal to the revenues actually collected during the experiment minus the number of declarations during the experiment multiplied by the average taxes and duties of 2009. The estimated additional revenues during the experiment are estimated to be €23.3 million (which is about 3% of the national customs revenue target).

The impact of the performance contracts on customs clearance time has been equally important. The share of declarations assessed by inspectors on the day they are lodged in the system by brokers was multiplied by 1.3 in office DP I (it is now around 84%), by 1.2 in Office DP V (77%) and by 0.9 in the counterfactual office DP VI (57%). The estimated gain in terms of time clearance is 8 hours for office DP I and 14 hours for office DP V.

In addition, the variance of time clearance has dramatically decreased since April 2010, and processing speeds have become more homogeneous. The standard deviation of daily processing speeds and the difference between the fastest and slowest clearance times have both been halved. Inspectors therefore no longer engage in stiff competition to process as many declarations as possible.

The impact on private operators has also been tangible. The time period between the broker's registration and the customs officer's assessment has been divided by 2.5 in office DP I and by 2 in office DP V. There had been no such evolution after the Asycuda launch.

The impact of the performance contracts on disputed claims is also interesting. In office DP V, the impact was substantial since the ratio of taxes adjusted to taxes assessed was multiplied by 1.7. In office DP I, the ratio of taxes adjusted to taxes assessed was multiplied by 1.003, which is an insignif-

---

[17]The results compared the period under contracts from February to November 2010 to the same period in 2009. December and January were excluded because of seasonality concerns: economic activity increases due to the Christmas period and so does the pressure on customs bureaus to achieve the annual revenue targets, which gives rise to specific procedures and low activity following Christmas.
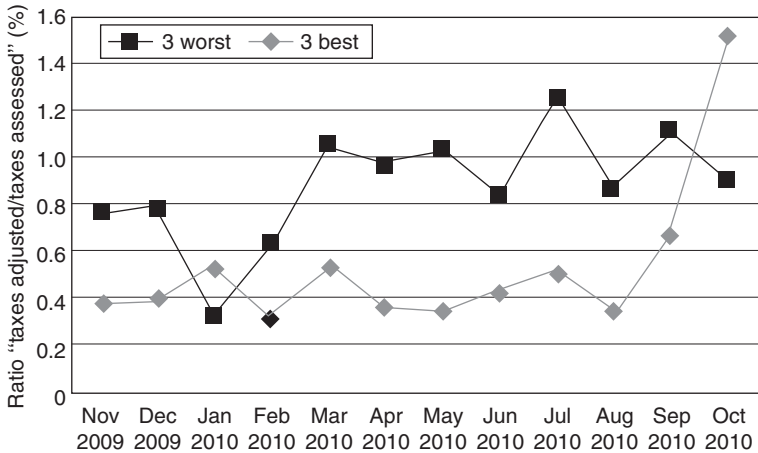
Figure 7.1: *Comparison of the most-efficient and least-efficient customs officers on the ratio of taxes adjusted to taxes assessed over the contract period (as a percentage). Source: authors' calculations based on data from Cameroon Customs. Note: the results shown refer to office DP I.*

icant increase, but the average taxes adjusted per fraud case was multiplied by 1.5.

In qualitative terms these results mean that the performance contracts marked a break: the inspectors abandoned low-level disputed claims to focus more on major ones. In office DP I, some inspectors did not give up their practices: among the eleven inspectors, three did not achieve the contract's target ratio of taxes adjusted to revenue assessed over 7–8 months as shown by Figure 7.1. The inspectors who had an important negative impact—by being less efficient and assessing a lot of declarations—on the bureau's performance in terms of fighting fraud have been transferred.

The performance contracts have also had a major impact on bad practices. First, rerouting from the yellow channel (documents control) to the red channel (physical inspection) is more effective in terms of disputed claims. From this point of view, the inspectors have shown discipline. At office DP I, 38% of rerouted declarations were the subject of litigation, while that figure was 10% in 2009. At office DP V, this value was 49% during the experiment relative to 0.7% in 2009.[18]

---

[18] These figures are almost too high. During a visit to the Cameroon Customs bureaus by the authors, the inspectors expressed caution about carrying out any rerouting of declarations. They had set themselves a very high threshold of disputed claims, much higher than the objective set in the performance contracts. The objective (one rerouted declaration out of six had to be the subject of a disputed claim) was reworded so as not to provoke, in the long run, a perverse effect which would see inspectors no longer rerouting declarations but carrying out, at their discretion, physical checks on declarations in the yellow channel.

In addition, the practice of systematically offsetting declarations in the yellow channel sharply declined, and some inspectors put an end to it entirely. On average, 80% of declarations adjusted in the yellow channel had been adjusted via an offsetting entry by the inspector who carried out the assessment. But this proportion fell to 7% for office DP I and 19% for office DP V by April 2010 and close to 0% in the following months.

5.2   Non-Quantifiable Results

Three non-quantifiable impacts of the performance contracts have also been detected. The operational managers used the contracts as an argument to organise greater fluidity in inspection procedures with the operator of the container terminal, a request that had previously gone unanswered for over two years.

The inspectors became more 'diligent', in their own words and those of their superiors. The strong constraint of trade facilitation and the end of the competition to attract most declarations requires a more constant presence of inspectors in the office.

Finally, the relationships between the inspectors and their office heads have improved by making the actors more aware of their responsibilities. Having become accountable to the director general for their litigation results, the inspectors refused to assume responsibility for rerouting declarations at the request of their superiors. The superiors themselves reroute declarations where their information shows this to be necessary.

In January 2011, one year after the experiment's start, the director general promoted the four best officers of the two bureaus (office DP I and office DP V) to better positions, including two positions as bureau heads. The poorer performing officers were transferred to minor bureaus. The impact of this decision is very powerful. Indeed, all customs officers know that the government, and not the head of customs, has the authority to nominate an inspector to a position of bureau head. These appointments meant therefore that the government considered the performance contract results in making its decision. This has affected the perception of performance contracts within the port and the customs community itself. Consequently, many of the interviewed customs officers working in other bureaus now wish their bureau to be under performance contracts too.

After three years of introducing quantification systems and a policy of patient implementation, these results mark a positive change in Cameroon Customs.[19] Front-line inspectors have a precise account of their actions in the system, giving them evidence and justifications in case of blacklisting.

