No 2013-33 – October

Working Paper

CEPII

Evaluating Aid for Trade: A Survey of Recent Studies

Olivier Cadot, Ana Fernandes, Julien Gourdon,
Aaditya Mattoo & Jaime de Melo

Highlights

- The demand for accountability in "Aid-for-Trade" (AFT) is increasing but monitoring has focused on case-studies and impressionistic narratives.
- The literature based on traditional econometric analysis has theoretic foundations but the wide spectrum of results reveals the difficulty of drawing robust policy conclusions because of confounding influences.
- Impact evaluation techniques provide "internal validity" as confounding influences can be controlled better.

RESEARCH AND EXPERTISE
ON THE WORLD ECONOMY

Abstract

The paper reviews recent evidence from a wide range of studies, recognizing that a multiplicity of approaches is needed to learn what works and what does not. The review concludes that there is some support for the emphasis on reducing trade costs through investments in hard infrastructure (like ports and roads) and soft infrastructure (like customs). But failure to implement complementary reform – especially the introduction of competition in transport services – may erode the benefits of these investments. Direct support to exporters does seem to lead to diversification across products and destinations, but it is not yet clear that these benefits are durable. In general, it is difficult to rely on cross-country studies to direct AFT. More rigorous impact evaluation (IE) is an under-utilized alternative, but situations of "clinical interventions" in trade are rare and adverse incentives (due to agency problems) and costs (due to the small size of project) are a hurdle in implementation.

Keywords

Aid for trade (AFT); trade performance; gravity; impact evaluation.

JEL

F15; F35.

Evaluating Aid for Trade: A Survey of Recent Studies[1]

Olivier Cadot, Ana Fernandes, Julien Gourdon
Aaditya Mattoo & Jaime de Melo[2]

1. Introduction

Since the Paris declaration of 2006, calling for an expansion of Aid-for-Trade (AFT) funding to reduce trade costs, a WTO AFT task force was set up to implement this 'positive agenda' to enhance competitiveness. Multiple goals were adopted, but clear guidelines on how to conduct evaluations were largely absent.[3] Evaluation has progressed slowly from accountability (making sure that infrastructure has been built) to outcomes (has performance improved), but no agreement has been reached so far as to the main yardsticks to be used to measure outcomes. Progress has also been slowed by donors (multilateral, bilateral and NGOs) using different evaluation frameworks, by lack of information, and by context-specificity. So far three biennial reviews have produced a useful discussion of approaches and methods[4] and a digest of a large collection of projects and case stories--many voluntarily supplied--feeding into meta-evaluations that have not yielded significant insights.[5]

This paper provides a selective review of some recent evidence. We begin in Section 2 with studies that examine the impact of aid directly on trade (Figure 1 shows how we decompose the channels of AFT's impact). Credible identification is a challenge and overall there are

[1] An earlier version of this paper entitled "Evaluation in Aid-for-Trade: From Case-study Counting to Measuring" was presented at the FERDI-ITC- World Bank workshop «Aid for Trade: What Do We Know ? Which Way Ahead?», Geneva, December 6, 2012. We thank Céline Carrère, Bernard Hoekman and seminar participants for comments on the earlier draft. Cadot and de Melo gratefully acknowledge support from France's Agence Nationale de la Recherche under "Investissement d'Avenir" grant ANR-10-LABX-14-01 and from Switzerland under NCCR work package 6, "Impact assessment".

[2] University of Lausanne, FERDI and CEPR; World Bank; CEPII; World Bank; FERDI and CEPR

[3] According to OECD (2011), the AFT agenda has been classified under six categories: (i) trade policy and regulation; (ii) trade development; (iii) trade-related infrastructure; (iv) building productive capacity; (v) trade-related adjustment; (vi) other trade-related needs. According to OECD (2011), 80% of donors use the DAC principles for evaluating programmes and projects

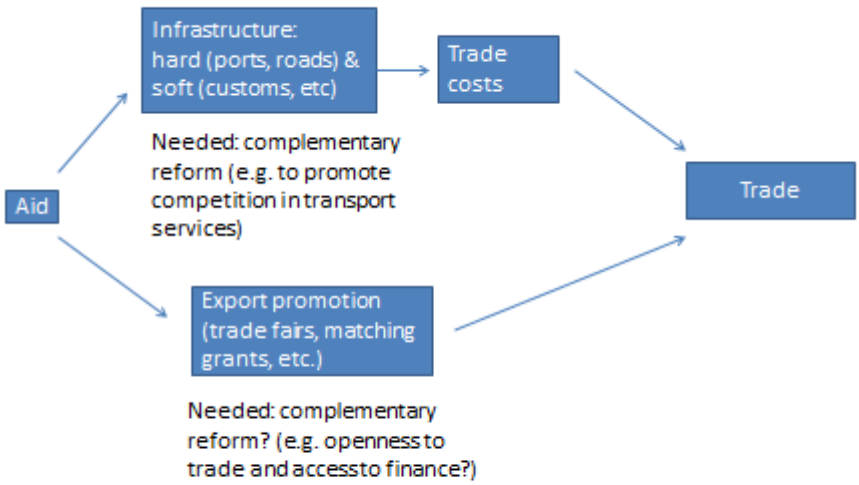
[4] To respond to the quest for accountability, the task force calls for 'managing for development results (MfDR) along a 'results chain'.

