Estimation of Missing Intra-African Trade

By

Nelson Villoria

GTAP Research Memorandum No. 12

November 2008

Center for Global Trade Analysis, Department of Agricultural Economics, Purdue University.
Email: nvillori@purdue.edu, Phone: (765) 494 – 8386

Table of Contents

Estimation of Missing Intra-African Trade

By

Nelson Villoria

Abstract

Missing trade is defined as the exports and imports that may have taken place between two potential trading partners, but which are unknown to the researcher because neither partner reported them to the United Nation's COMTRADE, the official global repository of trade statistics. In a comprehensive sample of African countries, over 40% of the potential trade flows fit this definition. For a continent whose trade integration remains an important avenue for development, this lack of information hinders the analysis of policy mechanisms -- such as the Economic Partnership Agreements with the EU -- that influence intra-regional trade patterns. This paper estimates the likely magnitude of the missing trade by modeling the manufacturing trade data in the GTAP Data Base using a gravity approach. The gravity approach employed here relates bilateral trade to country size, distance, and other trade costs while explicitly considering that high fixed costs can totally inhibit trade. This last feature provides an adequate framework to explain the numerous zero-valued flows that characterize intra-African trade. The predicted missing exports are valued at approximately 300 million USD. The incidence of missing trade is highest in the lowest income countries of Central and West Africa.

Estimation of Missing Intra-African Trade

By

Nelson Villoria

1. Introduction

A large amount of bilateral transactions between African countries are not reported to the United Nations Commodity Trade Statistics Database (COMTRADE). In the absence of a given country's statistics, trade specialists try to recover these unreported trade flows by using reconciliation techniques, as discussed below. However, when neither partner reports its trade, reconciliation is not possible and the trade is considered missing. The objective of this paper is to estimate the missing trade that arises from the lack of reporting in a comprehensive sample of African countries.

There are important reasons to study the extent of missing intra-African trade. For example, most countries in the region are signatories of agreements that seek to increase trade among their member countries. More recently, the Economic Partnership Agreements promoted by the European Union are likely to have significant effects on intra-African trade patterns. The design and evaluation of these efforts depends on the ability to capture effects inherent to discriminatory trade agreements such as trade diversion and creation. In the absence of reliable and up-to-date trade data, predictions of the missing trade provide a sense of the magnitudes involved, and as such, an opportunity to validate the results obtained with limited data.

There are a number of studies concerned with the question of whether intra-African trade is lower than expected, given the knowledge of some of the main determinants of trade. Foroutan and Pritchett (1993) posed the following question: "Compared to a sample of other countries with roughly similar economic characteristics, do SSA trade too little with each other?" They used a gravity equation framework and concluded that compared to other low income countries, SSA countries do not trade significantly less, and if anything, they trade slightly more than gravity-based predictions suggest. Their conclusions were supported by Rodrik (1998), who indicated that, as in any other region, intra-African trade is well explained by country size, per capita income, geography, and taxation of trade. From a different perspective, Coe and Hoffmaister (1999) asked whether trade between Africa and developed countries deviates from the average level of trade between other developing countries and countries in the north. Using a gravity framework, they found that trade between Africa and the north is "normal" as explained by the gravity factors. The bottom line of these studies is that if intra-African trade is low, it is because fundamentals such as economic size (i.e., GDP) are low.

More recent work has been concerned with the impediments to expanding intra-African trade. For example, Coulibaly and Fontagné (2006) argued that if economic size is the reason for low levels of trade, there would not be untapped potential for trade. They then explored the role of paved roads and internal geography in the West African Economic and Monetary Union countries and concluded that by improving infrastructure, intra-African trade would increase significantly. Longo and Sekkat (2004) explored the role of infrastructure availability, political conflicts, and economic policy on shaping trade among African countries, and with other countries. They concluded that by improving the infrastructure and economic policy, intra-African trade would expand.

In our framework, we do not explore whether intra-African trade is low or what are its determinants. Rather, we assume that a gravity equation reasonably explains intra-African trade and we use it as the basis for our estimation of the missing flows. In the next section, we discuss the data employed and the extent of the missing trade to be estimated. Our empirical strategy is explained in Section 3. Section 4 discusses our results and Section 5 concludes with some remarks.

2. Missing Trade in Africa: Definition and Extent of the Problem

Definition

Ideally, trade statistics are reported by both partners engaged in bilateral trade. For any two given countries, say A and B, the exports of country A must equal the imports (net of shipping and insurance costs) of country B. Conversely, country A's imports from country B should be identical to country B's exports to country A, once the relevant adjustments for transportation costs are made. In practice, however, it is common to find that the flows reported by countries differ.

An extreme case of discrepancies between reported flows is the case where only one country reports its trade. To illustrate, imagine that country A exports to country B, but only country A reports its statistics. In this case, trade specialists recover country B's imports from A by simply taking country A's exports and correcting them for any relevant costs. It is also common to find both countries reporting their bilateral transactions but with discrepancies that are larger than what might reasonably be explained by transportation and insurance costs. In these cases, specialists in trade statistics use reconciliation techniques based on the quality of the reporters, evidence of misclassification, transportation costs, and other factors[1].

Reconciliation is impossible, however, when neither A nor B report their trade. In this case, it is not possible to know whether the two countries engaged in trade. The potential trade that occurred but is not observable is known as *missing* trade. It is this form of missing trade that is the concern of this paper.

Despite reconciliation efforts, it is possible that reported statistics do not reflect the actual levels of trade. For Africa, this would be the case, for example, if there is significant smuggling, or because of institutional weaknesses that result in systematic misreporting of the transactions by the two partners. In these cases, the trade statistics are flawed and might result in *missing* trade. Because it is not feasible to know the real level of trade in these cases, reconciliation processes are not an answer to this problem.

Extent of missing trade

This study uses the data on merchandize trade that underlie the GTAP Data Base V.6 (reference year 2001). These data were originally sourced from the UN's COMTRADE database, and have already been subjected to reconciliation processes as discussed in Gehlhar (2007).

[1] Two leading examples are the GTAP Data Base documented in Gehlhar (2007) and the BACI database documented in Gaulier, Zignago, Sondjo, Sissoko and Paillacar (2007).

The GTAP Data Base gives us a matrix of bilateral exports (valued at FOB prices) for 54 countries in Africa[2] (listed in Table 1). From an aggregated perspective, there are 2,862 potential trade flows (i.e. 54x53 = 2,862). Of these flows, 1,607 are zero-valued. These zeroes can represent two distinct possibilities: trade did not occur at all, or trade is missing according to our definition in the first part of this section. To distinguish between these two possibilities, we used CEPII's BACI auxiliary dataset on missing trade, as documented in Gaulier, et al. (2007 p. 25). This dataset indicates when a trade flow is zero or missing based on the following two assumptions: 1) If a country exhibits a declaration for a particular year, it is assumed that this country is declaring its trade with all its partners; hence, all its flows are either positive or zero trade flows; 2) if neither partner declares its trade, the potential flow is missing.

An issue to consider is that CEPII's database treats the countries of the South African Customs Union (SACU) as a single entity, while GTAP has them disaggregated. According to CEPII, there are no missing exports originating in the SACU; hence, we assume that there are no missing exports for its individual members (Botswana, Lesotho, Namibia, South Africa, and Swaziland). Conversely, the BACI database indicates that SACU's imports from Guinea-Bissau, Libya, and Chad are missing, and so we assume that they are missing for the SACU members taken individually as well.

Our interest is to obtain estimates at the GTAP product level. In principle, the definition of missing trade employed here makes it an aggregated phenomenon: if trade is missing for a given country pair, it is missing for all the products that the country pair could potentially trade. Using an aggregated trade definition of missing trade poses some difficulties at the disaggregated level. In particular, in a limited number of cases, GTAP reports trade in some products between countries for which, at the national level, BACI considers trade missing. Specifically, out of the 2,862 potential intra-African transactions, BACI indicates that 1,200 are missing. However, if we aggregate over all the GTAP merchandise categories, 112 out of the 1,200 missing cases appear with trading in some products. To give an idea of the extent of missing trade at the aggregated level, Table 1 summarizes by country the value of

[2] The dataset employed here uses information on individual countries independently of whether these countries are part of a GTAP aggregated region.

exports, imports, and number of missing flows as an importer and as an exporter according to BACI. In Appendix 1, we have included a matrix that details, partner by partner, which flows are considered missing in the BACI dataset.

At the product level, we solve the discrepancy between GTAP and BACI by using the following decision rule: i) whenever GTAP indicates a positive export at the product level, we consider it as such, regardless of whether BACI classifies the country pair trade as missing or not, and ii) when GTAP indicates a zero valued flow, we classify it as missing or zero according to the information on the country-pair given by BACI. Proceeding in this way, we maximize the number of observations employed in the estimation process, and do not introduce further noise by estimating trade flows that have already been subjected to systematic reconciliation processes.

There are 41 merchandise sectors, or commodities, in the GTAP Data Base (Table 2 shows the total intra-African exports under each of these sectors). Adding the product dimension to the 54 countries, we have 117,342 potential trade flows (i.e., 2,862 bilateral flows* 41 merchandise commodities). Of these, 48,636 appear as missing[3], 58,523 are zero-valued trade flows, and only 10,183 are positive trade flows. In other words, 41.5% of the potential trade flows are considered missing. Our task is to estimate these flows using the econometric techniques described in the next section.

3. Methodological Framework and Data

The Gravity Equation and Sample Selection Bias

We rely on the statistical strength of the gravity equation to estimate the missing trade flows between African countries. The gravity equation conditions bilateral trade flows on variables that are country-specific and variables that change across trading partners. There are two ways of defining the country-specific variables: using measures of country size such as GDP, area, and population (e.g., Rose, 2004) or, alternatively, employing country fixed effects. The country fixed effects present a number of

[3] BACI flags 49,200 flows as missing. Of these, 564 appear as positive in GTAP; hence, we consider 49,200 - 564 = 48,636 missing flows.

theoretical, econometric, and data advantages: i) Theoretically, the fixed effects approximate structural components of the gravity equation in a variety of modeling settings. For example, the multilateral resistance terms of Anderson and Wincoop (2003) and the supply/demand components of price indexes in Redding and Venables (2004) can be estimated using fixed effects; ii) Econometrically, Feenstra (2002) indicated that fixed effects are sufficient to get consistent parameter estimates. This is in line with Hummels (2001), who used the fixed effects to take into account important factors in a demand function such as expenditure shares, changes in output, and prices, controlling the potential parameter inconsistency caused by omitted variables; iii) Due to data availability on country-specific variables, there is a trade-off between sample size and the number of country-specific relevant variables we can include; this is not an issue when using fixed effects. Because of these advantages, we employ fixed effects for controlling the country-specific effects of our gravity equation.