---

[19]This is confirmed by surveys of customs agents, in particular during the performance contracts experiment period.

6   TOWARDS A NEW PROFESSIONAL CULTURE?

This reform shows that the key role of the director general and his involvement is a strong component of the reforms linked to NPM. But, in this context, it also raises the question of how durable the policy may be once he leaves his post. The involvement of the organisation's senior managers is often essential to the success and implementation of a new culture (Behn 2002) to the extent that they are the main beneficiaries, having widened their authority from technical to management fields (Wholey and Hatry 1992; Franklin 2000; Julnes and Holzer 2001).

In Cameroon, the indicators and the performance contracts in customs have addressed the information asymmetry generally found in ministries of finances between headquarters and grassroots officials (Mascarenhas 1993; Raffinot 2001). At the same time, the contract's objectivity about an inspector's performance can support a manager in rejecting external requests to 'place' a protégé in an office with a high revenue potential. Despite these advantages, future directors general will only continue this quantification of performance under two conditions: if the quantification of action serves as a framework for establishing a new professional culture, and if this quantification is dynamic.

Next, we look at the conditions that shape a professional culture and examine whether a performance measurement system is needed. Performance measurement does not take root on virgin ground. Customs officials have formed professional associations and uphold an *esprit de corps* linked to their role in financing the developmental state (Cantens 2009). We must understand just how the performance contracts feed this professional culture.

First, one of the conditions for a professional culture is having a distinct professional identity (Elias 1950; Fisher 1966). Customs officials already have their own technical language, and the contracts reinforce this distinction: a common language of quantity, complex associations of technical terms, and a shared culture of presenting results in the form of graphs and tables, which become a vernacular of their own (Porter 1995).

Second, the contracts induce a new way of generating acceptable standards without disrupting hierarchical relations. The contracts recognise that administrative authority does not function well and depends mainly on the willingness of individuals to carry out orders. The contracts do not clash with this situation but exploit it. Based on the medians for recent years, the contractual thresholds calculate the behaviour of all, and thus establish a practical standard of behaviour which will be recognised and accepted because it is based on a median behaviour.

Third, the contracts strengthen the freedom to make decisions. Customs officials have the power to reach a compromise settlement for disputed claims, and the concept of 'risk analysis' is familiar to them. Faced with the size of flows and traders' demands for speed, the administration recognises that

it cannot counter all frauds or all forms of corruption comprehensively in controlling only those cargoes it deems to be high risk.

Thus, by emphasising two conflicting constraints—to release goods more rapidly and to impose more sanctions—inspectors must decide which cases should be investigated. Sociological analyses of public servants engaged in policing missions have shown that they prioritise their interventions, given that they cannot process all cases (Montjardet 1992, 1994; Favre 2001; Mouhanna 2001; Macci 2002).

Professional distinction, generation of 'acceptable' practical standards and freedom of decision-making are all conditions required for the development of a professional culture, but these same conditions raise the problem of the relationship with the law. Performance contracts, like all contracts, have the legal system in the background, which raises two questions.

First, the indicators meld fiscal policy into fiscal technique; they do not take into account that certain flows or operators are easier to tax than others. So, how can we consider the administration's efforts to extend the tax base of certain operators reputed to be difficult? Performance contracts set down global thresholds, with which it is assumed that inspectors in the field will somehow 'make do'.

Second, to what extent does this freedom exist within customs themselves? We are asking heads of customs offices to be 'managers' and distancing them from customs clearance functions, but every managerial function needs to be accompanied by a certain freedom of decision-making. However, this freedom is not guaranteed by any legal text, but rather linked to methods of appointment, which rely largely on political authority. Sociological issues combine with political issues—for instance, tribalism is invoked to explain why heads of customs offices are not allowed to choose their own subordinates, or that corruption prevents the development of preferential networks. Yet, just as inspectors and customs brokers are allowed to conduct customs clearance badly, should we not permit the heads of customs offices to make a poor choice of subordinates in order to judge their managerial capacity?

7   CONCLUSION

In this chapter we trace the Cameroon Customs reform from the introduction of the performance indicators to the preliminary measured results of performance contracts. We also demonstrate the positive, although preliminary, results of individual performance contracts implemented in two Douala port bureaus using indicators extracted from ASYCUDA. Several lessons can be drawn from the Cameroon case study.

Customs' performance contracts meet at the juncture of two different and controversial concepts—governance and new public management. Performance contracts penalise corruption and poor practice while distancing

themselves from any origins of corruption. As with a crime or offence (Crawford 2003), the contracts allow a return to the idea of a 'situation of governance' (Blundo 2002) and the need for empirical research; corruption is a question of opportunities and acts, not of predisposition or specific individuals. Given their use of a standard calculation method, the contracts therefore constitute a policy of corruption prevention and detection.

The fact that this is a pilot and not a vast, structured programme has two advantages. First, Cameroon Customs has controlled the risk that a major conceptual reform might pose to revenue collection. There was no question that the contracts would compromise the level of revenues collected. On the other hand, the hierarchy's commitment to reform is always questioned by grassroots agents (Behn 2002). The pilot imposed a rapid pace, which urged the head of customs to quickly demonstrate that she would fulfil her part of the contracts and convince officers under contracts that she was ready to grant them more flexibility: rewards as well as sanctions.

Wholey and Hatry (1992) defined four conditions for performance measurement:

- the right time (not necessarily when drawing up the financial balance sheet);
- a comparison (with the past or with an objective);
- a selection (it is not possible to measure everything);
- low cost.

In the case of Cameroon Customs, we have seen that timing was a key element in success: taking time to implement the contracts was the *leitmotif* of the directors general.

In addition, the indicators of the contracts have always been calculated on the basis of performances in previous years. The historic dimension is essential—any reform must also help clarify what change it is helping to bring about. And, finally, in terms of costs, all IT developments specific to performance measurement have been achieved using free software.

Establishing a policy of indicators and performance contracts has the advantage of giving more weight to the empirical knowledge of how the administration actually operates and of offering the framework for its own evaluation (Varone and Jacob 2004). However, this advantage is often perceived as a risk: quantification of public action is a leap into the unknown, and the heads of administration may be afraid of revealing the failings of their structure. In a way it is inevitable. The Cameroon pilot is interesting because it was done in a context where public servants were all labelled corrupt by the public. There was therefore little resistance on their part to investing in a performance culture that could prove this to be untrue and that could highlight the efforts made.