[5] For example, the meta-evaluation of 162 projects in Ghana and Vietnam (not all with a trade emphasis) revealed that what matters most for policy makers (terms like "imports", "exports" or "regulatory reform") were rarely mentioned. It also highlighted that project evaluators often lacked the baseline data against which to measure progress. See OECD (2009) and OECD (2011).

few convincing results that aid matters for trade. We then turn to evidence on the two main channels through which AFT could be expected to have an impact on trade (see Figure 1 below). In Section 3, we take a closer look at the first channel, through a reduction in trade costs. We examine the drivers of trade costs, focusing on the importance of improvements in hard infrastructure (such as ports and roads) and soft infrastructure (such as customs regulations and procedures), both of which have benefitted from AFT. A key insight here is that still-elusive complementary reform – particularly the introduction of greater competition in transport services – is needed to reap the full benefits of investment in infrastructure in terms of reduced trade costs.

In Section 4, we turn to the second channel of potential impact – through direct support to exporters. It is this class of directed assistance which is most amenable to rigorous impact evaluation because it may be feasible to distinguish between beneficiaries (the "treatment group") and non-beneficiaries (candidates for the "control group"). The few studies that follow this route suggest that AFT may indeed stimulate durable diversification but does not seem to have a durable effect on total exports of beneficiary firms. The review also highlights the inescapable trade-off between "internal validity" (the ability to identify impact effects net of confounding influences), which improves as one goes from (usually aggregate level) cross-country studies to impact evaluations, and "external validity" (the ability to draw general policy propositions from evaluation results) which may well worsen.

Figure 1 – Aid for Trade: Channels of Intended Impact



2. A direct impact of aid on trade?

We begin by considering the evidence of a link directly between aid and trade rather than through any specific transmission mechanism.

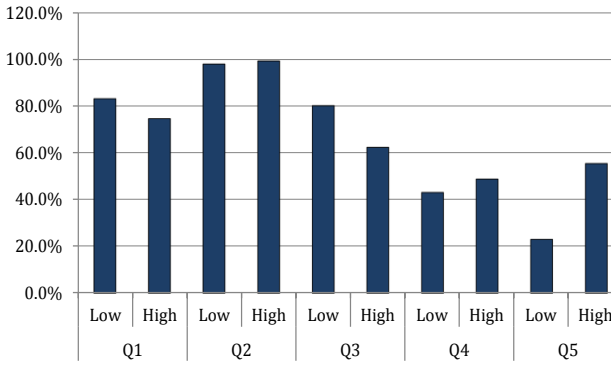
2.1. Prima-facie evidence

As a first pass, we ask whether there is any correlation between export growth and lagged AFT commitments. Figure 2 provides a very simple check on whether such a correlation is visible to the naked eye. We split the set of AFT recipients by the median into two cohorts, "low recipients" and "high recipients", based on average 2000-2005 receipts. We would want to see higher export growth in the latter group than in the former over the next five-year window (2005-2010), the lag being to leave room for delayed effects. In order to get some more information out of the data, Figure 2. looks separately at each quintile of the (baseline) export/capita distribution. Thus, Q1 is the worst-performing quintile in the baseline period, Q2 is the second-worst, and so on. Results are striking: only in the top two quintiles do we see a positive export-growth differential between high- and low-recipients (Panel a). On the possibility (see below) that AFT will have an indirect effect on export performance by working primarily through improved logistics markets, Panel b carries out the same exercise for the time to export, with similarly disappointing results.

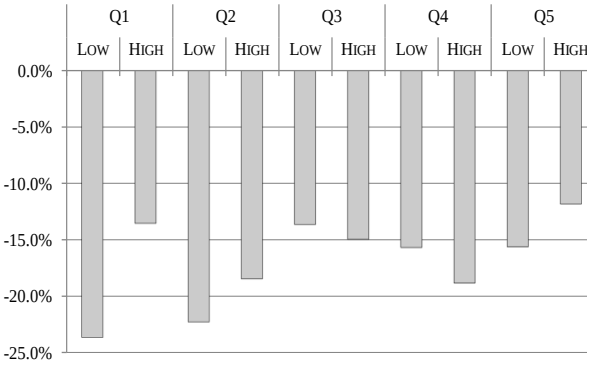
Although many confounding influences and channels of reverse causality should be filtered out before any firm conclusion is reached (see below), these results suggest that it will be difficult to tease out links between expenditures on AFT towards productive sectors and a final outcome like aggregate exports unless the channels are explicitly taken into account.

Figure 2 - Export growth and time to export vs. lagged AFT, by quintile of the export per capita distribution

(a) Export growth (five-year cumulative)



(b) Time to export



Source: Authors' calculations using OECD CRS database and WDI

2.2. Evidence from econometric studies

Using the OECD's CRS,[6] Cali and te Velde (2011) regress trading costs and aggregate export value on lagged AFT disbursements and control variables, on a panel of developing countries. Identification is based on recipient fixed effects (FE) and instrumenting AFT flows with the Freedom House's index of civil liberties, the authors arguing that the Millenium Challenge Corporation explicitly uses that index as an input in their aid allocation mechanism. For aid to infrastructure, coefficients are significant in some specifications, but with limited robustness. As for aid to productive capacity, it fails to correlate with exports whatever the specification, estimator, or lag structure. As for results on sectorally-targeted aid, they tend to confirm the profession's longstanding skepticism about targeted support. Cali and te Velde (2011) find significant effects only in some specifications, and they vanish as soon as comparative advantage is controlled by country-sector fixed effects.