The variables that change across trading partners are used to explain trade costs and other frictions that reduce trade. A key variable present in gravity applications is the distance between the trading partners. Border commonality is also an important variable explaining trade, and has been the subject of intensive research starting with McCallum (1995). Most applications include language commonality, whether the countries were colonized by the same nation, and some proxy for tariffs and trade policy (e.g., whether the two countries belong to the same trade agreement). In recent work, Helpman, Melitz, and Rubinstein (2007) employed information on whether two countries are landlocked. These authors also used religion commonality as a proxy for trading fixed costs in the context of a firm heterogeneity model. We employ information on each of these variables to model the bilateral frictions. Putting together the elements described above, our base model is:

$$\begin{aligned}\ln(X_{ijk}) = \alpha_0 + \sum_i \alpha_i^E E_i + \sum_j \alpha_j^I I_j + \sum_k \alpha_k^K C_k + \alpha_1 \ln(DIST_{ij}) + \alpha_2 BORD_{ij} + \alpha_3 LANG_{ij} + \\ \alpha_4 LOCK_{ij} + \alpha_5 COLY_{ij} + \alpha_6 RELG_{ij} + \alpha_7 TRTY_{ij} + \varepsilon_{ij}\end{aligned} \tag{1}$$

Where X_{ijk} are exports (in thousands of USD) of commodity k (k=1…41) from country i to country j (i,j=1…54), E_i and I_j are the exporter and importer fixed effects discussed above, and $DIST_{ij}$ is the distance (in km) from country *i*'s most important city to country *j*'s most important city. The rest of the variables are dummies that take the value of one when: a pair of trading partners share a border ($BORD_{ij}$); have the same official language, or at least 20% of the population speak the same language ($LANG_{ij}$); are landlocked ($LOCK_{ij}$); were colonized by the same nation ($COLY_{ij}$); in both countries, the dominant religion is the same ($RELG_{ij}$); and belong to the same regional economic group ($TRTY_{ij}$). Otherwise, these dummies are zero. The term ε_{ij} is a stochastic error that is assumed to have zero mean, a constant variance, and distributed independently from any other regressor. The variable C_k is a commodity fixed effect. The parameters to be estimated are represented by the Greek letters α , indexed as shown in Equation 1. The data and sources used in Equation 1 are described below.

An important limitation of Equation 1 is its inability to handle zero-valued trade flows. This inability is inherited from the underlying gravity models that predict only positive trade flows between all the countries. In practical terms, the log-linear form of Equation 1 is not able to take zero values of the dependent variable. However, zero trade flows are quite common and potentially contain important information. A good example is offered by the data we use in this paper, of which zero values represent 85.7%[4] of the known African trade[5]. These zero trade flows can be caused by country-specific features (i.e., very small countries), or by large trade barriers (e.g., large distances); in any case, if the zero trade flows are excluded from the estimation process, the parameters of Equation 1 are likely to be biased. Intuitively, the effects of the regressors that impede trade are likely to be underestimated if zero-valued exports are not included.

Recognizing the importance of zero trade flows, a number of authors have employed a variety of ad hoc measures to include the zero trade flows in the estimation of Equation 1. These measures, reviewed in Linders and De Groot (2006), include: simply ignoring the zero trade flows, substituting a

---

[4] From section 2, we have 58,523 zero-valued bilateral exports out of (58,123+10,183) non-missing exports

[5] An important number of works related to Africa explicitly deal with zero trade flows by using a Tobit estimator (Foroutan and Pritchett, 1993; Longo and Sekkat, 2004; Bora, Bouët and Roy, 2007) or a sample selection model (Bora, et al., 2007).

positive number for the zero trade flows and then taking the logarithms, or using count models such as Tobit or Poisson (Santos-Silva and Tenreyro, 2006).

At least partially motivated by the presence of zeroes in trade data, Helpman et al. (2007), developed a model that explains trade in the context of firm heterogeneity. Their model predicts a number of facts evidenced in the trade data such as zero trade flows, asymmetry between flows, and a small number of exporting firms. Their econometric implementation of the model is a gravity equation with two extra variables: one corrects for firm self-selection into trade and the other for sample selection bias. The sample selection bias is the bias to which we referred in the previous paragraph, resulting from ignoring the zero trade flows when estimating Equation 1. The estimation of Helpman et al.'s (2007) model proceeds in two steps: the first uses a Probit model to determine the probability of any two given countries engaging in trade, conditional on gravity variables (i.e., distance, borders, etc.). The second step fits an equation explaining the volume of trade, once countries decide to trade. This two-step procedure is similar to Heckman's 2-step solution to the sample selection model (Greene, 2003) and offers an elegant way of dealing with zero-trade flows.

The use of a sample selection model to deal with zero valued trade flows is also found in the work of Francois and Manchin (2007), who used a selection-based gravity approach to study the effects of infrastructure and institutions on bilateral patterns of trade. In the African context Bora, et al. (2007) mention robustness tests to the exclusion of zero-valued flows using a sample selection model. Linders and De Groot (2006) proposed to use a sample selection model, using a profitability argument. To illustrate, let profits be represented by π, and let the indicator variable Z=1 when trade is positive and Z=0 when it is zero. Then, the decision of country i to export product k to country j can be expressed as:

$$Z_{ijk} = \begin{Bmatrix} 1 \; if \;\; \pi_{ijk} > 0 \\ 0 \; otherwise \end{Bmatrix} \tag{2}$$

The implication of Expression 1 is that there is no trade unless it is profitable. This allows modeling the decision of exporting as a discrete choice problem using the following *selection equation*:

$$Z_{ijk} = \mathbf{w}'\boldsymbol{\gamma} + v_{ijk} \tag{3}$$

Where $\mathbf{w}$ is the matrix of the independent variables in Equation 1 (distance, dummy for borders, etc.), $\boldsymbol{\gamma}$ is a vector of parameters to be estimated, and v_{ijk} is an error term assumed to be standard normally distributed. Once a country finds it profitable to export (Z_{ijk}=1), the gravity equation can be estimated whenever Z_{ijk} =1. Because exports are observed only when Z_{ijk} =1, the dependent variable X_{ijk} is not randomly selected, or it is *incidentally truncated from below*. It is this incidental truncation what might cause parameter bias when the zero-valued exports are not properly taken into account. To correct the bias, Equation 3 is used to calculate the inverse Mills Ratio -- $\lambda(z) = \phi(z)/\Phi(z)$, where z is the predicted value of Z_{ijk}, $\phi(z)$ is the standard normal density of z, and $\Phi(z)$ is the standard normal cumulative distribution function. In the presence of sample selection, Equation 1 can be rewritten as an *outcome equation*:

$$\ln(X_{ijk}) = \mathbf{v}'\boldsymbol{\beta} + \beta_\lambda \lambda(\hat{z}) + u_{ijk} \tag{4}$$

Where $\beta_\lambda = \rho\sigma_u$ is the parameter that corrects the estimates from the incidental truncation mentioned above. Notice that if $\lambda(z)$ is statistically significant, Equation 1 suffers from an omitted variable bias whenever sample selection is important. In Equation 4 depending on exclusion restrictions, $\mathbf{v}$ is a subset of $\mathbf{w}$. The exclusion restrictions are needed to identify the model; failure to correctly specify these restrictions leads to imprecise parameter estimates (Wooldridge, 2002). $\boldsymbol{\beta}$ is a vector of parameters to be estimated, and u_{ijk} is a stochastic error assumed to be well-behaved with variance σ_u. Because of the dependence between X_{ijk} and Z_{ijk} (i.e., X_{ijk} > 0 if Z_{ijk} =1), the errors of both equations are potentially correlated as measured by the coefficient of correlation ρ. Formally, u_{ij} and v_{ij} are bivariate normally distributed: $(u_{ij}, v_{ij}) \sim$ bivariate normal $[0, 0, 1, \sigma_u, \rho]$.

Estimation of Equations 3 and 4 can be done using Heckman's 2-step procedure (see Greene, 2003 p.784), or, as we do, by using Maximum Likelihood and jointly estimating $\boldsymbol{\gamma}$, $\boldsymbol{\beta}$, σ_u and ρ. An added advantage of the sample selection model for the problem at hand is that, besides predicting trade values, it estimates the probability that currently missing values actually took place.

Data

The source of the export data is the GTAP Data Base discussed in the previous section. The variable religion is based on Sala-i-Martin (1997)[6]. This variable indicates whether the same religion prevails in any two countries. In Appendix 2, we show for each country the dominant religion and the percentage of the population that practices it. There are 13 regional economic groups in Africa; in most cases, a given country belongs to two or more of these groups. In order to create the variable $TRTY_{ij}$, each country was assigned to only one agreement[7], as shown in Appendix 2. The chosen agreements are the ones with the largest trade values according to The World Bank (2006). Appendix 2 also shows, for each country, the closest and farthest partner (indicating minimum and maximum distance), the number of different countries with which it shares a border, the language spoken, whether it is landlocked or not, and the colonizer. This information was used to construct the dyadic variables ($DIST_{ij}$, $BORD_{ij}$, $LANG_{ij}$, $LOCK_{ij}$, $COLY_{ij}$) and comes from CEPII, as documented in Mayer and Zignago (2006)[8].

Regarding the importance of each regressor, we note that 7.06% of the country pairs share a border, 43.68% speak the same language, 7.33% are landlocked, 26.14% were colonized by the same country, 39.34% share the same religion, and 13.48% belong to the same agreement. The mean distance between countries is 3,711.42 km with a standard deviation of 35.77 km. From Appendix 2, we see that the least distance between economic centers is 10.48 km (between Congo and the Democratic Republic of Congo), while the largest distance is 9,677.82 km (between Mauritius and Cape Verde).

---

[6] Available at http://www.columbia.edu/~xs23/data.htm

[7] Madagascar, Mauritius, Malawi, Congo DR, Zambia and Zimbabwe are part of both COMESA and SADC; because both of these groups are important, we created a group (COMESADC) for these countries.

[8] Available at http://www.cepii.fr/anglaisgraph/bdd/distances.htm.

4. Results

Ordinary Least Squares vs. Sample Selection Model

The first column in Table 3 shows the estimates of Equation 1 using OLS. Recall that this estimation considers only positive exports. Except for sharing a common colonizer, all the included regressors have a statistically significant effect. The coefficient on distance implies that exports tend to be larger when they are destined for closer countries: a 1% increase in distance reduces exports by 0.862%. For a clearer interpretation of the dummy variables (commonality of borders, languages, etc.), next to each OLS estimate, we have calculated the effect of the dummy on exports in percentage terms[9]. For instance, for any given country, exports to an adjacent neighbor are 111.07% higher than exports to non-neighboring countries. Likewise, on average, a given country's exports to a country with the same language are 53.57% higher than exports to countries with different languages. In contrast, the exports of a landlocked country to another landlocked country are lower than the average level of intra-African exports by 41.02%. Another cultural trait with positive effects on export volume is religion: on average, exports to a country with the same dominant religion are 11.96% larger than exports to countries with another religion. In the trade policy front, any given country exports 40.35% more to countries that are members of the same regional agreement than to non-member countries.

The second panel in Table 3 shows the coefficients of the selection equation (Equation 3 in the preceding section). Recall that the dependent variable in this case is an indicator variable, taking the value of one whenever exports are greater than zero. Notice that the same variables that affect the *volume* of exports from *i* to *j* affect (in the same direction) the *probability* of *i* exporting to *j*. That is, the likelihood of any two countries engaging in trade reduces as distance increases and it increases by sharing a border, a language, a colonizer, a religion, or by belonging to the same treaty.

[9] Using $\exp(\hat{\alpha})-1$, where $\hat{\alpha}$ is the parameter estimate.

Interestingly, the coefficient for landlocked partners has a significant effect on increasing the probability of these countries exporting to each other. However, turning our attention to the *outcome equation* (Equation 4 in the previous section) -- the fourth column in Table 3 shows that the same variable has a significantly negative effect on export volumes. A potential explanation of this finding is that to export their merchandise, landlocked countries are more dependent than coastal countries on their immediate neighbors; it is reasonable to think that this dependence is even higher when both countries are landlocked, as the results of the selection equation suggest. However, because of this same dependency, the exports of landlocked countries are disproportionately affected by factors that might make trade difficult such as lack of infrastructure, war or political conflicts, economic shocks, etc. caused by their neighbors. Again, the impeding role on trade of these factors is likely to be exacerbated when both countries are landlocked, which is reflected in the outcome equation.

Notice that in the estimation of Equation 4, we have excluded the variable religion. This is our exclusion restriction, selected by following Helpman, et al. (2007). These authors argued that "trade barriers that affect fixed trade costs but do not affect variable trade cost should only be used as explanatory variables in the selection equation." Although we have not explicitly used Helpman et al.'s model, their argument is valid in the context of our model as well. From an econometric viewpoint, having a smaller set of explanatory variables in the outcome equation helps to improve model identification (Wooldridge, 2002). Helpman et al. pointed out that common language is a good candidate for the exclusion restrictions too. We tried using language and common colonizer as exclusion restrictions; however, they did not change the parameter estimates in a meaningful way, so we kept religion as the exclusion restriction.