Even if the evaluation of the pilot performance contracts has not adopted the gold statistical standards because of the context, doing before–after com-

parisons is feasible and desirable. Performance contracts, as designed in the Cameroon case, are not replicable as they are in other contexts; they only demonstrate what is possible. We cannot advocate that there is no 'one size fits all' solution in development while, at the same, always be looking for simplistic replication. What is replicable—although it is not a new idea—is the introduction of performance measurement at individual levels based on the existence of IT systems to fight bad practices and make reforms effective on the ground.

After almost two years of work on this experiment, we can advocate the quantification of public action in customs (which is useful for impact evaluation) but we should also advocate the need for a qualitative understanding before, during and after the reform in order to evaluate change. From our point of view, this is a major difference with most impact evaluation of donors' interventions in trade areas. In customs reforms, as in any administrative reform, we have to understand what makes sense before designing any reform and evaluation plan. Therefore, empirical knowledge is a pre-requisite for a positive impact.

Thomas Cantens is a Researcher at the World Customs Organization.

Gael Raballand is the Senior Economist in the Public Sector Reform and Capacity Department, Africa Region at the World Bank.

Samson Bilangna is Head of the Cameroon Customs Information Technology Division.

Marcellin Djeuwo is Head of the Cameroon Customs Risk Analysis Unit.

REFERENCES

Behn, R. (2002). The psychological barriers to performance management: or why isn't everyone jumping on the performance-management bandwagon? *Public Performance & Management Review* **26**, 5–25.

Besley, T., and M. Ghatak (2005). Competition and incentives with motivated agents. *American Economic Review* **95**, 616–636.

Bilangna, S. (2009). La réforme des douanes camerounaises: entre les contraintes locales et internationales. *Afrique Contemporaine* **230**, 19–31.

Blundo, G. (2002). La gouvernance entre technique de gouvernement et méthode d'exploration empirique. Editorial, *Bulletin de l'APAD*, no 23–24, http://apad.revues.org/129.

Cantens, T. (2007). La réforme de la douane camerounaise à l'aide d'un logiciel des Nations Unies ou l'appropriation d'un outil de finances publiques. *Afrique Contemporaine* **223**, 289–307.

Cantens, T. (2009). Être chef dans les douanes camerounaises, entre titular chief, idéaltype et big katika. *Afrique Contemporaine* **230**, 83–100.

Cantens, T. (2011). Is it possible to reform a customs administration? The role of the customs elite on the reform process in Cameroon. UNU WIDER Working Paper 2010/118.

Cantens, T., G. Raballand, and S. Bilangna (2010). Reforming Customs by measuring performance: a Cameroon case study. *World Customs Journal* **4**(2), 55–74.

Considine, M., and J. M. Lewis. (2003). Bureaucracy, network, or enterprise? Comparing models of governance in Australia, Britain, the Netherlands, and New Zealand. *Public Administration Review* **63**(2), 131–140.

Crawford, A. (2003). Contractual governance of deviant behavior. *Journal of Law and Society* **30**, 479–505.

de Balzac, H. (1844). *Les Employés* (1985 edition). London: Folio Society.

Duflo, E., and M. Kremer (2003). Use of randomization in the evaluation of development effectiveness. Paper prepared for the World Bank Operations Evaluation Department (OED) Conference on Evaluation and Development Effectiveness in Washington, DC.

Elias, N. (1950). Studies in the genesis of the naval profession. *British Journal of Sociology* **1**(4), 291–309.

Farrington, D. (2003). British randomized experiments on crime and justice. *Annals of the American Academy of Political and Social Science* **589**, 150–167.

Favre, P. (2001). Around Dominique Monjardet's 'Sociologie de la force publique': recent books and articles in French in the field of sociology of the police. *Revue Française de Sociologie* **42**, (Supplement: An Annual English Selection 2001), 175–186.

Fisher, G. (1966). The foreign service officer. *Annals of the American Academy of Political and Social Science* **368**, 71–82.

Franklin, A. L. (2000). An examination of bureaucratic reactions to institutional controls. *Public Performance & Management Review* **24**(1), 8–21.

Julnes, P.d.L, and M. Holzer (2001). Promoting the utilization of performance measures in public organizations: an empirical study of factors affecting adoption and implementation. *Public Administration Review* **61**, 693–708.

Libom, M., T. Cantens, and S. Bilangna (2009). Gazing into the mirror: operational internal audit in Cameroon Customs. Discussion Paper 8, World Bank, Washington, DC.

Macci, O. (2002). Dominique Monjardet: ce que fait la police, sociologie de la force. *Annales Histoire Sciences Sociales* **57**(6), pp 1718–1721.

Mascarenhas, R. C. (1993). Building an enterprise culture in the public sector: reform of the public sector in Australia, Britain, and New Zealand. *Public Administration Review* **53**, 319–328.

Mbembe, A. (1999). Du gouvernement privé indirect. *Politique Africaine* **73**, 103–121.

Mbonji, E. (1999). Les 'déflatés' du développement. De la tradition de dépendance à l'autogestion. *Bulletin de l'APAD* **18**, http://apad.revues.org/455.

Merrien, F. X. (1999). La Nouvelle Gestion publique: un concept mythique. *Lien Social et Politiques* **41**, 95–103.

Monjardet, D. (1992). Quelques conditions d'un professionnalisme discipliné. *Déviance et Société* **16**, 399–403.

Monjardet, D. (1994). La culture professionnelle des policiers. *Revue Française de Sociologie* **35**, 393–411.

Mouhanna, C. (2001). Faire le gendarme: de la souplesse informelle a la rigueur bureaucratique. *Revue Française de Sociologie* **42**, 31–55.

Ogien, A. (2010). La valeur sociale du chiffre. La quantification de l'action publique entre performance et démocratie. *Revue Française de Socio-économie* **5**(1), 19–40.

Porter, T. M. (1995). *Trust in Numbers: The Pursuit of Objectivity in Science and Public Life.* Princeton University Press.