Brenton and von Uexkul (2009) find that, in a simple before-after comparison, sectors that receive aid support perform better, but a difference-in-differences regression of country-sector exports on aid flows controlling for heterogeneity through matching does not show significant effects (in particular once outliers are eliminated), suggesting that crude comparisons that fail to control for aid endogeneity pick up reverse causation.

Ferro, Portugal-Perez, and Wilson (2011) exploit the differential intensities of service use across manufacturing sectors (based on input-output tables from the U.S. and Argentina) to evaluate the impact of aid for trade flows directed at five services 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. AFT: Impact through Infrastructure and Trade Costs

We begin with studies exploring the determinants of trade costs, and then consider the effects through improvements in hard infrastructure, such as roads and ports, and soft infrastructure, such as customs. The key constraints to estimating the effects of such trade-facilitating interventions are the endogeneity of program 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). Therefore, most of the studies that we review in this section do not involve rigorous impact evaluation.

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

3.1. Drivers of trade costs: What the gravity equation tells us

That international trade costs are very large has long been established. Almost all comparisons of aggregate trade costs are based on some version of the ubiquitous gravity equation. Two recent estimates (Novy (2012) and Arvis et al. (2013)) invert the structural gravity equation to compute bilateral trade costs. In their approach, changes in bilateral trade costs are inferred from changes in the ratio of bilateral trade flows relative to domestic trade (approximated by GDP "purged" of trade and services).

Thus, if one is willing to accept that structural gravity holds on the data (and that income and trade are jointly determined), the inverted gravity approach provides an estimate (rather a calibration) of aggregate trade costs directly obtained from observable data. The resulting ad-valorem estimate of total bilateral trade costs (including the effects of tariffs, language barriers, currency barriers, the equivalent of NTMs, etc..) has two advantages over common proxies. First, it does not rely on a functional form for trade costs; second, it varies over time while typical proxies in the standard gravity approach (e.g. distance) do not vary over time.

From a sample of 13 OECD countries covering 1970 to 2000, Novy (2012) estimates that Canada's and Korea's average trade costs fell respectively from 131% to 101% and from 246% to 146%. He also estimates that trade costs between the US and its NAFTA partners dropped more rapidly during the period of NAFTA implementation, showing the benefits of market integration. Comfortingly, he also shows that his constructed measure of trade costs is correlated with expected determinants (e.g. distance increases trade costs and adjacency reduces them).

Arvis et al. (2013) carry out the same calibration over the period 1995-2010 for a large sample of 178 countries to show that relative trade costs have fallen less rapidly in low-income countries, (especially in SSA) than in developed countries. Their decomposition of between-country trade costs shows that geography (distance, contiguity, etc.) and that policy variables (tariffs, RTAs, entry costs into a new business, logistics and Liner shipping connectivity indexes) all contribute significantly to trade costs along expected lines with the quantitatively most significant contributions coming from distance, and the composite liner and logistics indexes. Their results suggest a broad-based approach to policy reform that takes into account the interconnections among the various sources of trade costs. There are, however, two problems with their analysis. First, the inverted gravity approach provides an estimate of trade costs that is consistent with observed trade volumes but not independent of these volumes. In fact, we know from other research (e.g. Hummels, 1998, and Fink et al., 2002) that bilateral trade costs are highly sensitive to bilateral trade volumes. Therefore, assessing the determinants of trade costs using a specification which does not take into account the influence of trade volumes (suitably instrumented) suffers from a serious omitted variable problem. Second, the strong multi-collinearity across components and the aggregative nature of these proxies is problematic: is it customs, roads, telecoms, or competition among providers that is the major bottleneck?

A large literature has relied on the gravity model to disentangle some of the components of the trade costs identified in figure 4 in the introduction to this symposium. As reviewed below, these have only examined some of the components of trade costs and have usually been carried out on a cross-section basis as the variables are not available on a time-series basis. This means that they cannot examine how changes in AFT flows affect trade costs. The main contributions are reviewed below drawing a distinction between 'hard' and 'soft' trade costs.

3.2. The 'Hard' Side: Roads and Ports

Beyond longstanding interest in the question of how transport costs--especially maritime costs that account for 80% of world trade—have evolved (see Moneta, 1959, or more recently Hummels, 1998), attention has turned to the constraints on LDC exports created by poor infrastructure. This emphasis arose from the observation of Africa's poor export performance in the late 1990s in spite of wide-ranging structural adjustment reforms.[7] For instance, in an early study, Amjadi and Yeats (1995) found that over 40% of the export earnings of some of Africa's landlocked countries were absorbed by freight and insurance payments, with a continent-wide average of 15% (against 5.8% for all developing countries).

As we consider this channel of impact, the first question is whether there is any evidence that aid affects infrastructure? Vijil and Wagner (2012) look for the effect of infrastructure-aid commitments on an index of infrastructure quality composed of roads and telecom densities in a cross-section of 91 countries for which they take average values of all variables over 2002-2007. They control for overall ODA inflows, geography and institutions (proxied by a rule-of-law index), and deal with endogeneity and measurement error by instrumenting aid to infrastructure by the number of privatizations in the infrastructure sector between 2000 and 2007. They find that when all country controls are included, the quality of infrastructure is significantly positively correlated with aid to infrastructure in all 2SLS specifications.