The case for using the sample selection model lies in having OLS estimates that are biased. To compare the OLS and the outcome equation estimates, we estimate again Equation 1 but excluding religion; these new estimates, along with the dummy effects in percentage terms, are shown in the last two columns of Table 3. Notice that the outcome equation estimate on distance (-1.595) is almost two times larger than the correspondent OLS estimate (-0.883). This is an indication of the important role of distance in totally suppressing trade flows within the African continent. Just as in the case of distance, the estimated border effect in the outcome equation is almost two times larger than the effect under OLS. The difference is dramatic: when zero-trade flows are taken into account, sharing a neighbor increases exports by 242.12%. However, if we only look at the countries that already export, the role of vicinity is more limited (it increases trade by 110.22%). Language commonality also plays a larger role in determining export values when the zeroes are factored in (103.20% vs. 55.27% when using OLS with positive flows). The effect of both countries being landlocked is lower in the outcome equation (-30.72% vs. -41.49% in the OLS case). A difference with the OLS estimates is that while having the same colonizer has a negative but insignificant impact in the OLS case (-10.06%), it appears positive and significant once that sample selection is controlled for. The difference is far from trivial: sharing the same colonizer is associated with bilateral exports 25.61% higher. Finally, membership in the same economic regional agreement appears to be higher in the outcome equation than in the OLS equation (47.70% vs. 39.93%). The differences between the parameter estimates of the outcome equation and the OLS equation confirm that by ignoring the zero-valued exports, the OLS coefficients underestimate the role of distance, border, language, same colonizer, sea access, and economic agreements on bilateral exports.

A more formal test of the sample selection effects is provided in the last row of the Selection Equation in Table 3. The coefficient of correlation (ρ) between the residuals of Equations 3 and Equation 4 is positive and significant, implying that there would be a significant sample selection bias by using positive exports only.

Predicting missing trade

The equations explained above are used to predict the missing trade flows described in Section 2. In order to gauge the predictive power of both the OLS and the sample selection model, we calculated simple correlation coefficients between observed log-exports (positive) and the corresponding predicted values. The correlation between observed values and the predicted values obtained using the OLS estimates is 50.96%, while the correlation between observed values and predicted values using the sample selection model is 56.60%. Although these coefficients are rather low, the sample selection estimates seem to be superior to the OLS estimates.

The estimating equation shown in Table 3 imposes the same parameters for all the different commodities. It is important then to explore if there are any differences between the predictive power of the sample selection equation across commodities. In the fourth column of Table 2, we show the correlation coefficient between observed and predicted values for each commodity group. In order to facilitate comparisons, we have grouped the different GTAP commodities in three groups: agricultural products, extractive industries, and manufacturing. The results clearly show that the estimates discussed before do a poor job in predicting most of the raw agricultural products. For instance, our predicted values are negatively correlated with the observed exports in sugar cane (c_b), paddy rice (pdr), fibers (pfb), and wool (wol). In general, except for vegetables, fruits and nuts (v_f), and cereal grains (gro), the correlation coefficients are rather low. In the extractive industries, the predictions are poor for coal and oil (0.10 and -0.21 correlation coefficient respectively) but better for gas and minerals (omn). In the broader group of manufactures, the correlation between observed and predicted values ranges from a low of 0.03 for other metals (nfm) and 0.15 for bovine meat products (cmt) to a high of 0.93 for metal products (fmp). Over a third of the manufacturing commodities show a correlation greater than 75%, and in most cases, the correlations are larger than 50%.

In order to explore whether the behavior of the correlations is a consequence of imposing the same set of parameters on disparate products, we estimated new equations for the three groups shown in Table 2: agricultural products, extractive industries, and manufacturing[10]. The results are shown in Table 4. Starting from the left, the outcome equation for the extractive industries shows that, except for language, none of the included regressors has explanatory power. This is a clear indication of the inadequacy of our gravity model to explain trade in this sector. Focusing on the coefficients of the outcome equations using agricultural products and manufactures, we verify that the underlying models are

[10] Due to their very low correlation coefficients, we include bovine meat products (cmt) as an agricultural product and other metals (nfm) as an extractive industry product.

also different. For instance, distance has a greater effect on manufacturing (-1.681) than on agriculture (-1.237). Sharing a border, on the other hand, is more important for agricultural exports than for manufacturing (1.867 vs. 1.042). Also, exports between landlocked countries are lower in agricultural products than in manufactures (-0.736 vs. -0.265). The effect of having the same colonizer is almost negligible in the case of agricultural products, while membership in the same treaty is more important for agriculture than for manufacturing (0.486 vs. 0.348).

Regarding the quality of the predictions, we observe that the correlation coefficient between observed and predicted values is 1.75% for products of extractive industries, 24.38% for agricultural products, and 69.75% for manufacturing. The last column in Table 2 shows the correlations by commodity category, using group-specific equations. The results in this column are nearly equal to the correlations obtained using the full sample. They support the idea that the model performs better for the manufacturing sectors than for the agricultural and extractive industries. Because of this evidence, our final predictions are only for manufactures, and are calculated using the parameters of the "Manufacturing" model, shown in the first two columns of Table 4.

A closer look at the predicted values

Focusing on manufactures reduces the number of the missing trade flows to be estimated. There are 22 manufacturing sectors included in the estimation process. This translates into 62,964 potential trade flows (i.e., 54 exporters x 53 importers x 22 commodities). Of these, 25,942 appear as missing[11], 29,078 are zero-valued trade flows, and only 7,944 are positive trade flows. In other words, 41.20 percent of the potential trade flows are considered missing.

An important point to keep in mind is that the predictions from the outcome equation are in logarithms. From Equation 4:

$$\hat{l}_{ijk} = E(\ln(X_{ijk}) \mid \mathbf{v}; \widehat{\boldsymbol{\beta}}) = \mathbf{v}'\widehat{\boldsymbol{\beta}} \tag{5}$$

---

[11] BACI indicates 26,400 flows as missing. Of these, 458 appear as positive in GTAP; hence, we consider 26,400 - 458 = 25,942 missing flows.

Where $E(\ln(X_{ijk}) \mid \mathbf{v}; \widehat{\boldsymbol{\beta}})$ is the expected value of the log-exports (some of them considered missing) conditional on gravity variables $\mathbf{v}$ obtained using estimated parameters $\widehat{\boldsymbol{\beta}}$. We rewrite these expected conditional values as $\hat{l}_{ijk}$, to emphasize that they are the predicted log-exports of product *k* from country *i* to country *j*. We are interested in the exports, not in their logarithmic equivalent. Wooldridge (2006 p.218) points out that simple exponentiation of $\hat{l}_{ijk}$ will systematically underestimate the predicted value and suggests the following transformation[12]:

$$\widehat{X}_{ijk}^{s} = e^{\sigma^2/2} \times e^{\hat{l}_{ijk}} \tag{6}$$

Where σ^2 is the variance of the estimator $\hat{l}_{ijk}$. Expression 6 is the mean of a log-normal distributed random variable, implying the that $\hat{l}_{ijk}$ is assumed to be normally distributed (Greene, 2003 p. 112). The normal distribution of $\hat{l}_{ijk}$ is in turn a consequence of assuming the residuals of Equation 4 to be normally distributed.

The predicted log-exports $\hat{l}_{ijk}$ are subject to two sources of uncertainty. First, they inherit the uncertainty of the estimated $\widehat{\boldsymbol{\beta}}$ of Equation 4. Second, the distance between the missing true value and the predicted value is, of course, unknown. The first source of uncertainty is summarized by $\hat{\sigma}_u^2$, the variance of the residuals of Equation 4. The second takes into account how far are the explanatory variables of the unknown observations from the explanatory variables employed during estimation. This presents us with two candidates for an estimator of σ^2 in Expression 6. Greene (2003 p.112) suggests that in large

[12] Let *y* = ln(Y) and *y** be a consistent and unbiased estimator of *y*. Assume *y* is distributed normally, hence Y is distributed lognormal. Just as in our problem assume that we are interested in recovering the expected value of Y from *y**. Goldberger (1968) notes that exp(*y**) is an estimator of the conditional median of Y; because of the skewness of the log-normal distribution, this median is always lower than the expected value (mean). The correction suggested by Wooldridge (2006 p. 218) aims to recover the expected conditional mean. An important point is that both the conditional expected mean (from an expression such as 6 in the text) and median exp(y*) are biased. Wooldridge (2006) notes that there is not an unbiased estimator for the expected value case. Kennedy (1983) discusses a correction for the median.

samples the variance of the predicted values will converge to $\hat{\sigma}_u^2$. We scaled our predictions using both, the estimated variance $\hat{\sigma}_u^2$ (see last row of the manufacturing equations in Table 4) and the variance of the predicted $\hat{l}_{ijk}$. The results are similar and we decide to use $\hat{\sigma}_u^2$ to estimate the $\widehat{X}_{ijk}^s$ discussed here[13].

In what follows, we refer to the $\widehat{X}_{ijk}^s$ as the point estimates of the predicted exports of product *k* from *i* to *j*. A reasonable question then is if in light of the uncertainty discussed above, our predicted values are plausible given the underlying data generating process that we tried to capture with the regressions. We used the prediction standard deviation of each $\hat{l}_{ijk}$ and built 95% confidence intervals. All the predicted values fall between the boundaries of the interval suggesting that the out-of-sample predictions are consistent with the same data generating process that underlies the estimates in Table 4. It should be noticed that the confidence interval are rather wide, a point to which we will come back below after discussing the size of the estimates.

Before discussing the magnitude of the estimated missing trade by country and for Africa as a whole, Table 5 helps in an understanding of the predictions of our model in more detail. There, we show the information on missing exports for Kenya, chosen because it has a reduced number of potentially missing exports. This can be seen in Appendix 1, where Kenya has missing exports with five partners: Congo (COG), Cape Verde (CPV), Guinea Bissau (GNB), Saint Helena (SHN), and Sao Tome and Principe (STP). Each country is a panel in Table 5. The first column of each panel shows the probabilities of Kenya exporting each of the commodity categories to each of its partners. These probabilities are obtained from the selection equation (Equation 3 with results shown in Table 4 under the label "manufacturing") and are expected values conditional on the gravity factors used as regressors. Next to the probabilities, we show the predicted exports (point estimates) followed by the upper and lower bound of the 95% confidence interval.

[13] The standard deviations and the upper and lower limits of the prediction intervals are included in a HAR file described at the end of this section.

Starting from the first row, notice that the probability of Kenya exporting beverages and tobacco products to Congo is 15%. The expected value of the exports is 104 thousand USD. The exports with the highest probability of actually had taken place are those of chemical, rubber, and plastics (57%) and other machinery and equipment (56%). The values of these exports are 935 and 567 thousand USD. A glance at the probabilities of Kenya exporting individual products to the other three countries reveals that they are quite low. At the extreme is Guinea-Bissau, for which all the probabilities are close to zero. Consistent with the low probabilities of trade taking place, we see that the estimated exports from Kenya to Guinea-Bissau are only 9 thousand USD. A similar situation occurs with Sao Tome and Principe and Cape Verde.