Raffinot, M. (2001). 'Motiver' et 'chicoter': l'économie politique de la pression fiscale en Afrique subsaharienne. *Autrepart* **20**, 91–106.

Rouban, L. (1998). Les états occidentaux d'une gouvernementalité à l'autre. *Critique Internationale* **1**, 131–149.

Spanou, C. (2003). Abandonner ou renforcer l'état wébérien. *Revue Française d'Administration Publique* **1-2**(105), 109–120.

Strathern, M. (2001). Blowing hot and cold. *Anthropology Today* **17**(1), 1–2.

Turner, H., R. Boruch, A. Petrosino, J. Lavenberg, D. de Moya, and H. Rothstein (2003). Populating an international web-based randomized trials register in the social, behavioral, criminological, and education sciences. *Annals of the American Academy of Political and Social Science* **589**, 203–223.

Varone, F., and S. Jacob (2004). Institutionnalisation de l'évaluation et nouvelle gestion publique: un état des lieux comparatifs. *Revue Internationale de Politique Comparée* **11**, 271–292.

Wholey, J., and H. Hatry (1992). The case for performance monitoring. *Public Administration Review* **52**, 604–610.

World Bank (2005). Bâtir des états performants. Créer des sociétés engages. Groupe de travail sur le renforcement des capacités en Afrique. Working Paper, World Bank, Washington, DC.

Ymbert, J.-G. (1825). *Mœurs administratives*, Volume 1. Paris: Ladvocat.

Aid for Trade and Export Performance:
The Case of Aid in Services

ESTEBAN FERRO, ALBERTO PORTUGAL-PÉREZ AND JOHN S. WILSON[1]

1   INTRODUCTION

The response of developed countries to the unprecedented global recession has placed their budgets under enormous strain. Unfortunately, foreign aid tops the list of expenses to be cut in donor countries, particularly in the eyes of taxpayers. As donors' willingness to provide foreign aid could fall, even if its need remains vital, there is a greater necessity to identify the projects that more efficiently pursue the intended goals.

Aid for trade has rapidly gained importance in trade and development circles, as well as in the donor community. Despite enjoying preferential market access and facing lower tariffs, several developing countries (especially least developing countries) have seen their share of world exports diminish over the past few years. Clearly, the reduction of tariff and non-tariff barriers is an essential condition for export growth, but not a sufficient one. These countries face supply-side constraints that severely limit their ability to reap the benefits from global trade integration.

Launched at the Hong Kong WTO Ministerial Conference in December 2005, the Aid for Trade (AfT) Initiative aims at helping developing countries, particularly least developing countries (LDCs), to overcome their supply-side constraints and expand trade. Despite the clearness of this main goal, its assessment has been problematic. Evidence on the positive relationship of AfT and export performance has been found by estimating equations where exports variables are on the left-hand side, and AfT variables are on the right-hand side, as well as other covariates.

However, potential reverse causality may arise, as AfT may also be determined by trade performance. For instance, if better performing countries tend to receive more aid, estimates of AfT coefficients would be biased upwards.

In this chapter, we propose a new identification strategy that overcomes the reverse-causality problem related to the literature on aid. Using input–output data, we exploit the differential service intensities of industrial sectors to evaluate the impact of aid in five service sectors (transport, communications, energy, banking/financial services and business services) on exports of downstream manufacturing sectors for 106 countries between 1990 and 2008. Our results show that assistance to the energy and banking sectors has consistently had a significant and positive impact on downstream manufacturing exports.

The rest of the chapter is organised as follows: in Section 2 we provide a review of the literature on aid and AfT, and in Section 3 we explain our identification strategy. In Section 4 we describe our data, and then present the results in Section 5. Finally, in Section 6 we give conclusions and discuss potential avenues for further research.

2   LITERATURE REVIEW

Our review is divided into two parts. We start with a brief account of the extensive literature on aid and growth. The second part defines briefly aid for trade and reviews the literature that is more specific to it.

2.1   Aid and Growth

The literature on the impact of aid on recipient countries' level of development is large and keeps growing; and we provide a small and selective review in this chapter. Unfortunately, the results of the studies have been mixed.[2] There is no robust evidence of either a positive or negative correlation between foreign aid inflows and the economic growth of poor countries.

Most studies that find a positive effect between aid and growth show that this relation holds only under specific conditions. For instance, Burnside and Dollar (2000) stipulate that the quality of institutions in the recipient country is key for the effectiveness of aid. Dalgaard *et al* (2004) find that location (geography) is also an important determinant of aid effectiveness. Angeles and Neanidis (2009) show that the structure of social elites plays an important part in shaping the effectiveness of aid. Clemens *et al* (2004) study the effect of a specific type of aid, and discover that short term aid—including budget and balance-of-payments support, investments in infrastructure and aid for productive sectors such as agriculture and industry—has a significant positive effect on growth.

---

[2]See for example Burnside and Dollar (2000); Collier and Dollar (2002); Easterly (2003); Easterly *et al* (2003); Clemens *et al* (2005); Bourguignon and Sundberg (2007); Rajan and Subramanian (2008).
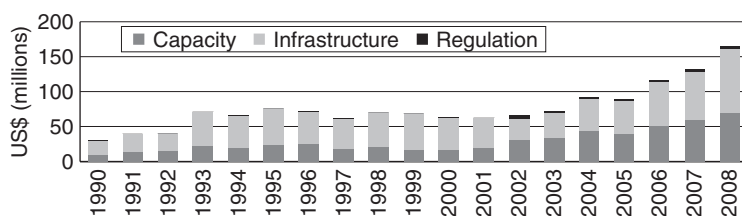
Figure 8.1: *Aid for trade 1990–2008.*
Source: authors' own calculations using OECD's CRS database.

However, Rajan and Subramanian (2011) argue that the costs emanating from foreign aid offset its benefits. The authors find that aid has a Dutch disease effect on the terms of trade of recipient countries resulting in a negative impact on tradables and on growth.[3]

The existing literature also stresses the reverse causality between aid and growth that could lead to misleading results. A few studies attempt to address this problem using instrumental variables. Rajan and Subramanian (2008) estimate a bilateral aid specification including variables traditionally used in the gravity model of bilateral trade (*ie* common language and colony relationships, among others) to generate a fitted aid measure (first stage) that is used as an instrument for aid in a GDP growth equation (second stage). They find little robust evidence of a positive (or negative) relationship between aid and GDP growth.