A second question is whether infrastructure affects trade costs. A breakthrough came with the pioneering study of Limão and Venables' (2000), where they introduced shipping company quotes for a 40ft container carrying standard good as a measure of trade costs alternative to cif/fob price comparisons. They approximated 'hard' infrastructure by a composite index of roads, rail and telephone lines which they showed contributed 50 percent of the total variation in container rates across destinations while distance only contributed

7 Frankel (1997) found that "under-trading" was particularly acute in the case of intra-regional trade. Classic papers by Collier (1995) and Collier and Gunning (1999) attributed Africa's under-trading to the disastrous policies including (inter alia) protectionism, currency overvaluation and export monopolies, adopted roughly between the mid-70s and mid-90s. However, Foroutan and Pritchett (1993), Coe and Hoffmaister (1998), and Rodrik (1998) argued that size, income and other gravity determinants largely explained Africa's low trade volumes.

10 percent of that variation.[8] They showed that an improvement from the 75th percentile to the median for their infrastructure index would be equivalent to a distance reduction of 3'466km by sea or 419km overland. In addition to confirming the high costs of being landlocked, they detected additional costs to overland distance (1'000 km of overland distance added on average $1'380 to container freight costs, against only $190 by sea) for landlocked countries compounded by border delays, uncertainty, higher insurance costs, and charges by transit countries. They showed that an improvement from the 75th percentile to the median for their infrastructure index would be equivalent to a distance reduction of 3'466km by sea or 419km overland. Finally, they also showed that this estimated transport cost estimate performed very well in a standard gravity equation, estimating that a 10% reduction in trade costs increased trade by 30%. Their key finding was that 'hard' infrastructure accounted for nearly half of the transport cost penalty borne by intra-SSA trade. This poor infrastructure over-explained the under-performance of the continent's trade.[9]

The policy implications of this body of work were clear: It provided intellectual support to a return of the "big-push" view, according to which donors should build roads and ports in order to unlock Africa's trade and, by implication, its growth. Indeed, Buys, Deichmann and Wheeler (2010) explored the returns on a pan-African program of road infrastructure development on inter-city corridors. Combining gravity coefficients to estimate the program's trade impact with World Bank data on road improvement costs ($127'000/km for the median project) they found that the payback horizon would be barely over one year, with $254 billion of additional trade generated over the project's lifetime at a cost of about $32 billion. A similar exercise performed by Shepherd and Wilson (2006) for the ECA (Europe and Central Asia) region reached similar conclusions: A complete upgrading of the road infrastructure in Hungary, Romania and Albania (at about $227'000/km) would generate an "on-impact" (annual) trade increase of over $35 billion for a one-time cost of $3 billion.

The extremely high rates of return on road investments identified in the trade literature were substantially above, but broadly consistent with, high rate-of-return estimates found in the macro growth literature—typically over 200%—which put road investments on top of other types of infrastructure investments such as telecommunications or energy (see Estache 2007).[10]

[8] As noted by Limaõ and Venables who were the first to introduce a composite index of infrastructure, taking a linear combination of these components assumes that these inputs are perfect substitutes. Bundled up with capital and labor in a Cobb-Douglas function gives a cost function for transport costs.

[9] Coulibaly and Fontagné (2006) confirmed Limão and Venables' results on aggregate and disaggregated trade flows in West Africa, predicting that if all roads were paved in the region, trade would almost treble.

[10] These too-good-to-be-true rates of return were reminiscent of the "Aschauer debate" on infrastructure and growth (see Estache and Fay 2007 and references therein for an overview). An internal evaluation of World Bank infrastructure projects over 1999-2003 produced an economic rate of return of 43%, by all means a respectable rate but nowhere near the miracles suggested by the literature (see Estache 2007). However, the ranking of rates of return also put road investments on top, suggesting the same lending priorities.

Thus, after almost two decades of multilateral donor emphasis on structural adjustment and policy reform, by the mid-2000s empirical research was suggesting that the pendulum should swing back toward (infrastructure) capital accumulation.

3.3. The 'Soft' Side: Customs and Regulation

For interventions such as customs reform, it may be possible in principle to generate a control group by introducing 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 clearance for trusted operators, can be restricted and randomly allocated in an early phase, using non-eligible operators as controls. However, in practice, there are few examples of such programs.

Cantens, Raballand, Bilangna, and Djeuwo (2011) describe a recent pilot for customs reform in Cameroon that involved the introduction of contracts with performance indicators for frontline 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. This project is an interesting example of a trade intervention that in principle is non-targeted, but where targeting could have been introduced by focusing on a sub-set of frontline customs inspectors. But it was not. Therefore, evaluation of the customs performance contracts project was conducted as a comparison of inspectors' behavior 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 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 re-routings from the yellow channel (documents control) to the red channel (physical inspection) declined tremendously.

3.4. The Neglected Soft Power of Competition

While evidence accumulated on the strong effect of infrastructure on trade costs, whether the right policy response was a "big push" in infrastructure investment was questioned by Teravalinthorn and Raballand (2008). This preoccupation reflects a new awareness of the importance of the 'logistics markets' (see figure 4 in the introduction to this symposium issue)

inspired by Chile's deregulation of its shipping industry, which until 1979 had been regulated by a cargo reservation system dating back from the 1950s.