At the country level, the largest predicted values are for Congo (COG) and the Democratic Republic of Congo (ZAR). Indeed, they are too large to be credible. In the case of Congo, the predicted exports surpass 23 billion USD – this is equivalent to 1,210 times its observed manufacturing exports. For the DR of Congo, the predicted values are around 16 billion USD, or 5,307 times its observed value. To put these values in perspective, consider that South Africa, the largest exporter of manufactures in sub-Saharan Africa, sells approximately 7 billion USD in manufacturing. Consider further that the next highest predicted value is 47 million USD. The vast majority of these predicted exports are bilateral (from Congo to DR Congo, and vice versa). The reason for these extremely high values is that the distance between the economic centers of the two countries is extremely short in comparison with other distances: according to our sources, the economic centers of Congo and the DR Congo are 10.48 km. apart, while the next closest pair of partners (Benin and Nigeria) are separated by 105.18 km (See Appendix 2 for the closest and farthest partners).

In Table 6, we show the predicted exports at the country level. The first four columns of Table 4 show the exports for each country (total and manufacturing), the share of manufacturing exports in total exports, and the number of flows for which trade is considered missing as a percentage of potential flows. The fifth column contains the predicted exports using the parameter estimates of the sample selection model for manufactures; we have sorted the countries using the values of this column. In the last column we show the ratio of predicted missing exports to the sum of observed exports plus the predicted exports themselves. This value ranges from 0 to 100. The close this value is to a 100, the larger the estimated missing trade relative to what is currently known, and vice versa. The idea is that if we consider the estimates a reasonable measure of the missing trade, the ratio just described measures missing trade as a percentage of total trade (missing and not missing). In Table 6, first row, we can see that the largest estimates of missing trade are for Morocco, with 46,864 thousand USD; if we consider this estimate as a reasonable measure of Morocco's missing exports and add them to Morocco's observed exports (293,845 thousand USD) , the last column of Table 6 shows that 14% of Morocco's manufacturing exports are missing. Next is Egypt, with 29,486 thousand USD, this would be 10% of total Egypt's manufacturing exports. For some countries, predicted trade is a high share of what we currently observe. For instance, the prediction for Angola is 21,273 thousand USD; this means that 81 percent of Angola's manufacturing exports would be missing. For the vast majority of countries, the predicted trade is below 10 million USD; for a third of the countries, the predictions are less than 1 million USD.

At the aggregate level the bottom of Table 6 shows that, excluding the predicted bilateral trade between Congo and the DR of Congo, and adding the predicted missing exports for all the countries, our point estimate of missing trade in Africa is 297,066 thousand USD. This implies that 2% of intra-African trade in manufactures is missing. The upper bound of our predicted interval is 2,097,704 thousand USD and the lower bound is 37,261 thousand USD. In general the upper bound for individual predictions is between 7 and 10 times larger than the point-estimates. For individual predictions the lower limit of the interval is bounded by zero because of the anti-log transformation. The anti-log transformation also gives asymmetric predicted intervals. In Appendix 3 we show the upper and lower bounds of the predicted exports by country.

Table 6 evidences how the importance of missing exports in total exports varies widely. In one extreme are the countries of the SACU (last five rows of Table 6) that do not have missing exports, while in the other extreme are countries such as Guinea-Bissau, Equatorial Guinea, Comoros and Eritrea. For all these countries over 96% of their manufacturing exports are missing (defining total exports as the sum of estimated missing exports and observed exports). A pattern that emerges from Table 6 is that, the larger the number of missing exports, the larger is the estimated missing trade as a percentage of total trade. This is seen in a clearer way in Figure 1 where we have plotted the count of missing exports over potential exports against missing trade over total trade. As the figure suggests, there is a positive and important degree of correlation between these two measures. It is worth to emphasize that these are relative measures, and they say little about the estimated volume of exports. For example, back in table 6 we see that for Equatorial-Guinea the predicted exports are only 1,047 thousand USD, even though it would represent 98% of total Equatorial-Guinea's exports. Because 75% of Equatorial-Guinea's exports are missing, this country appears high in the upper-right plot in Figure 1 (point EQN in the plot).

Just as the importance of missing exports in total exports is expected to be correlated with a high number of missing transactions, one would expect that larger missing volumes should be associated with larger countries. To explore this relationship, Figure 2 plots the exporter fixed effects estimated from Equation 4 against the predicted export values. We use the fixed effects because they reflect the relative effect of country size on the exporting role of the country. For example in Figure 3 we can see that Kenya (KEN) and Sudan (SDN) have comparable GDP levels[14], however, Sudan's fixed effect is much lower than Kenya's due to a lower average level of exports. Back in Figure 2, the plot confirms that country with higher fixed effects have higher estimated values. We also notice that many of the countries to the left of the zero mark in the horizontal access of Figure 2 (zero corresponds to Angola, the omitted category in the estimation) are those with a high share of missing trade in their total trade.

Another dimension to explore is whether there seems to be any spatial pattern in the incidence of missing trade. In Figure 2 we have geographically mapped our two relative measures of missing trade: the ratio of missing exports to total exports and the ratio of the number of missing flows to potential number of exports. We use black dots for the former and grey points for the latter. Each point represents a tenth of the relevant ratio, or 10%. For example, in Madagascar we see three grey points and one black point. This means that 30% of the potential exports that Madagascar can send to her African partners are

[14] Data on GDP is from the World Bank's World Development Indicators Online, series GDP (constant 2000 USD), year 2001.

missing (this can be verified in Table 6, in the column "% Missing Exports"). The presence of only one black point implies that our estimated missing trade value for Madagascar is equivalent to 10% of its total trade. Table 6 confirms that the exact count of Madagascar's missing exports is 6%. The distribution of grey and black points in Figure 2 suggests that the incidence of missing trade is higher in the countries of Central and West Africa, moderate in North Africa and unimportant in Southern Africa.

Table 7 shows the value of predicted exports by manufacturing sector. Almost a sixth of the predicted exports are in the chemical, rubber, and plastics category. This sector, together with food products and machinery and equipment, account for a third of the predicted values. In the last two columns of Table 7, we show the percentiles 50 and 99 for the individual probabilities of trade taking place, calculated using the selection model (Equation 3 with results in Table 4). The percentile 99 indicates that 99% of the estimated probabilities are below the value shown in Table 7. Likewise, the percentile 50 indicates that half of the estimated probabilities are below the value shown in Table 7. The percentile 50 is the median value. The general pattern that emerges is that the probabilities of trade taking place are generally low. The highest percentile 99 probability is 0.74 (in the machinery and equipment sector), implying that 99 percent of the estimated probabilities are below this value. The sector with the highest percentile 50 probability is chemical/rubber/plastics, with just 6%. In other words, half of the estimated probabilities are below 6%[15].

5. Closing Remarks

The objective of this paper was to get an approximation of the size of unknown trade flows among African countries. Our definition of missing trade is simply the trade that might have taken place but about which we do not have information because neither partner reported it.

[15] This paper accompanies a HAR file with three dimensions: exporters (54), importers (54), and GTAP manufacturing categories (22) containing the following headers: estimated individual probabilities of each flow, predicted values (point estimates), lower bounds of expected values, upper bounds of expected values, and the standard deviations of the predictions.

Using GTAP export data and the information on missing trade contained in the CEPII's BACI dataset, we determined that approximately 41% of the potential bilateral transactions between 54 African countries were missing. For estimating the trade flows, we employed a gravity model. A feature of intra-African trade is that 85% of the known transactions are zero valued; however, the standard gravity model only handles positive trade flows. The zero-valued flows are likely to contain important information on the determinants of trade; thus, their exclusion might generate biased estimates. To overcome this limitation, we employed a sample selection model that explicitly takes into account the effects of the gravity variables on the size of the trade flows, even when they are zero. Our results strongly support the sample selection model estimates against standard OLS estimates. At the product level, we found that the gravity model poorly fits the data on agricultural products and extractive industries. For these reasons, we focused on manufacturing.

The predicted manufacturing exports are close to 300 million USD, approximately 2.5% of what is currently observed. The importance of missing exports varies across countries. In general, the larger the number of missing transactions, the larger the predicted estimates relative to the observed. Geographically, Southern Africa has a very low incidence of missing exports; in turn, missing exports are more important for the countries of West and Central Africa than in North Africa. In most cases, the predicted export values are moderate in size.

6. References

Anderson, J. E., and E. v. Wincoop. 2003. "Gravity with Gravitas: A Solution to the Border Puzzle." *The American Economic Review* 93, no. 1: 23.

Bora, S., A. Bouët, and D. Roy. 2007. "The Marginalization of Africa in World Trade." *International Food Policy Research Institute Research Brief*, no. 7.

Coe, D. T., and A. W. Hoffmaister. 1999. "North-South Trade: Is Africa Unusual?" *Journal of African Economies* 8 2: 228-56.

Coulibaly, S., and L. Fontagne. 2006. "South-South Trade: Geography Matters." *Journal of African Economies* 15, no. 2: 313-341.

Feenstra, R. 2002. "Border Effects and the Gravity Equation: Consistent Methods for Estimation." *Scottish Journal of Political Economy* 49, no. 5: 491-506.

Foroutan, F., and L. Pritchett. 1993. "Intra-Sub-Saharan African Trade: Is It Too Little?" *Journal of African Economies* 2, no. 1: 74-105.

Francois, J. F., and M. Manchin. 2007. "Institutions, Infrastructure and Trade." Working Papers No. 224 (January, 2007), Centro Studi Luca D'Agliano Development Studies. Milan, Italy.

Gaulier, G., S. Zignago, D. Sondjo, A. A. Sissoko, and R. Paillacar. 2007. "Baci: A World Database of International Trade at the Product-Level, 1995-2004 Version." CEPII Working Paper No. 2007, Centre d'Etudes Prospectives et d'Informations Internationales. Paris, France.

Gehlhar, M. (2007) "Reconciling Merchandise Trade Data (Chapter 15-B)." In B. Dimaranan, ed. *Global Trade, Assistance, and Production: The Gtap 6 Data Base*. West Lafayette, IN, Center for Global Trade Analysis, Purdue University.

Goldberger, A. S. 1968. "The Interpretation and Estimation of Cobb-Douglas Functions." *Econometrica* 36, no. 3/4: 464-472.

Greene, W. H. 2003. *Econometric Analysis*. Fifth Edition ed. New Jersey: Prentice Hall.

Helpman, E., M. Melitz, and Y. Rubinstein. 2007. "Estimating Trade Flows: Trading Partners and Trading Volumes." Working Paper No. 12927, National Bureau of Economic Research Cambridge, MA.

Hummels, D. 2001. "Toward a Geography of Trade Costs." Working Paper, Purdue University.

Kennedy, P. 1983. "Logarithmic Dependent Variables and Prediction Bias." *Oxford Bulletin of Economics & Statistics* 45, no. 4: 389-392.

Linders, G.-J., and H. L. De Groot. 2006. "Estimation of the Gravity Equation in the Presence of Zero Flows." No. TI 2006-072/3, Tinbergen Institute Discussion Paper.

Longo, R., and K. Sekkat. 2004. "Economic Obstacles to Expanding Intra-African Trade." *World Development* 32 8: 1309-21.

Mayer, T., and S. Zignago. 2006. "Notes on Cepii's Distances Measures." Centre d'Etudes Prospectives et d'Informations Internationales. Retrieved March 20, 2007 from http://www.cepii.fr/anglaisgraph/bdd/distances.htm

McCallum, J. 1995. "National Borders Matter: Canada-U.S. Regional Trade Patterns." *The American Economic Review* 85, no. 3: 615-623.

Redding, S., and A. Venables. 2004. "Economic Geography and International Inequality." *Journal of International Economics* 62: 53-82.

Rodrik, D. 1998. "Trade Policy and Economic Performance in Sub-Saharan Africa." Working Paper No. 6562, National Bureau of Economic Research. Cambridge, MA.

Rose, A. K. 2004. "Do We Really Know That the Wto Increases Trade?" *The American Economic Review* 94, no. 1: 98-114.

Sala-i-Martin, X. 1997. "I Just Ran Two Million Regressions." *American Economic Review* 87, no. 2: 178-83.