Bruckner (2011) adopts a different two-stage strategy. In the first stage, he estimates an equation explaining aid using rainfall, commodity price shocks, GDP growth and other controls. In the second stage, the fitted residuals are used as an instrument for aid on a per capita GDP growth specification. Bruckner finds that aid has a significant and positive effect on real per capita GDP growth.

Finally, Bourguignon and Sundberg (2007) point to the wide heterogeneity of aid motives and the limitations of the tools of analysis. They suggest that the complex causality chain linking external aid to final outcomes has been handled mostly as a kind of 'black box', and that progress on estimating rigorously aid effectiveness requires opening that box.

2.2   Aid for Trade

According to the WTO, aid for trade aims to help developing countries, particularly least-developed countries, develop the trade-related skills and infrastructure that are needed to implement and benefit from WTO agreements

---

[3]Some other studies of aid and exchange rates include Younger (1992); Arellano *et al* (2005); Berg *et al* (2005); Prati and Tressel (2006).

and to expand their trade.[4] Aid for trade, as defined by the OECD, nearly tripled between 2001 and 2008, as shown in Figure 8.1; however, evidence of its effectiveness is still scant.

A limited number of studies focus on the impact of aid for trade.[5] Brenton and von Uexkuhll (2009) analyse the effectiveness of export-development programmes. Using a difference-in-differences approach, they aim at isolating the impact of the policy interventions and draw four main conclusions.

- Most export-development programmes have coincided with or pre-dated stronger export performance.
- Such programmes appear to be more effective where there is already significant export activity.
- There is some concern about the 'additionality' of the programmes as support may be channelled to sectors that would have prospered anyway.
- Conclusions strongly depend on what we postulate would have happened in the absence of the policy intervention, so the definition of a credible counterfactual is critical for the evaluation of technical assistance for exports.

Helble *et al* (2009) made one of the first attempts to analyse how foreign aid spent on trade facilitation increases trade flows in developing countries. The authors used a gravity model of bilateral trade and found that the bulk of the relationship between aid and trade appears to come from a narrow set of aid flows directed towards trade policy and regulatory reform, rather than broader AfT categories directed toward sectoral trade development or infrastructure development. Other studies on the effect of aid on trade have found similar positive results to those found by Helble *et al* (2009), including Calì and te Velde (2011), the latter being the closest study to our own.

Indeed, Calì and te Velde (2010) evaluate whether AfT has improved export performance. They find that AfT facilitation, and to some extent AfT policy and regulations, helps to reduce the cost of trading (in terms of both exports and imports). In addition, their results suggest that aid to economic infrastructure increases exports, whereas aid to productive capacity appears to have no significant impact on exports. They correctly point out that AfT is possibly endogenous to exports, particularly aid to productive capacity. For instance, if better performing sectors tended to receive more AfT than other sectors, this would generate an upward bias in the aid coefficient.

To address this endogeneity, Calì and te Velde instrument AfT with indexes measuring both the degree of respect for civil and political liberties compiled by Freedom House (2009) and the 'affinity of nations' compiled by Gartzke

---

[4]See http://www.wto.org/english/tratop_e/devel_e/a4t_e/aid4trade_e.htm.

[5]For and extended review of the literature on the relationship of aid and trade, see Suwa-Eisenmann and Verdier (2007).

(2009). They argue that many donors choose recipients based on development and democratic measures like those captured by those indexes that correlate with aid but not with exports. However, they acknowledge that their instruments are not appropriate for sectoral analysis, as they vary only across country and year, and not across country, sector and year.

We propose an alternative identification strategy that allows us to analyse the impact of sectoral aid. To address reverse causality, we exploit the links between the service sector and the manufacturing sector relying on input–output tables. Although input–output data has not been used in the analysis of aid effectiveness so far, we found a number of studies, particularly in the FDI literature, that use such data. Given the difficulty of finding consistent input–output matrices across countries, most studies rely on input–output data from the USA to describe the technological possibilities of firms in a given economy. Acemoglu *et al* (2009) investigate how contracting costs and financial development determine the extent of vertical integration across countries. Alfaro and Charlton (2009) use new firm-level data at the four-digit sector level and US input–output tables to classify firms between horizontal and vertical subsidiaries.

The authors find that, at the two-digit industry level, there are considerably more horizontal (subsidiaries in the same industry as their parents) than vertical (subsidiaries that supply their parents with inputs) FDI. However, disaggregating to the four-digit level reveals that many of the foreign subsidiaries in the same two-digit industry as their parents are, in fact, located in sectors that produce highly specialised inputs to their parents' production. Thus, contrary to the conventional wisdom, the authors find that the number of vertical multinational subsidiaries is larger than commonly thought.

Among the few studies using input–output data for countries other than the USA, Hummels *et al* (2001) define a measure of vertical specialisation that captures a country's role in the fragmentation of production into multiple stages in multiple locations. They use input–output tables from ten OECD nations and four other countries to measure a country's vertical specialisation as its exports weighted by the share of imported inputs in its total output. Trefler and Zhu (2010) use 20 input–output tables from the Global Trade Analysis Project (GTAP) to reassess Vanek's (1968) factor content of trade predictions in 41 countries.

3   TACKLING REVERSE CAUSALITY: AN IDENTIFICATION STRATEGY

Aid flows are expected to have an impact on exports, but a country's exports may also affect the aid the country receives. It is plausible that donors target industries in recipient countries where exports are expanding or declining (see, for example, Brenton and von Uexkuhll 2009).

To address this issue, we propose a new identification strategy. Instead of directly analysing how aid that targets a specific manufacturing sector affects

its own exports, we analyse how aid that targets services—such as banking and energy services—affects exports of downstream manufacturing making use of those services. The following offers an example to illustrate this idea.

The USA is likely to grant a loan to Ecuador to continue developing the cut-flower industry—a major source of export revenue for that country—which is likely to result in greater exports. In this case, exports determine aid flows, and aid also affects exports.

But the USA is less likely to grant a loan to the Ecuadorian government to develop the banking or energy sector in order to build up the cut-flower industry in the country. In this case exports would not determine aid flows even though aid to these services will indirectly have an impact on exports not only of cut flowers but of all other industries. This link between services and manufacturing will help us to quantify the impact of aid on exports.