Since maritime transport still accounts for 80% of world transport, it is important to explore how generalizable the Chilean experience is. Fink et al. (2002) and Clark et al (2004) explored the impact of efficiency on shipping costs to the US. Relying on efficiency estimates of ports drawn from interviews, Fink et al. regressed freight-rate data for U.S. seaborne imports on the existence of maritime cartels (so-called "shipping conferences") as well as various restrictive regulations applying to shipping (cargo reservation schemes) and port operations. Their cross-section estimates suggest that cartels pushed up freight rates by about a third but the evidence on policy restrictions was inconclusive.

Drawing on reliable US data on bilateral import charges at the HS-6 level over the period 1991-2003, Blonigen and Wilson (2008) regressed for each product import charges on all relevant characteristics except changes in product composition. After controlling for all other factors affecting charges, their port fixed effects provided an efficiency ranking of US and foreign ports. Overall, they estimated that a 10 percent increase in port efficiency increased trade between a country-pair by 3.2 percent, or alternatively a change in port efficiency from the 75% percentile to the 25% percentile led to a 5 percent increase in trade.

Further progress came from studies digging deeper into cartel behavior, long known to be prevalent among "shipping conferences". Inspired by the observation that Caribbean and Central American countries trade far less than would be predicted by the gravity model (Guatemala's exports of manufactures to Caribbean partners are far less than 1%, yet they are close and have easy access to each other by sea), Wilmeister and Hoffman (2008) analyze freight rates charged by one major liner shipping company on 189 routes in the region. Their estimates show that distance is trumped by the number of liner shipping companies providing services between pairs of countries, a result that would likely also carry over to Sub-Saharan Africa where transshipments are frequent.[11]

Again focusing on US ports and maritime traffic to Latin America, Hummels, Luggovsky and Skiba (2009) estimate the market power of shipping companies by using the cross-product variation of tariffs to identify unobserved market power.[12] They estimate that eliminating market power in shipping would boost trade volumes by 5.9% for the US and 15.2% for Latin America. Furthermore, high tariffs on trade give market power to shippers: a 1% increase in tariffs leads to a 1-2% increase in shipping prices per kilo.

[11] Their model of liner shipping freight rates controls for transshipment vs. direct services; the number of competing carriers; UNCTAD's liner shipping connectivity index; transit time; port infrastructure endowment in the importing and exporting countries. The model accounts for three fifth of the variance of liner shipping freight rates across the Caribbean.

[12] When tariffs are high, the share of freight costs in consumer prices is lower, and so is the price elasticity of demand perceived by the shipping lines, which will, if they have market power, induce them to raise freight rates. Thus, the co-movement of tariffs and freight rates identifies market power.

Turning to road transport, Teravaninthorn and Raballand (2008) showed that trucking deregulation in Rwanda after the civil war had effects similar to those of shipping deregulation in Chile: nominal rates dropped by 30% and the domestic trucking fleet expanded instead of shrinking. By contrast, countries like Malawi where domestic truckers were protected by restrictive entry regulations, ended up essentially penalizing farmers—a common policy outcome in Africa. They also highlighted the deleterious effects of cartels and regulations through "freight bureaus" on Central African corridors where freight rates per ton/km were about 80% more and truck-utilization rates 40% less than on East African corridors.[13] Throughout West Africa, they found that bilateral agreements, queuing systems and quotas stifled competition. Even on the most competitive trucking corridors of East Africa, anticompetitive regulations abounded, with e.g. Kenya prohibiting international transit trucks on the Mombasa-Kigali corridor from taking domestic freight on the return trip, forcing them to cover 1'700km empty.[14]

In fact, a new cross-country database on services policy reveals a perverse pattern: many landlocked countries restrict trade in the very services that connect them with the rest of the world (Borchert, et al. 2012). In particular, air-transport policies are significantly more restrictive on average in landlocked countries than elsewhere. The phenomenon is most starkly visible in Sub-Saharan Africa and is associated with lower levels of political accountability. This paper finds evidence that these policies lead to more concentrated market structures and more limited access to services than these countries would otherwise have, even after taking into account the influence of geography and incomes, and the possibility that policy is endogenous. In the aviation sectors, moving from an intermediate level of restrictiveness to an open regime could lead to a 25 percent increase in flight connections per airline.

For donors, the implications of this work were starkly different from those of previous pieces of empirical research on infrastructure. Rather than build more roads, ports and airports, they should pursue policy dialogue with recipient governments to improve regulatory frameworks and ensure competition in service provision. The burden of action is not just on aid-recipient countries, but on the donors themselves, because they too maintain restrictive arrangements in areas like air transport and condone (through exemptions from competition law) anti-competitive practices by private providers in areas like maritime transport.

13 Interestingly, when regressing transport prices on road condition, they found negative and significant effects in East Africa, but insignificant or positive effects in West and Central Africa (Table 4.3 p. 42), suggesting pricing formulas based on anticompetitive arrangements rather than marginal costs in those regions.

14 They collected data on costs (Vehicle Operating Costs (VOCs), transport costs incurred by transport providers) and prices paid by end users from a sample of trucking companies operating across the continent. They then simulated the effects of a reduction in: (i) fuel price; (ii) informal payments; (iii) reduction in border crossing time; (iv) rehabilitation of corridors. Their simulations showed that for West Africa (and to a lesser extent Central Africa), a reduction in fuel prices and a rehabilitation of roads would have no effect on prices paid by end users because of barriers to entry. By contrast in Eastern Africa, the same policies would reduce prices paid by end users.