Santos-Silva, J. M. C., and S. Tenreyro. 2006. "The Log of Gravity." *The Review of Economics and Statistics* 88, no. 4: 641-658.

The World Bank. 2006. "Africa Development Indicators." The World Bank. Retrieved Dec 30, 2005 from http://www4.worldbank.org/afr/stats/adi2005/default.cfm

Wooldridge, J. M. 2002. *Econometric Analysis of Cross Section and Panel Data*. Cambridge, Mass.: MIT Press.

Wooldridge, J. M. 2006. *Introductory Econometrics : A Modern Approach*. 3rd ed. Mason, OH: Thomson/South-Western.

Table 1. Intra-African imports and exports (in USD 1000) and missing transactions (counts), by country.

Country (ISO code)	Imports	Missing M.	Exports	Missing X
Algeria (DZA)	134,428	13	284,115	24
Angola (AGO)	407,528	28	5,607	31
Benin (BEN)	100,592	12	46,536	20
Botswana (BWA)	1,421,271	3	214,475	-
Burkina Faso (BFA)	126,653	20	69,696	25
Burundi (BDI)	37,814	39	11,125	37
Cameroon (CMR)	450,560	11	131,196	8
Cape Verde (CPV)	5,640	32	149	34
Central African Republic (CAF)	12,819	30	4,546	29
Chad (TCD)	12,821	33	6,139	38
Comoros (COM)	18,818	43	50	40
Congo (COG)	123,402	28	28,771	28
DR Congo (ZAR)	146,542	31	5,195	30
Côte d'Ivoire (CIV)	135,603	9	421,323	5
Djibouti (DJI)	89,003	38	51,024	39
Egypt (EGY)	378,706	29	323,143	19
Equatorial Guinea (GNQ)	6,654	40	2,812	40
Eritrea (ERI)	3,762	44	280	42
Ethiopia (ETH)	104,901	10	94,957	22
Gabon (GAB)	51,350	19	95,681	16
Gambia (GMB)	41,833	18	2,766	23
Ghana (GHA)	691,454	25	176,950	21
Guinea (GIN)	71,447	13	115,973	14
Guinea-Bissau (GNB)	944	42	114	44
Kenya (KEN)	396,649	12	483,913	5
Lesotho (LSO)	676	3	506	-
Liberia (LBR)	11,107	35	20,240	35
Libyan Arab Jamahiriya (LBY)	332,255	37	386,862	36
Madagascar (MDG)	169,241	14	43,275	16
Malawi (MWI)	300,936	21	136,679	15
Mali (MLI)	216,780	21	237,815	15
Mauritania (MRT)	62,563	23	32,832	28
Mauritius (MUS)	353,795	16	197,543	12
Morocco (MAR)	567,373	29	309,810	20
Mozambique (MOZ)	719,313	27	54,468	26
Namibia (NAM)	1,189,512	3	549,534	-
Niger (NER)	134,248	18	110,141	17
Nigeria (NGA)	354,845	2	1,047,845	14
Rwanda (RWA)	108,635	20	41,315	36
Saint Helena (SHN)	5,203	46	1,223	41
Sao Tome and Principe (STP)	767	38	303	37
Senegal (SEN)	344,479	9	244,027	8
Seychelles (SYC)	32,210	34	9,271	35
Sierra Leone (SLE)	15,069	32	5,375	30
Somalia (SOM)	13,890	43	1,793	40
South Africa (ZAF)	1,983,978	3	7,466,584	-
Sudan (SDN)	183,289	21	57,836	23
Swaziland (SWZ)	797,497	3	522,640	-
Tanzania, United Rep. of (TZA)	344,482	10	116,133	11
Togo (TGO)	92,336	17	158,890	14
Tunisia (TUN)	513,916	8	457,798	7
Uganda (UGA)	221,159	13	150,651	15
Zambia (ZMB)	775,690	14	193,029	17
Zimbabwe (ZWE)	716,374	18	401,858	18

Source: The trade values are from GTAP. The missing flows are from BACI as documented in Gaulier, et al. (2007, p. 25).

Notes: This table summarizes the total imports (in USD 1000) for each African country, the number of partners for which imports are considered missing (column Missing M), the total exports (in USD 1000), and the number of partners for which exports are missing (column Missing X). Both the missing exports and imports are out of a total of 53 potential partners.

Table 2. Intra-African export values (in USD 1000), by GTAP merchandise sector.

Commodity Description (GTAP Code)	Group	Export Value	Correlation of in-sample predictions of regressions using:	
			All Commodities	By Groups
Sugar cane, sugar beet (c_b)	Agricultural Products	308	-0.21	-0.23
Bovine cattle, sheep and goats, horses (ctl)		97,145	0.03	0.08
Forestry (frs)		57,621	0.17	0.22
Fishing (fsh)		13,607	0.16	0.16
Cereal grains nec (gro)		108,272	0.71	0.65
Animal products nec (oap)		55,645	0.45	0.56
Crops nec (ocr)		615,908	0.10	0.20
Oil seeds (osd)		69,486	0.02	0.20
Paddy rice (pdr)		22,945	-0.04	-0.04
Plant-based fibers (pfb)		144,360	-0.01	0.32
Vegetables, fruit, nuts (v_f)		186,217	0.80	0.66
Wheat (wht)		20,074	0.33	0.46
Wool, silk-worm cocoons (wol)		1,182	-0.04	0.01
Coal (coa)	Extractive Industries	140,573	0.10	0.00
Gas (gas)		14,152	0.61	0.64
Oil (oil)		1,256,779	-0.21	-0.16
Minerals nec (omn)		150,725	0.57	0.62
Beverages and tobacco products (b_t)	Manufacturing	396,452	0.49	0.50
Bovine meat products (cmt) *		70,918	0.15	0.32
Chemical, rubber, plastic products (crp)		2,143,272	0.85	0.84
Electronic equipment (ele)		348,162	0.86	0.87
Metal products (fmp)		491,509	0.93	0.94
Ferrous metals (i_s)		535,438	0.74	0.74
Leather products (lea)		107,499	0.83	0.80
Wood products (lum)		320,512	0.87	0.84
Dairy products (mil)		120,421	0.88	0.85
Motor vehicles and parts (mvh)		584,165	0.84	0.85
Metals nec (nfm) **		556,372	0.03	0.07
Mineral products nec (nmm)		594,315	0.79	0.78
Food products nec (ofd)		978,242	0.79	0.76
Machinery and equipment nec (ome)		1,718,479	0.92	0.91
Manufactures nec (omf)		167,780	0.73	0.70
Meat products nec (omt)		70,206	0.64	0.66
Transport equipment nec (otn)		138,187	0.50	0.52
Petroleum, coal products (p_c)		1,281,829	0.55	0.53
Processed rice (pcr)		86,494	0.47	0.43
Paper products, publishing (ppp)		764,645	0.63	0.59
Sugar (sgr)		272,821	0.29	0.29
Textiles (tex)		506,237	0.39	0.37
Vegetable oils and fats (vol)		121,216	0.45	0.44
Wearing apparel (wap)		202,642	0.69	0.67

Notes: This table presents the 41 GTAP commodities grouped in three categories: agricultural products, extractive industries, and manufacturing. For each product, the first column shows total exports. The two last columns show the coefficient of correlation between observed export values and the export values predicted by using the ML parameter estimates from the outcome equation (see equation 4 in the text) using different samples. That is, the column labeled "All commodities" contains the correlation coefficients when all the products are included in the regression (see Table 3 below) while the column "By groups" displays the correlation coefficients from group-specific regressions (see Table 4 below). * Included in the "Agricultural Products" category for the estimations in Table 4 (see "Results" section for more details). ** Included in the "Extractive Industries" category for the estimations in Table 4 (see "Results" section). Source: Export values are from GTAP. Correlations from author's calculation based on regression results.

Table 3. Regression Results: OLS vs. Sample Selection Model

	OLS		Selection Equation	Outcome Equation		OLS (no religion)	
Coefficients	Log(exports)	FE in %	Exports (1,0)	Log(exports)	FE in %	Log(exports)	FE in %
Constant	9.272***		3.668***	11.68***		9.373***	
	(0.54)		(0.17)	(0.56)		(0.54)	
Log(Distance)	-0.862***		-0.692***	-1.595***		-0.883***	
	(0.045)		(0.016)	(0.053)		(0.044)	
Same Border	0.747***	111.07	0.585***	1.230***	242.12	0.743***	110.22
	(0.073)		(0.029)	(0.078)		(0.073)	
Same Language	0.429***	53.57	0.354***	0.709***	103.20	0.440***	55.27
	(0.068)		(0.024)	(0.070)		(0.068)	
Both Landlocked	-0.528***	-41.02	0.0781**	-0.367***	-30.72	-0.536***	-41.49
	(0.11)		(0.039)	(0.12)		(0.11)	
Same Colonizer	-0.102	-9.70	0.191***	0.228***	25.61	-0.106	-10.06
	(0.076)		(0.026)	(0.081)		(0.076)	
Same Religion	0.113**	11.96	0.0510***				0.00
	(0.052)		(0.017)				
Same Treaty	0.339***	40.35	0.119***	0.390***	47.70	0.336***	39.93
	(0.063)		(0.023)	(0.066)		(0.063)	
ρ			0.868***				
			(0.038)				
Observations	10,183		68,706	10,183		10,183	
R-squared	0.33		.	.		0.33	

Notes: This table shows the results of estimating Equation 1 (OLS) and Equations 3 and 4 (selection and outcome equation of the sample selection model). In the OLS and outcome equations, the regressand is the log of exports. In the selection equation, the regressand is a binary variable (1, 0) that equals 1 when exports are observed. The leftmost panel shows the results of using OLS to estimate Equation (1). The column "FE in %" shows the effect of each dummy in percentage terms by using exp(B) -1, where B is the value of the coefficient. The second panel contains the selection and outcome equation estimated using ML. The last panel repeats the estimation of equation (1) but omits the variable, religion. This omission permits us to compare the coefficient of the outcome question and Equation (1). Robust standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1

Table 4. Sample selection model results for commodity groups (agriculture, extractive industries and manufactures)

	Manufacturing		Agricultural Products		Extractive Industries	
Coefficients	Outcome Equation	Selection Equation	Outcome Equation	Selection Equation	Outcome Equation	Selection Equation
Constant	12.91***	4.038***	0.949	0.745*	2.534	1.540***
	(0.59)	(0.20)	(1.93)	(0.40)	(2.71)	(0.55)
Log(Distance)	-1.681***	-0.758***	-1.237***	-0.615***	-1.536	-0.661***
	(0.057)	(0.020)	(0.17)	(0.037)	(2.06)	(0.069)
Same Border	1.042***	0.531***	1.867***	0.781***	1.922	0.728***
	(0.082)	(0.036)	(0.26)	(0.063)	(1.64)	(0.11)
Same Language	0.716***	0.379***	0.736***	0.377***	0.799**	0.158*
	(0.073)	(0.029)	(0.19)	(0.050)	(0.39)	(0.088)
Both Landlocked	-0.265**	0.140***	-0.736**	-0.0655	0.0708	-0.0326
	(0.13)	(0.048)	(0.30)	(0.085)	(0.78)	(0.16)
Same Colonizer	0.234***	0.246***	-0.147	0.0293	0.267	0.162*
	(0.085)	(0.031)	(0.22)	(0.057)	(0.72)	(0.096)
Same Treaty	0.348***	0.0740***	0.486***	0.167***	0.935	0.365***
	(0.070)	(0.029)	(0.17)	(0.048)	(1.13)	(0.10)
Same Religion		0.0431**		0.0510		0.130**
		(0.021)		(0.039)		(0.062)
ρ		0.775***		1.002***		1.317
		(0.042)		(0.17)		(2.54)
Observations	37,022	37,022	23,348	23,348	14,310	14,310
Sigma	2.278594	.	.	.	.	.