For our identification strategy, we need to control not only for the aid received by each service sector but also for how intensively each manufacturing sector uses the different services. We use the total input requirements matrix to measure the service intensity of each manufacturing sector. Thus, exports are determined by

$$\ln X_{ijt} = \alpha_{ij} + \gamma_{it} + \delta_{jt} + \sum_k \beta_k (\ln \text{aid}_{ikt}\, \text{intensity}_{ijk}) + \varepsilon_{ijt} \qquad (8.1)$$

where X_{ijt} is exports of sector j in country i in year t, aid_{ikt} is the amount of aid in service sector k received by country i in year t and intensity_{ijk} is the intensity with which the manufacturing sector j uses service sector k in country i. We also include country–sector effects, α_{ij}, to control for factors such as taxes or subsidies specific to a manufacturing sector in a given country. Country–year effects, γ_{it}, control for inflation, exchange rates, political or economic shocks and climate shocks such as natural disasters. Sector–year effects, δ_{jt}, control for shocks specific to a product worldwide in a given year, such as any supply or demand shock having an impact on world market prices.

4   DATA DESCRIPTION

Data is compiled from three main sources. Aid flow data were gathered from the OECD's Creditor Reporting System (CRS); exports, specifically mirrored imports, were taken from UNCTAD's Comtrade database; and input–output tables were compiled from the US Bureau of Economic Activity and from Argentina's INDEC. Since each of these sources uses different classification schemes, we first merged all databases together through the concordances described in Table 8.1, which also describes the five input service sectors and the nine manufacturing sectors used in our analysis. The final sample consists of 106 developing countries over the period 1990–2008. Table 8.2 lists all the countries in the sample.

Table 8.1: *Databases' concordances.*

Sector	Industry	CRS (aid) purpose code	COMTRADE (exports) SITC	US I-O 2008 (input intensities) NAICS
Agriculture	Agriculture	311	11	111CA
Forestry & fishing	Forestry & fishing	312, 313	12, 13	113FF
Mining	Mining	322	21, 22, 23, 29	211, 212, 213
Construction	Construction	323		23
Manufacturing	Agro-industries	32161	31	311FT
	Forest industries	32162	33, 34	321, 322, 323, 337
	Textile	32163	32	313TT, 315AL
	Chemicals	32164, 32165, 32168	35	324, 325, 326
	Non-metallic mineral prod.	32166	36	327
	Primary metals	32169, 32170	371, 372	331, 332
	Machinery	32171	382, 383, 385	333, 334,335,
	Transport equipment	32172	384	3361MV, 3364OT
	Other manufacturing	Other under sector 321	381, 39	339
Services	Transportation	210		481, 482, 483, 484, 485, 486, 487OS, 493
	ICT	220		511, 512, 513, 514
	Energy	230		22
	Banking	240		521CI, 523, 524, 525
	Business	250		5411, 5415, 5412OP, 55, 561

Table 8.2: *Sample of countries.*

|			
Algeria	Cyprus	Lebanon	Saudi Arabia
Angola	Djibouti	Lesotho	Senegal
Antigua & Barbuda	Dominica	Liberia	Seychelles
Argentina	Dominican Rep.	Libya	Sierra Leone
Bahrain	Ecuador	Madagascar	Solomon Islands
Bangladesh	Egypt, Arab Rep.	Malawi	South Africa
Barbados	El Salvador	Malaysia	Sri Lanka
Belize	Equatorial Guinea	Maldives	St. Kitts & Nevis
Benin	Fiji	Mali	St. Lucia
Bhutan	Gabon	Mauritius	Vincent & Grenadines
Bolivia	Gambia, The	Mexico	Sudan
Botswana	Ghana	Morocco	Suriname
Brazil	Grenada	Mozambique	Swaziland
Burkina Faso	Guatemala	Namibia	Tanzania
Burundi	Guinea	Nepal	Thailand
Cameroon	Guinea-Bissau	Nicaragua	Togo
Cape Verde	Guyana	Niger	Tonga
Central African Rep.	Honduras	Nigeria	Trinidad & Tobago
Chad	India	Oman	Tunisia
Chile	Indonesia	Pakistan	Turkey
China	Israel	Panama	Uganda
Colombia	Jamaica	Papua New Guinea	Uruguay
Comoros	Jordan	Paraguay	Vanuatu
Congo, Dem. Rep.	Kenya	Peru	Vietnam
Congo, Rep.	Kiribati	Philippines	Zambia
Costa Rica	Korea, Rep.	Rwanda	
Côte d'Ivoire	Lao PDR	Samoa	

The key ingredient in our analysis is the link between inputs and outputs across sectors in an economy, as described in the input–output matrix. We use the total input requirement matrix of the USA for 2008. We are interested in the total effect as any change in a service sector will also affect all the other inputs of any manufacturing sector.[6] The total effect can be defined as the *direct* and *indirect* effects. The total requirement matrix is estimated using a simple multiplier, $(I - A)^{-1}$, where I is the identity matrix and A is a matrix of direct input coefficients. Figure 8.2 summarises the service intensities of each manufacturing sector.

5   RESULTS

Table 8.3 displays the results of estimating Equation (6.1) using OLS. Columns (1)–(5) show the individual effect that aid to each service sector has on man-ufacturing exports. Aid to energy and banking are the only two types of

---

[6]For a detailed discussion on input–output analysis, see Miller and Blair (2009).
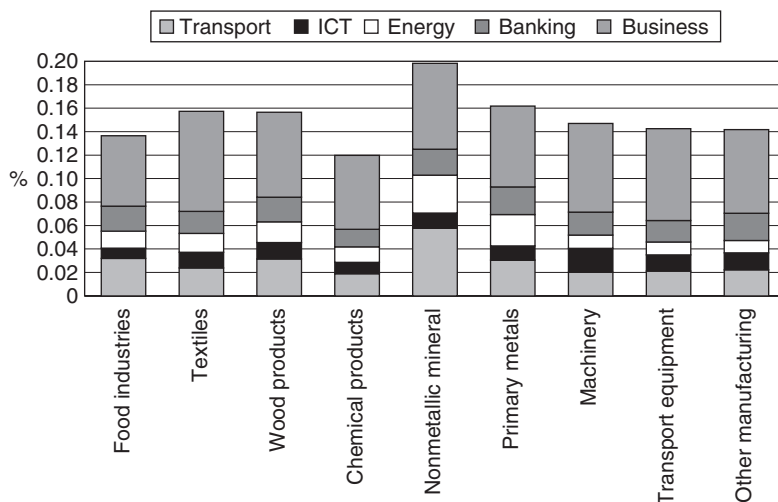
Figure 8.2: *Service intensities by manufacturing sector.*
Source: authors' own calculations using Bureau of Economic Analysis (BEA) 2008 input–output tables.