4. AFT: Impact through Direct Support to Exporters

Direct support to exporters includes "clinical" trade competitiveness programs 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. The key feature of these interventions is that the programs are assigned exclusively to certain units, often firms. Because these interventions operate at the level of the firm, non-assisted firms can in principle serve as the control group and more rigorous evaluation is feasible.

4.1. Approaches to Evaluation and Data Needs

Targeted programs of assistance could, in principle, be amenable to Randomized Control Trial (RCT) design provided that the decision to randomize assignment was taken ex ante. Since in practice, only a minority of policies fit into this category, the alternative is to rely on Quasi-Experimental (QE) methods (e.g. matching, difference-in-differences, instrumental variables, or regression discontinuity design). In that case, ideal data for QE methods will typically include

- *Trade data* at the transaction level from customs, which is available from ASYCUDA raw files. The data can be easily anonymized by deleting firm names and keeping only TINs (tax identification numbers) and will provide information on firm-level outcomes;
- *Program data* including enrolment status, dropouts, and rejects;
- *Firm characteristics data* from an industrial survey (typically balance-sheet information including employment, turnover, age, as well as ownership, number of establishments, etc.); with the survey's key for the classification of firms being compatible with that of customs data for reconciliation, precluding the use of "dummy" firm identifiers.

Clearly, these data requirements are heavy and raise confidentiality issues; whether the data will actually be made available to the evaluation team by government authorities depends on interest (buy-in) for the IE's results, donor involvement, and quality of the dialogue.[15]

4.2. The Cost of Implementation

In practice, efforts to generalize the use of IE in trade interventions face two types of constraints: incentives and resources.

[15] The World Bank has recently launched the « Exporter Dynamics » project which aims at collecting precisely this type of data (at least the customs data) from Customs administrations around the world. However, sharing the data with researchers has proved a difficult and laborious process because of the confidentiality issues involved.

In terms of incentives, an IE risks slowing down project roll-out and diverting managerial attention for results that are unlikely to be available within a manager's tenure horizon; and if they did, they might do more harm than good. Moreover, in order to be incentive-compatible, IE should be used only to generate new knowledge and should be fully decoupled from the evaluation of project managers. However, it is not clear that an organization could make such a claim credible, as it would obviously suffer from a time-inconsistency problem.

As for funding, the basic issue is that IEs have tended to be on the expensive side, although there may be scope to reduce costs. Gertler et al. (2011) show that, for a sample of World Bank-supported programs for which IE and program cost data was available, IE costs represented on average between 4% and 5% of total program costs, ranging between 0.2% and 13.3%. This is because project costs in the sample ranged between $11 million (Rwanda) and $86 million (Colombia). Trade-related projects rarely attain such levels. If we take DIME's estimate of a minimum of $300'000 for a feasible IE, a ratio of 5% would require a project of $6 million. By the standards of World Bank projects in social development, poverty, or health, this is a small project. By the standards of trade-related assistance, it is very large.

These rough calculations provide one reason for the slow spread of IE in trade-related assistance and suggest that IE templates must be adapted to the area of trade assistance in order to make IE an acceptable proposition for donors. Clearly, quasi-experimental methods using statistical data instead of original household surveys are the way to go. We now turn to a few examples of recent IEs in that spirit and how they have contributed to our understanding of the effectiveness of trade interventions.

4.3. Early results: Does export promotion make a difference?

So far, there have been very few impact evaluations of trade-related interventions, and only a thin, "early" literature can be reported on. However, the performance of export-promotion agencies, which is one of the few areas of "clinical-type" interventions that have been extensively studied, provides a good testing ground to evaluate the contribution that IE can bring to policy debates and dialogue with developing countries.

A new strand of literature, surveyed in Volpe (2011), has turned to "clinical" (firm-level) evaluation of Export Promotion Agencies (EPAs). Using "difference-in-difference" (DID) estimation at the firm level, Alvarez and Crespi (2000) found that Chile's EPA, PROCHILE, had an impact on the beneficiaries' number of destinations, although not on their number of export products. Since then, a number of firm-level studies have shown that export promotion seems to be more successful at affecting the performance of established exporters than at encouraging non-exporting firms to start exporting (Bernard and Jensen, 2004; Görg, Henry and Strobl, 2008; Girma, Gong, Görg and Yu, 2009), as exporters differ from non-exporters in terms of productivity and other characteristics (see, e.g. Bernard, Jensen, Redding and Schott, 2007), which export promotion may not be able to offset. The impact seems stronger

along the extensive margin than along the intensive one (Alvarez and Crespi, 2000, Volpe and Carballo, 2008,).

Cadot et al. (2012) evaluated the effects of the FAMEX export promotion program in Tunisia on the performance of beneficiary firms. While much of the literature assesses only the short-term impact of such programs, their paper also considers the longer-term impact. Propensity-score matching, DID and weighted least squares estimates suggest that beneficiaries initially see faster export growth and greater diversification across destination markets and products. However, three years after the intervention, the growth rates and the export levels of beneficiaries are not significantly different from those of non-beneficiary firms. Exports of beneficiaries do remain more diversified, but the diversification does not translate into lower volatility of exports. The authors also did not find evidence that the program produced spillover benefits for non-beneficiary firms. However, the results on the longer-term impact of export promotion must be interpreted cautiously because the later years of the sample period saw a collapse in world trade, which may not have affected all firms equally.