Notes: This table shows the sample selection models for the GTAP commodities grouped in three categories: manufacturing, agricultural products, and extractive industries. In the outcome equations, the regressand is the log of exports. In the selection equation, the regressand is a binary variable (1, 0) that equals 1 when exports are observed and zero otherwise. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table 5. Predictions for Kenya: Predicted export values (in USD 1000) and probabilities (%), by partner.

Description	Congo (COG)		Cape Verde (CPV)		Guinea-Bissau (GNB)		Saint Helen (SHN)		S. Tome & Prince (STP)	
	Predicted (1000 USD)	Probability (%)	Predicted (1000 USD)	Probability (%)	Predicted (1000 USD)	Probability (%)	Predicted (1000 USD)	Probability (%)	Predicted (1000 USD)	Probability (%)
Beverages and tobacco products (b_t)	104	15	2	1	0	0	4	4	1	0
Chemical, rubber, plastic products (crp)	935	57	14	11	2	0	38	28	8	6
Electronic equipment (ele)	80	33	1	3	0	0	3	11	1	2
Metal products (fmp)	143	35	2	3	0	0	6	12	1	2
Ferrous metals (i_s)	167	16	3	1	0	0	7	4	1	0
Leather products (lea)	43	21	1	1	0	0	2	5	0	1
Wood products (lum)	121	32	2	3	0	0	5	11	1	1
Dairy products (mil)	37	7	1	0	0	0	2	1	0	0
Motor vehicles and parts (mvh)	139	37	2	4	0	0	6	13	1	2
Mineral products nec (nmm)	188	31	3	3	0	0	8	10	2	1
Food products nec (ofd)	643	40	10	5	1	0	26	15	5	2
Machinery and equipment nec (ome)	567	56	9	10	1	0	23	27	5	6
Manufactures nec (omf)	85	37	1	4	0	0	3	14	1	2
Meat products nec (omt)	17	4	0	0	0	0	1	1	0	0
Transport equipment nec (otn)	44	17	1	1	0	0	2	4	0	0
Petroleum, coal products (p_c)	189	11	3	0	0	0	8	2	2	0
Processed rice (pcr)	32	3	0	0	0	0	1	0	0	0
Paper products, publishing (ppp)	142	32	2	3	0	0	6	11	1	1
Sugar (sgr)	77	5	1	0	0	0	3	1	1	0
Textiles (tex)	229	37	3	4	1	0	9	13	2	2
Vegetable oils and fats (vol)	55	7	1	0	0	0	2	1	0	0
Wearing apparel (wap)	44	24	1	2	0	0	2	7	0	1
Total	4,082.70		61.63		9.19		164.07		32.74	

Notes: This table shows, for each manufacturing category, the predicted exports of Kenya that are currently considered missing. Each panel is a country for which Kenya's exports are considered missing. In each panel, the first column shows the estimated probability of Kenya exporting an individual manufacturing product. For example, the probability of Kenya exporting beverages and tobacco products to Congo is 15%. These probabilities are the expected probabilities, conditional on gravity factors obtained using the values of the selection equation (Equation 3 in the text) with parameters shown in Table 4 above. Next, we show the point estimate of predicted trade. These are the expected values for exports, conditional on the gravity variables, using the parameter estimates of the outcome equation (Equation 4 in the text). For example, for Guinea Bissau, the point estimated predicted exports of chemical/rubber/plastics are 2,000 USD.

Table 6. Total exports, and manufacturing exports (in USD 1000) destined to African countries by African exporters.

Country (ISO Code)	Total Exports	Manufacturing Exports	% Manf/Total	% Missing Exports	Predicted	Predicted/(Predicted + Observed)
Morocco (MAR)	309,810	293,845	95	29	46,864	14
Egypt (EGY)	323,143	275,042	85	25	29,486	10
Djibouti (DJI)	51,024	41,767	82	74	24,441	37
Angola (AGO)	5,607	4,916	88	58	21,273	81
Liberia (LBR)	20,240	10,155	50	64	14,700	59
Congo (Democratic Republic of the) (ZAR)	5,195	3,033	58	54	14,454	83
Libyan Arab Jamahiriya (LBY)	386,862	181,222	47	67	13,483	7
Tunisia (TUN)	457,798	423,525	93	13	12,278	3
Congo (COG)	28,771	19,218	67	51	10,896	36
Ghana (GHA)	176,950	105,581	60	35	9,928	9
Saint Helena (SHN)	1,223	1,223	100	76	6,961	85
Sierra Leone (SLE)	5,375	4,521	84	56	6,866	60
Nigeria (NGA)	1,047,845	91,577	9	25	6,119	6
Eritrea (ERI)	280	230	82	79	5,366	96
Niger (NER)	110,141	49,308	45	32	5,338	10
Chad (TCD)	6,139	1,276	21	71	4,443	78
Kenya (KEN)	483,913	313,776	65	9	4,350	1
Zimbabwe (ZWE)	401,858	234,523	58	34	3,656	2
Uganda (UGA)	150,651	47,694	32	28	3,594	7
Mauritania (MRT)	32,832	27,785	85	52	3,468	11
Malawi (MWI)	136,679	64,008	47	28	3,373	5
Guinea-Bissau (GNB)	114	38	33	83	2,814	99
Togo (TGO)	158,890	144,689	91	26	2,570	2
Madagascar (MDG)	43,275	39,132	90	30	2,470	6
Gabon (GAB)	95,681	13,731	14	30	2,403	15
Mauritius (MUS)	197,543	193,988	98	23	2,212	1
Guinea (GIN)	115,973	31,448	27	26	2,083	6
Burkina Faso (BFA)	69,696	56,502	81	47	2,078	4
Sudan (SDN)	57,836	17,462	30	43	2,034	10
Cameroon (CMR)	131,196	84,664	65	15	1,962	2
Ethiopia (ETH)	94,957	7,662	8	41	1,954	20
Sao Tome and Principe (STP)	303	303	100	70	1,872	86
Seychelles (SYC)	9,271	9,231	100	66	1,721	16
Mali (MLI)	237,815	31,342	13	28	1,562	5
Senegal (SEN)	244,027	217,082	89	15	1,538	1
Zambia (ZMB)	193,029	99,282	51	32	1,527	2

Country (ISO Code)	Total Exports	Manufacturing Exports	% Manf/Total	% Missing Exports	Predicted	Predicted/(Predicted + Observed)
Mozambique (MOZ)	54,468	36,328	67	49	1,513	4
Benin (BEN)	46,536	25,606	55	38	1,449	5
Burundi (BDI)	11,125	3,377	30	69	1,393	29
Somalia (SOM)	1,793	1,322	74	75	1,385	51
Central African Republic (CAF)	4,546	1,680	37	54	1,375	45
Gambia (GMB)	2,766	1,651	60	43	1,370	45
Côte d'Ivoire (CIV)	421,323	388,363	92	9	1,323	0
Algeria (DZA)	284,115	276,482	97	45	1,071	0
Equatorial Guinea (GNQ)	2,812	21	1	75	1,047	98
Tanzania (TZA)	116,133	73,069	63	21	976	1
Cape Verde (CPV)	149	147	99	64	909	86
Comoros (COM)	50	26	52	75	625	96
Rwanda (RWA)	41,315	1,504	4	68	494	25
Botswana (BWA)	214,475	173,000	81	-	-	0
Namibia (NAM)	549,534	411,759	75	-	-	0
South Africa (ZAF)	7,466,584	6,916,844	93	-	-	0
Swaziland (SWZ)	522,640	498,108	95	-	-	0
Lesotho (LSO)	506	455	90	-	-	0
Total	**15,532,812**	**11,950,523**	**77**	**-**	**297,066**	**2**

Table 6. Total exports, and manufacturing exports (in thousand USD) destined to African countries by African exporters.

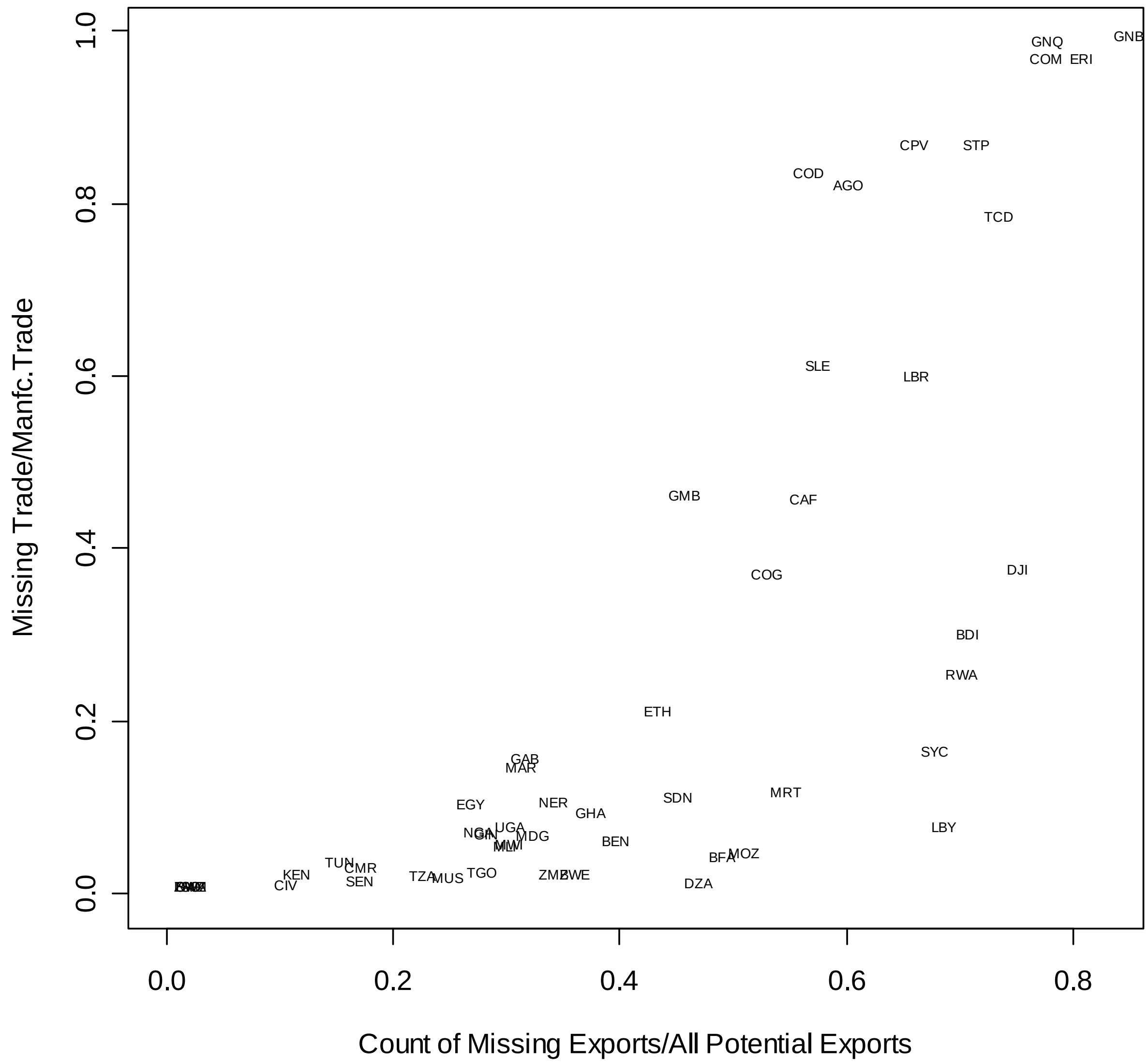
Notes: This table summarizes the predicted values from the sample selection model applied to manufactures (See Table 4 for parameter estimates). In the first two columns, we show total and manufacturing exports (in 1000 USD) by country. The third column is the ratio of manufacturing exports to total exports already expressed in percent terms. The fourth column is the count of missing exports expressed as percentage of all the potential exports; i.e. each country could have up to 53 partners*22 manufacturing commodities = 1,166 exports -- the column shows the number of missing flows as a percentage of these 1,166 exports. The column labeled "Predicted" is the sum, over products, of the predicted exports using the outcome equation of the sample selection model. The last column shows the ratio of the predicted missing exports to the sum of predicted missing exports and observed exports, already in percentage terms. The closer this value is to 100, the larger the size of the estimated missing exports relative to what is currently one. In the limit, if observed exports are zero, the ratio takes the value of 100. For example, towards the end of the table, in the row for Rwanda, we see that Rwanda exports 41,135 thousand USD; out of those, 1,504 thousand USD are manufactures. These manufactures represent 4% of Rwanda's total manufacturing exports. Rwanda has a large number of missing exports: 68%. After we use our gravity estimates for manufactures (see Table 4), to estimate the expected value of Rwanda exporting each GTAP manufacture to each one of its partners, we found that Rwanda has an estimated of 494 thousand USD of missing trade. This missing trade represents 25% of total Rwanda's exports (missing plus observed).