Table 8.3: *Impact of aid to services on manufacturing exports.*

	(1)	(2)	(3)	(4)	(5)	(6)
(transport_intensity) × (aid_transport)	0.21 [0.154]					0.13 [0.155]
(ICT_intensity) × (aid_ICT)		−0.259 [0.383]				−0.231 [0.381]
(energy_intensity) × (aid_energy)			0.549** [0.243]			0.483* [0.247]
(banking_intensity) × (aid_banking)				1.018* [0.556]		0.912* [0.560]
(business-services_intensity)) × (aid_ business-services					0.351 [0.222]	0.371* [0.220]
Observations	17,678	17,678	17,678	17,678	17,678	17,678
R^2	0.95	0.95	0.95	0.95	0.95	0.950

Notes: robust standard errors are given in square brackets; *significant at 10%, **significant at 5%; ***significant at 1%. Dependent variable is ln(exports). Service sector intensities are estimated using US total input requirements. All regressions control for country–sector, country–year and sector–year effects.

aid to services that are significantly associated with higher levels of exports in manufacturing. The coefficient of information and communications technologies (ICT) is negative but it is statistically non-significant. Column (6) in Table 8.3 includes simultaneously the variables of aid to all service sectors.

Table 8.4: *Robustness checks.*

	Baseline (1)	No country–sector effect (2)	Year >1999 (3)	Argentinean intensities (4)
(transport_intensity) × (aid_transport)	0.13 [0.155]	0.817*** [0.169]	0.343 [0.267]	0.015 [0.168]
(ICT_intensity) × (aid_ICT)	−0.231 [0.381]	−0.959* [0.569]	−1.617** [0.807]	0.156 [1.493]
(energy_intensity) × (aid_energy)	0.483* [0.247]	1.716*** [0.269]	0.819** [0.393]	0.491*** [0.183]
(banking_intensity) × (aid_banking)	0.912* [0.560]	2.482*** [0.788]	2.691*** [0.906]	3.453** [1.452]
(business-services_intensity) × (aid_ business-services)	0.371* [0.220]	1.349*** [0.334]	−0.211 [0.342]	−0.679*** [0.252]
Observations	17,678	17,678	17,678	9,480
R^2	0.95	0.82	0.95	0.96

Notes: robust standard errors are given in square brackets. *significant at 10%; **significant at 5%; ***significant at 1%. Dependent variable is ln(exports). Service sector intensities are estimated using US total input requirements, except for column (4), where Argentina's total input requirements are used. All regressions control for country–sector, country–year and sector–year effects, except for column (2), where no country–sector effect is included.

Multicollinearity among sector variables does not seem to be an issue, as all coefficients retain their magnitude and significance compared with the first set of regressions. Coefficients for aid to energy and banking remain positive and significant, whereas the coefficient for aid to business services remains positive but is now statistically significant at the 10%. Aid that targets the three services is positively associated with a higher level of exports in manufacturing.

   We test the robustness of our results to a number of different specifications, and results are presented in Table 8.4. In column (2) we drop the country–sector effects ($106 \times 9 = 954$ dummy variables) as we fear that our model is over-determined and driven by too many dummy variables. When we exclude country–sector dummies, the effect of aid becomes larger in absolute terms. However, the conclusions from the baseline regression remain unchanged— aid to energy, banking and business services is associated with better export performance in manufacturing.

   To check whether aid has been more effective in the last decade, we restrict the sample to the period between 1998 and 2008. Column (3) of Table 8.4 displays the results. In this case ICT is negative and statistically significant. On the other hand, business services become insignificant. Aid to the energy and banking sector is still significantly correlated with higher levels of exports.

Since our sample includes only developing nations, the US input–output matrix may incorrectly depict their economies. We replace the service sector intensities obtained from the US input–output tables with intensities obtained from Argentina's 1997 input–output tables. Column (4) of Table 8.4 exhibits the results. Aid to the business sector is now significantly associated with lower levels of exports. This result is driven by the fact that Argentina's economy, compared with the USA's economy, is less intensive in the use of services, particularly business related services. However, aid to energy and banking remains positive and significant, confirming that aid to these two sectors is associated with higher levels of manufacturing exports.

6   CONCLUSIONS

Evaluating trade-related projects is a challenging task. Rigorous methods of impact evaluation used more extensively in education, cash transfers and health programmes are less easily implementable in trade projects. In the absence of micro-level impact evaluation, macro-level evaluations can provide a general assessment on whether aid has had the expected impact on a specific variable. Of course, macro-level exercises are not free from econometric evils, such as endogeneity, and results have to be taken carefully for policy recommendations.

We propose a new identification strategy to measure the impact of aid on sector exports. We use the links that exist between inputs and outputs in an economy to analyse how aid to services affects exports in the manufacturing sector. We feel that our estimating strategy is econometrically sound, and thus our results have clear policy implications. We find that aid to the energy and banking sectors is the most effective when the objective is to increase exports of a recipient country. Our estimations show a positive correlation between aid to these two service sectors and higher levels of exports. These results are robust to a number of different specifications. We also find that aid to the business sector is positive and significant in most of our specifications, but it is not nearly as robust as aid to energy and banking.

Future research can improve on our current work. It is important to use input–output tables for as many countries as possible in our sample, as input–output intensities do vary across countries. Assuming that all countries have the same technology, as we did in this study, is a simplification because of data availability. But the validity of our results may not be robust to a better representation of the input–output linkages.

In addition, we have not exploited all the input–output linkages available from the input–output tables. In future research, we expect to estimate the impact of aid in inputs other than services on exports of downstream goods. Among other possible avenues for future research, we would like to expand the sample of countries and test the effect of aid for different subsamples of countries.