Did this literature produce any insight that the cross-country literature did not? On one hand, it did not *overturn* the qualitative result of Lederman, Olarreaga and Payton (2010) that EPAs do make a difference. On the other hand, the result is qualified; for instance, estimated effects tend to be substantially smaller at the firm level (Cadot et al. (2012) find only six dollars of additional exports for one dollar of export promotion). Second, the level of detail in the decomposition of EPA activities tends to be higher in the clinical studies than in survey-based cross-country studies, allowing for closer examination of which "treatment arms" seem to be most effective; finally, the decomposition of impacts along various margins of firm performance (extensive or intensive) is necessarily richer at the firm level. However, clinical studies have little external validity; for instance, the success of PROCHILE in fostering diversification and innovation may have to do with many features of the Chilean business and Government environment that could not be transplanted easily.

In sum, as Rodrik (2008) put it, there is an inescapable trade-off between "internal validity" (the ability to identify impact effects net of confounding influences), which improves as one goes from cross-country studies to impact evaluations, and "external validity" (the ability to draw general policy propositions from evaluation results), which may well worsen.

5. Concluding remarks

The literature we surveyed based on traditional econometric analysis has theoretic foundations but the wide spectrum of results reveals the difficulty of drawing robust policy conclusions because of confounding influences. Impact evaluation (IE) techniques provide "internal validity" as confounding influences can be controlled better. How promising, then, is the use of IE to trade interventions?

First, although IE is "a-theoretic", most of the practical IE literature pays at least lip service to the need for evaluation to be backed by some sort of "theory of change" (see e.g. Gertler et al. 2011).

Take "hard" infrastructure which often plays a twin role. Apart from its direct effect on trade costs, it also provides an opportunity to start or maintain dialogue with recipient governments on policy reforms, e.g. in terms of regulation of related services (trucking, maritime transport etc.) or even on broader agendas (introduction of competition). How much donors actually use this leverage effect of infrastructure investments varies, depending on the depth of their dialogue, their own economic sophistication, and their willingness to look critically at their own competition-inhibiting policies. But in this "soft" area as well, the theory of change is there, as the IO-and-trade literature has long established the inter-relationship between trade performance and regulatory/competition policy (the so-called "behind-the-border" agenda).

Second, in order to generalize the use of IE in trade-related interventions, what is needed is to make it practically feasible in terms of design (project and evaluation), incentives, and resources. In terms of design, the message of our brief overview of methods is that there is substantial scope for adapting methods to the particular context of trade interventions, especially with quasi-experimental approaches. In terms of incentives, we argued that if the decision to launch an IE and budget for it out of project resources is left to project managers, there is an agency problem. Part of the problem is the potential for IE to bring bad news. Thus, IE results should be decoupled from individual performance evaluation, but promises to keep a firewall between the two are unlikely to be time-consistent. One solution might be to set up an independent IE center for AFT projects as suggested by Hoekman Wilson (2010). However, ultimately government buy-in would be a crucial ingredient, and it would be unlikely with a complete separation of IE from project management. There is clearly a need for further thinking on this issue.

Finally, and perhaps most importantly, adopting IE as routine practice in AFT projects requires the "evaluation community" to work on reducing IE costs. Although experienced IE practitioners like to warn newcomers against "doing IE on a shoestring", the currently very high cost of IEs acts as a powerful deterrent. In trade policy, there should be scope for better use of existing statistics and, crucially, for more dialogue with governments to ensure the availability of firm-level statistics. That is where the issues of cost and buy-in converge: Governments will be more willing to relinquish semi-confidential data to researchers if they understand the value of the results generated.

References

Alvarez, R., and G. Crespi (2000). 'Exporter Performance and promotion Instruments: Chilean Emprical Evidence', *Estudios de Economia* 27, 2, 225-24.

Amjadi, A. and A. Yeats (1995), 'Have Transport Costs Contributed to the Relative Decline of Sub-Saharan African Exports?', Policy Research Working Paper Series 1559, World Bank, Washington, DC.

Bernard, A. and B. Jensen (2004), 'Why Some Firms Export', *Review of Economics and Statistics* 86, 2, 561-569.

Bernard, A., S. Redding and P. Schott (2007). 'Firms in International Trade', *Journal of Economic Perspectives* 21, 3,105-130.

Bloningen, B. and W. Wilson (2008) 'Port Efficiency and Trade Flows', *Review of International Economics*, 16, 1, 12-36.

Borchert, I., B. Gootiiz, A. Grover and A. Mattoo (2012), 'Landlocked or Policy Locked? How Services Trade Protection Deepens Economic Isolation' World Bank Policy Research Working Paper 5942.

Brenton, P. and E. von Uexkull (2009). 'Product-Specific Technical Assistance for Exports—Has it Been Effective?', *Journal of International Trade and Economic Development* 18, 2, 235-254.

Buys, P., U. Deichmann and D. Wheeler (2010), 'Road Network Upgrading and Overland Trade Expansion in Sub-Saharan Africa-super- †', *Journal of African Economies*, Centre for the Study of African Economies (CSAE), 19, 3, 399-432.

Cali, M. and D. te Velde (2011), 'Does Aid for Trade Really Improve Trade Performance?', *World Development*, 39, 5, 725-40.

Cadot, O., A. Fernandes, J. Gourdon, and A. Mattoo (2012), 'Are the Benefits of Export Support Durable? Evidence from Tunisia', World Bank Policy Research Working Paper No. 6295.