Table 7. Manufacturing exports (observed and predicted, in USD 1000) destined to African countries, by GTAP commodity category.

Description	GTAP Code	Export Value	Predicted Exports	%(Predicted/Observed)	Probability perc. 50	Probability perc. 99
Chemical rubber plastic products	crp	2,143,272.0	54,609.09	0.03	0.06	0.72
Food products nec	ofd	978,242.0	39,440.50	0.04	0.03	0.59
Machinery and equipment nec	ome	1,718,479.0	38,204.60	0.02	0.06	0.74
Petroleum coal products	p_c	1,281,829.0	24,317.29	0.02	0.00	0.32
Textiles	tex	506,237.0	16,115.94	0.03	0.02	0.58
Mineral products nec	nmm	594,315.0	12,577.29	0.02	0.01	0.50
Ferrous metals	i_s	535,438.0	12,252.66	0.02	0.00	0.37
Motor vehicles and parts	mvh	584,165.0	10,945.07	0.02	0.02	0.65
Metal products	fmp	491,509.0	10,704.18	0.02	0.02	0.55
Paper products publishing	ppp	764,645.0	9,974.37	0.01	0.02	0.53
Beverages and tobacco products	b_t	396,452.0	9,608.41	0.02	0.00	0.41
Sugar	sgr	272,821.0	9,116.72	0.03	0.00	0.20
Wood products	lum	320,512.0	8,985.09	0.03	0.02	0.52
Manufactures nec	omf	167,780.0	6,873.40	0.04	0.02	0.65
Vegetable oils and fats	vol	121,216.0	6,800.66	0.06	0.00	0.26
Electronic equipment	ele	348,162.0	6,121.46	0.02	0.02	0.60
Transport equipment nec	otn	138,187.0	4,977.33	0.04	0.00	0.45
Wearing apparel	wap	202,642.0	3,530.72	0.02	0.01	0.53
Processed rice	pcr	86,494.0	3,443.33	0.04	0.00	0.13
Dairy products	mil	120,421.0	3,432.67	0.03	0.00	0.24
Leather products	lea	107,499.0	3,143.30	0.03	0.01	0.39
Meat products nec	omt	70,206.0	1,892.16	0.03	0.00	0.17
Total		**11,950,523.00**	**297,066.25**	**0.02**	-	-

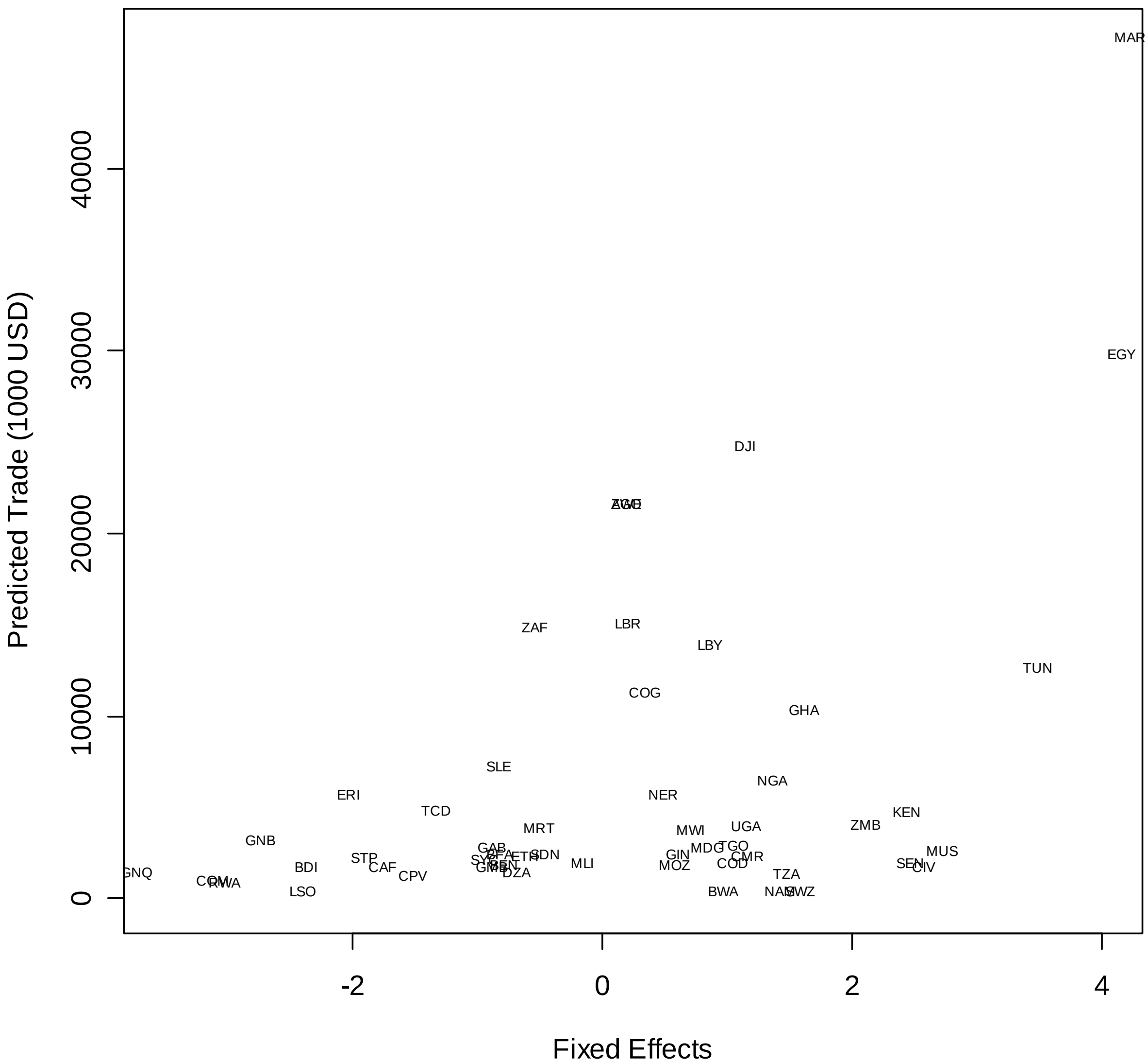
Notes: The first column of this table shows, for each GTAP manufacturing commodity, the observed export value (in thousand USD). The second column shows the sum over countries of the predicted exports. In the third column, we divide the predicted exports by the observed exports to have an idea of the importance of the missing trade at the product level. The column labeled "Probability perc. 50" contains the percentile 50 of the distribution of the estimated probabilities of exports taking place in a given commodity. These probabilities are estimated using the selection equation (Equation 3 in the text) for manufactures (see Table 4). For example, in the first row, for chemical, rubber, and plastic products, the percentile-50 value of 0.06 implies that half of the estimated probabilities of missing exports in this sector actually taking place are below 6%. Another way of looking at this is that, for half of the flows considered missing in the chemical, rubber, and plastics commodity, the probabilities of these flows to be missing (and not be true zeroes) is lower than 6 %. Similarly, in the last column, the percentile 99 implies that 99% of the estimated probabilities are below the value shown in the table. For the chemical, rubber, and plastics, we found that 99% of the probabilities of missing trade actually taking place were below 0.72 (72%).

Figure 1. Manufacturing Missing Exports: Counts vs. Volume



Notes: Figure 1 plots the ratio of the count of missing exports to all potential exports (horizontal axis), vs. the share of missing trade in total trade (vertical axis). The maximum count of exports that a given country is 1166 (i.e. 53 importers x 22 manufacturing products = 1166), then, for each country the horizontal axis shows, out of 1166, how many potential exports are unknown. In the vertical axis we assume that the estimated missing trade (using parameters in Table 4 and results in Table 6) is a reasonable estimate and add this to the observed trade obtaining a measure of total trade. The importance of missing trade on total trade is given by the ratio.

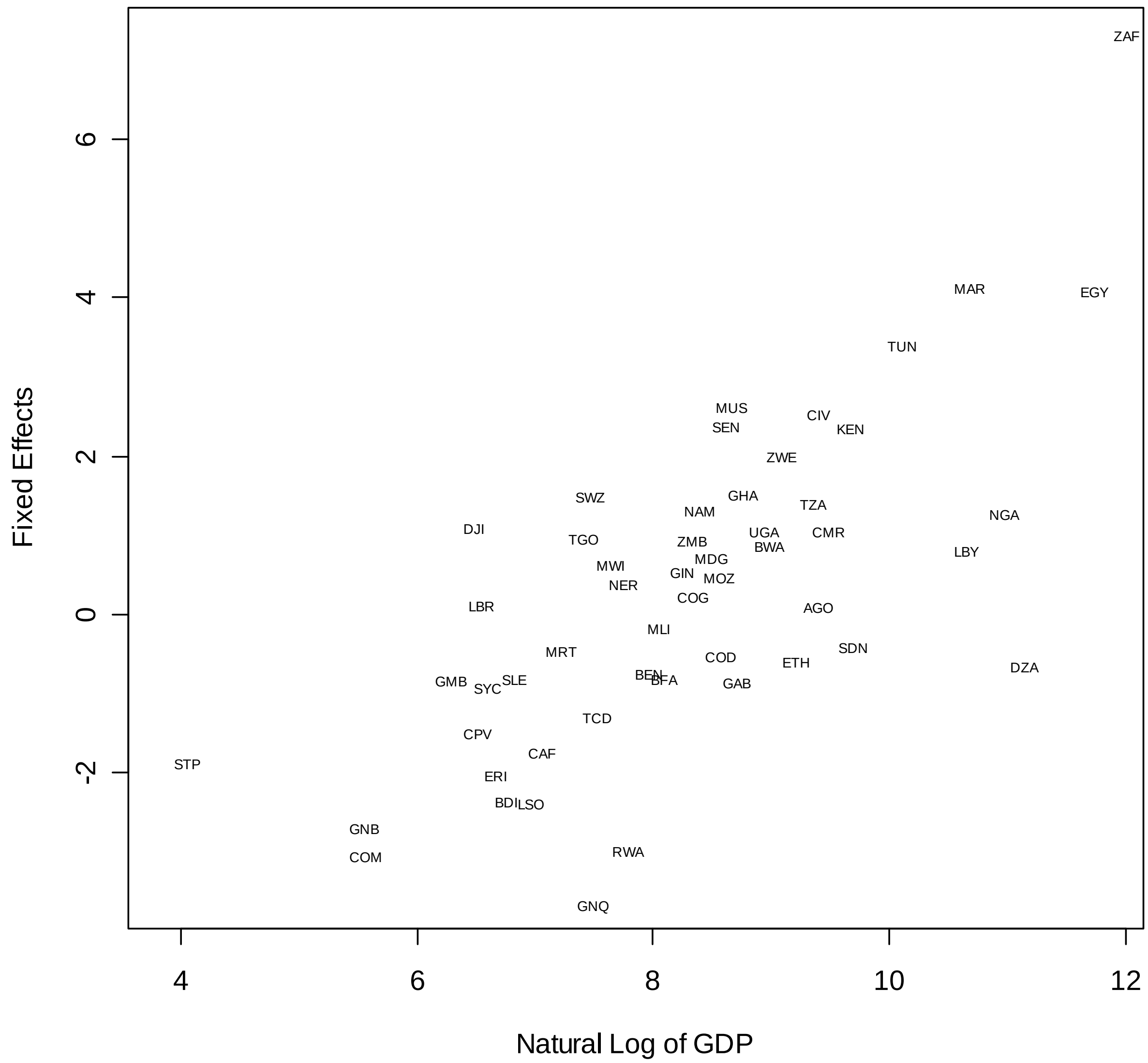
Figure 2. Country Size and Predicted Trade



Notes: Figure 2 plots the manufacturing exporter fixed effects (horizontal axis) against the predicted export values. The exporter fixed effects were estimated using the outcome equation (Equation 4). We argue (in Figure 3) that the fixed effects are a better surrogate for country size in the context of this work. The plot suggests a positive correlation between the size of the prediction and the size of the country.