Finally, we stress that results presented here are targeted at stimulating discussion and helping policymakers and stakeholders arrive at a tentative prioritisation of their efforts regarding AfT. More detailed analysis involving rigorous impact evaluation of specific projects covering costs and benefits for a developing country can help to spend the next million dollars in a wiser way.

Esteban Ferro is a Consultant in the Trade and International Integration Team of the Development Research Group at the World Bank.

Alberto Portugal-Pérez is an Economist in the Trade and International Integration Team of the Development Research Group at the World Bank.

John S. Wilson is the Lead Economist in the Trade and International Integration Team of the Development Research Group at the World Bank.

REFERENCES

Acemoglu, D., S. Johnson, and T. Mitton (2009). Determinants of vertical integration: financial development and contracting costs. *Journal of Finance* **64**, 1251–1290.

Alfaro, L., and A. Charlton (2009). Intra-industry foreign direct investment. *American Economic Review* **99**, 2096–2119.

Angeles, L., and K. Neanidis (2009). Aid effectiveness: the role of the local elite. *Journal of Development Economics* **90**, 120–134.

Arellano, C., A. Bulir, T. Lane, and L. Lipschitz (2005). The dynamic implications of foreign aid and its variability. IMF Working Paper 05/119. International Monetary Fund, Washington, DC.

Berg, A., M. Hussain, S. Aiyar, S. Roache, and A. Mahone (2005). *The Macroeconomics of Managing Increased Aid Inflows: Experiences of Low-Income Countries and Policy Implications* Washington, DC: International Monetary Fund.

Bourguignon, F., and M. Sundberg (2007). Aid effectiveness: opening the black box. *American Economic Review* **97**, 316–321.

Brenton P., and E. von Uexkuhll (2009). Product specific technical assistance for exports: has it been effective? *The Journal of International Trade and Economic Development* **18**, 235–254.

Bruckner, M. (2011). On the simultaneity problem in the aid and growth debate. Research Paper 2011-01, University of Adelaide School of Economics.

Burnside, C., and D. Dollar (2000). Aid, policies, and growth. *American Economic Review* **90**, 847–868.

Calì, M., and D. te Velde (2011). Towards a quantitative assessment of aid for trade. *World Development*, forthcoming.

Clemens, M., S. Radelet, and R. Bhavnani (2004). Counting chickens when they hatch: the short-term effect of aid on growth. Working Paper 44, Center for Global Development, Washington, DC.

Collier, P., and D. Dollar (2002). Aid allocation and poverty reduction. *European Economic Review* **46**, 1475–1500.

Dalgaard, C., H. Hansen, and F. Tarp (2004). On the empirics of foreign aid and growth. *The Economic Journal* **114**, 191–216.

Easterly, W. (2003). Can foreign aid buy growth? *Journal of Economic Perspectives* **17**, 23–48.

Easterly, W., R. Levine, and D. Roodman (2003). New data, new doubts: revisiting aid, policies, and growth. Working Paper 26, Center for Global Development, Washington, DC.

Freedom House (2009). Freedom in the World country ratings (online database). http://www.freedomhouse.org.

Gamberoni, E., and R. Newfarmer (2009). Aid for trade: matching supply and demand. Working Paper 4991, World Bank Policy Research.

Gartzke, E. (2009). The Affinity of Nations Index 1946–2002, version 4.0 and related documents. http://dss.ucsd.edu/˜egartzke/htmlpages/data.html.

Helble, M., C. L. Mann, and J. S. Wilson (2009). Aid for trade facilitation. Working Paper 5064, World Bank Policy Research.

Hummels, D., J. Ishii, and K. Yi (2001). The nature and growth of vertical specialization in world trade. *Journal of International Economics* **54**, 75–96.

Miller, R., and P. Blair (2009). *Input–Output Analysis Foundations and Extensions*. Cambridge University Press.

Mitton, T. (2008). Institutions and concentration. *Journal of Development Economics* **86**, 367–394.

Prati, A., and T. Tressel (2006). Aid volatility and Dutch disease. Is there a role for macroeconomic policies? Mimeo, International Monetary Fund, Washington, DC.

Radelet, S., M. Clemens, and R. Bhavnani (2006). Aid and growth: the current debate and some new evidence. In P. Isard, L. Lipschitz, A. Mourmouras and B. Yontcheva (eds), *The Macroeconomic Management of Foreign Aid: Opportunities and Pitfalls*, pp 43–60. Washington, DC: International Monetary Fund.

Rajan, R., and A. Subramanian (2008). Aid and growth: what does the cross-country evidence really show? *The Review of Economics and Statistics* **90**, 643–665.

Rajan, R., and A. Subramanian (2011). Aid, Dutch disease, and manufacturing growth. *Journal of Development Economics* **94**, 106–118.

Rajan, R., and L. Zingales (1998). Financial dependence and growth. *American Economic Review* **88**, 559–586.

Suwa-Eisenmann, A., and T. Verdier (2007). Aid and trade. *Oxford Review of Economic Policy* **23**, 481–507.

Trefler, D., and S. Zhu (2010). The structure of factor content predictions. *Journal of International Economics* **82**, 195–207.

Vanek, J. (1968). The factor proportions theory: the *N*-factor case. *Kyklos* **21**, 749–756.

Younger, S. (1992). Aid and Dutch disease: macroeconomic management when everybody loves you. *World Development* **20**, 1587–1597.

"A welcome trend is emerging towards more clinical and thoughtful approaches to addressing constraints faced by developing countries as they seek to benefit from the gains from trade. But this evolving approach brings with it formidable analytical challenges that we have yet to surmount. We need to know more about available options for evaluating Aid for Trade, which interventions yield the highest returns, and whether experiences in one development area can be transplanted to another. These are some of the issues addressed in this excellent volume."
Pascal Lamy, Director-General, World Trade Organization

"Five years into the Aid for Trade project, we still need to learn much more about what works and what does not. Our initiatives offer excellent opportunities to evaluate impacts rigorously. That is the way to better connect aid to results. The collection of essays in this well-timed volume shows that the new approaches to evaluation that we are applying to education, poverty, or health programs can also be used to assess the results of policies to promote or assist trade. This book offers a valuable contribution to the drive to ensure value for aid money."
Robert Zoellick, President, The World Bank