Cantens, T., G. Raballand, S. Bilangna and M. Djeuwo (2011), 'Reforming Customs by Measuring Performance: A Cameroon Case Study', in O. Cadot, A. Fernandes, J. Gourdon and A. Mattoo, eds., *Where to spend the next million: Applying impact evaluation to trade*; London/Washington, DC: CEPR/The World Bank, pp. 183-206.

Clark, X., D. Dollar, and A. Micco (2004), 'Port efficiency, maritime transport costs, and bilateral trade', *Journal of Development Economics*, 75, 2, 417-450.

Coe, D.T. and A.W. Hoffmaister (1998), 'North-South Trade: Is Africa Unusual?', Working Paper of the International Monetary Fund, Washington, D.C.

Collier, P. and J.W. Gunning (1999), 'Explaining African Economic Performance', *Journal of Economic Literature*, 37, 1, 64-111.

Collier, P. (1995), 'The Marginalization of Africa', *International Labour Review* 134, 4-5,541-57.

Coulibaly, S. and L. Fontagné (2006), 'South-South Trade: Geography Matters', *Journal of African Economies*, 15, 2, 313-341.

Estache, A. (2007), 'Infrastructure and Development: A survey of Recent and Upcoming Issues', ULB Institutional Repository 2013/44060, Brussels: Université Libre de Bruxelles.

Estache, A. and M. Fay (2007), 'Current debates on infrastructure policy', Policy Research Working Paper 4410; Washington, DC: The World Bank.

Ferro, E., A. Portugal-Perez, and J. Wilson (2012), 'Aid to the Services Sector: Does it Affect Manufacturing Exports?', Policy Research Working Paper Series 5728, World Bank, Washington, DC.

Fink, C., A. Mattoo and C. I. Neagu (2002), 'Trade in International Maritime Services: How Much Does Policy Matter?', *World Bank Economic Review*, 16, 1, 81–108.

Foroutan, F. and L. Pritchett (1993), 'Intra-Sub-Saharan African Trade: Is It Too Little?', *Journal of African Economies* 2,1,74-105.

Gertler, P., S. Martinez, P. Premand, L. Rawlings and C. Vermeersch (2011), Impact evaluation in practice; Washington, DC: The World Bank.

Girma, S., Y. Gong, H. Görg and Z. Yu (2009), 'Can Production Subsidies Explain China's Export Performance? Evidence from Firm-level Data', *Scandinavian Journal of Economics* 111, 4, 863-891.

Görg, H., M. Henry and E. Strobl (2008), 'Grant Support and Exporting Activity', *Review of Economics and Statistics* 90, 1, 168-174.

Görg, H. and J. Wilson (2010), 'Aid for Trade: Building on Progress to Date', paper prepared for the Korea-World Bank High Level Conference on Post-Crisis Growth and Development, Busan, June 3-4, 2010.

Hummels, D. (1998), 'Toward a Geography of Transport Costs', mimeo, University of Chicago.

Hummels, D, V. Lugovskyy and A. Skiba (2009), 'The Trade Reducing Effects of Market Power in International Shipping', *Journal of Development Economics*, 89, 1, 84-97.

Lederman, D., M. Olarreaga and L. Payton (2010), 'Export Promotion Agencies Revisited', *Journal of Development Economics* 91, 2, 257-265.

Limao, N. and A.J. Venables (2000), 'Infrastructure, Geographical Disadvantage, Transport Costs, and Trade', *World Bank Economic Review*, 15, 3, 451–479.

Moneta, C. (1959), 'The Estimation of Transport Costs in International Trade', *Journal of Political Economy* , 67, 1, 41-58.

Novy, D. (2012), 'Gravity Redux: Measuring International Trade Costs with Panel Data', *Economic Inquiry,* 51, 1, 101-121.

OECD (2009), How to Evaluate Aid for Trade? A Scoping Note, (Paris: Organisation for Economic Cooperation and Development)

OECD (2011), Strengthening Accountability in Aid for Trade, (Paris: Organisation for Economic Cooperation and Development)

Rodrik, D. (2008), 'The New Development Economics: We Shall Experiment, But How Shall We Learn?', Mimeo, Kennedy School of Government, Harvard University.

Shepherd, B. and J. Wilson (2006), 'Road infrastructure in Europe and central asia: does network quality affect trade?', Policy Research Working Paper 4104; Washington, DC: The World Bank.

Teravaninthorn, S. and G. Raballand (2008), 'Transport Prices and Costs in Africa: A Review of the Main International Corridors', Washington, DC: World Bank.

Vijil, M. and L. Wagner (2012), 'Does Aid for Trade enhance export performance? Investigating the infrastructure channel', *The World Economy*, 35, 7, 838-68.

Volpe, C. (2011), 'Assessing the Impact of Trade Promotion in Latin America', in O. Cadot, A. Fernandes, J. Gourdon and A. Mattoo, eds., *Where to spend the next million: Applying impact evaluation to trade*; London/Washington, DC: CEPR/The World Bank, pp. 39-81.

Volpe, C. and J. Carballo (2010), 'Beyond the Average Effects: The Distributional Impacts of Export-Promotion Programs in Developing Countries', *Journal of Development Economics* 92, 2, 201-214.

Wilmeister, G. and J. Hoffman (2008), 'Liner Shipping Connectivity and Port Infrastructure as Determinants of Freight Rates in the Caribbean', *Maritime Economics and Logistics* 10, 1-2, 131-151