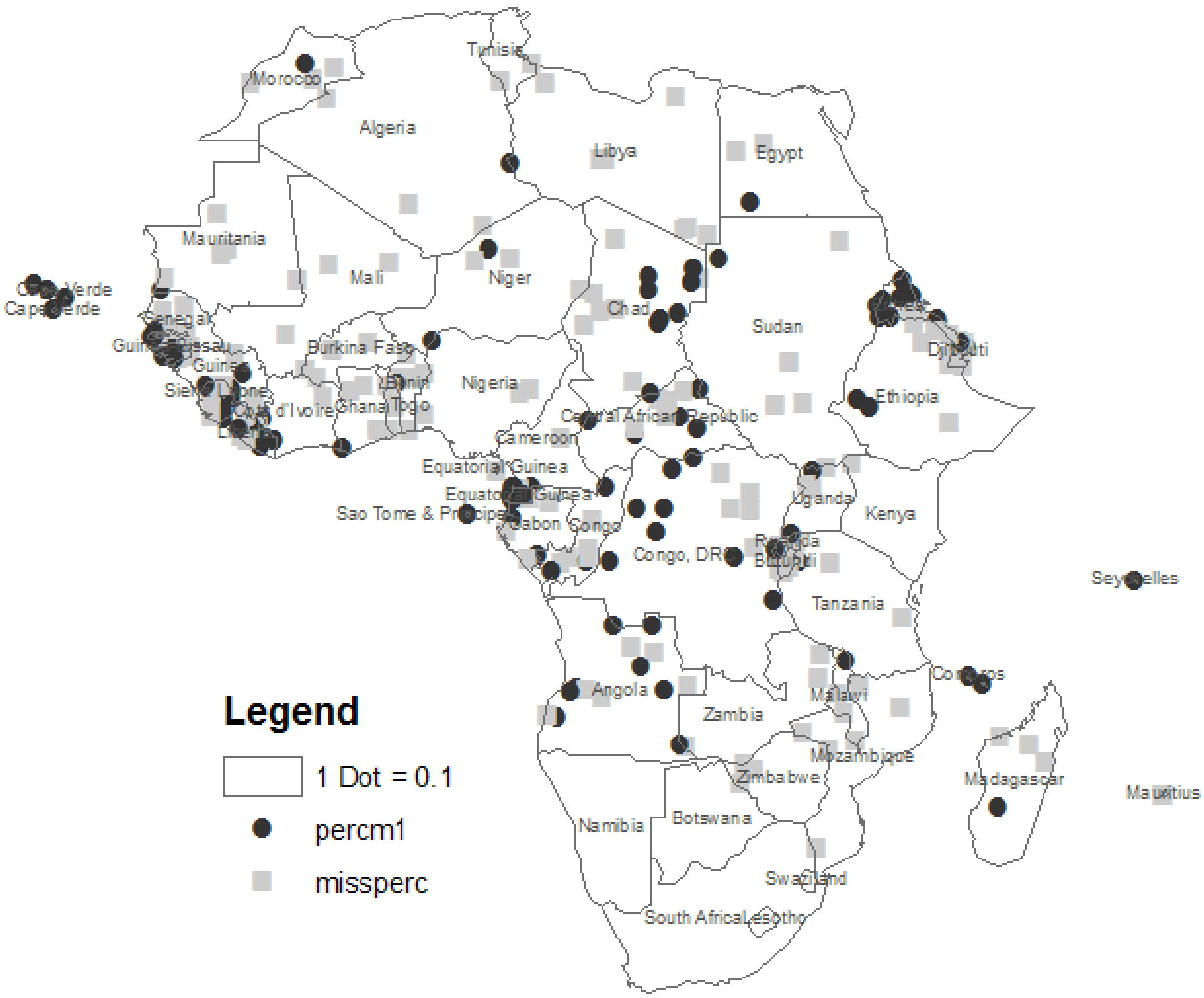
.

Figure 3: GDP and Exporter Fixed Effects



Notes: Figure 3 plots the natural log of GDP against the exporter fixed effects estimated using Equation 4 in the text. Fixed effects and country size are positively correlated. After all, country fixe effects are used to capture the effects of country size in trade. The fixed effects have the added advantage of including the effect of country size on trade, that is, some countries such as Sudan (SDN) and Kenya (KEN) can have the same size (as measured by GDP), but one is a larger exporter of Manufactures (Kenya) as revealed by the fixed effects.

Figure 4. Geographic distribution of missing exports.



Notes: This figure shows the geographic importance of missing exports. We use grey points for the ratio of the number of missing flows to potential exports and black points for the ratio of missing exports to total exports. Each point represents a tenth of the relevant ratio, or 10%. For example, in Madagascar we see three grey points and one black point. This means that 30% of the potential exports that Madagascar can send to her African partners are missing (this can be verified in Table 6, in the column "% Missing Exports"). The presence of only one black point implies that our estimated missing trade value for Madagascar is equivalent to 10% of its total trade (Table 6 confirms that the exact count of Madagascar's missing exports is 6%). Figure 2 suggests that the incidence of missing trade is higher in the countries of Central and West Africa, moderate in North Africa and unimportant in Southern Africa.

APPENDIX 1. Matrix of intra-African missing trade flows

Importer (columns) / Exporter (rows)

Exporter	AGO	BDI	BEN	BFA	BWA	CAF	CIV	CMR	COG	COM	CPV	DJI	DZA	EGY	ERI	ETH	GAB	GHA	GIN	GMB	GNB	GNQ	KEN	LBR	LBY	LSO	MAR
AGO		M				M	M		M	M		M		M	M		M	M			M	M		M	M		M
BDI	M		M	M		M	M	M	M	M	M	M		M	M	M		M	M	M	M	M	M	M	M		M
BEN		M								M	M	M			M	M				M	M		M	M	M		
BFA	M	M								M	M	M		M	M					M	M	M		M	M		M
BWA																											
CAF	M	M								M	M	M		M	M			M	M	M	M	M	M	M			M
CIV										M					M												
CMR		M								M					M						M						
COG	M	M								M	M	M		M	M			M		M	M	M	M	M	M		M
COM	M	M	M	M		M	M	M	M		M	M		M	M		M	M	M	M	M	M		M	M		M
CPV		M		M		M		M	M	M		M		M	M	M	M	M			M	M	M	M	M		M
DJI	M	M	M	M		M	M	M	M	M	M		M	M	M			M		M	M	M		M	M		M
DZA	M	M	M			M			M	M		M			M	M		M			M	M					
EGY	M	M							M	M	M	M			M			M			M	M		M	M		M
ERI	M	M	M	M		M	M	M	M	M	M	M		M		M	M	M	M	M	M	M		M	M		M
ETH	M	M		M		M		M	M	M					M				M	M	M	M		M	M		
GAB		M								M	M	M			M						M	M	M		M		
GHA	M	M				M			M	M	M	M		M	M						M	M		M	M		M
GIN		M				M				M		M			M						M	M			M		
GMB	M	M		M		M			M	M	M	M	M	M	M							M		M	M		M
GNB	M	M	M	M	M	M			M	M		M	M	M	M	M	M	M	M			M	M	M	M	M	M
GNQ	M	M	M			M			M	M	M	M	M	M	M	M	M	M	M		M		M	M	M		M
KEN									M		M										M						
LBR	M	M		M		M			M	M	M	M		M	M			M			M	M	M		M		M
LBY	M	M			M	M	M		M	M		M		M	M		M	M		M	M	M		M		M	M
LSO																											
MAR	M	M							M	M		M		M	M			M			M	M		M	M		
MDG	M	M				M									M			M			M	M			M		
MLI		M				M				M	M	M		M	M							M		M	M		
MOZ		M				M		M		M	M	M	M	M	M		M	M		M	M	M		M			M
MRT	M	M							M	M	M	M		M	M		M				M	M		M	M		M
MUS											M		M		M						M	M			M		
MWI		M				M				M	M	M			M						M	M			M		
NAM																											
NER						M			M	M		M		M	M						M	M		M			
NGA		M								M		M			M						M	M		M	M		M
RWA	M		M	M		M		M	M	M	M	M	M	M	M		M	M		M	M	M		M	M		M
SDN	M	M					M		M	M	M						M		M		M	M		M			
SEN		M								M		M			M												
SHN	M	M	M	M		M	M	M	M	M	M	M	M	M	M		M	M	M	M	M	M	M	M	M		M
SLE	M	M		M		M		M	M	M	M	M	M	M	M			M			M	M		M	M		M
SOM	M	M	M	M		M		M	M	M	M	M		M	M		M	M	M	M	M	M		M	M		M
STP		M		M		M			M	M	M	M	M	M	M	M		M	M	M	M			M	M		M
SWZ																											
SYC	M	M	M	M		M			M	M	M	M	M	M	M		M	M	M	M	M	M		M	M		M
TCD	M	M		M	M				M	M	M	M		M	M	M	M	M		M	M	M	M	M	M	M	M
TGO		M								M			M	M	M						M		M		M		
TUN		M																									
TZA				M		M					M						M				M	M		M	M		
UGA	M		M	M		M				M	M	M			M							M			M		M
ZAF																											
ZAR	M	M		M					M	M	M	M		M	M	M	M	M	M	M	M	M		M	M		M
ZMB				M		M	M			M		M	M		M		M				M	M		M			M
ZWE						M				M	M	M			M		M				M	M		M	M		
#	28	39	12	20	3	30	9	11	28	43	32	38	13	29	44	10	19	25	13	18	42	40	12	35	37	3	29

Exporter	MDG	MLI	MOZ	MRT	MUS	MWI	NAM	NER	NGA	RWA	SDN	SEN	SHN	SLE	SOM	STP	SWZ	SYC	TCD	TGO	TUN	TZA	UGA	ZAF	ZAR	ZMB	ZWE	#
AGO	M	M		M	M	M		M			M		M	M	M			M	M	M		M	M		M			31
BDI	M	M			M	M			M		M		M	M	M	M		M	M	M					M		M	37
BEN	M				M	M					M		M		M	M		M							M			20
BFA			M			M				M	M		M	M	M	M		M					M			M	M	25
BWA																												0
CAF		M	M			M		M		M			M	M	M	M		M				M	M			M	M	29
CIV											M		M		M													5
CMR			M								M		M		M													8
COG		M	M	M				M			M		M	M	M	M		M	M						M	M		28
COM		M	M	M		M				M	M	M	M	M	M	M		M	M	M		M	M		M	M	M	40
CPV		M	M		M	M		M		M	M		M	M	M			M	M	M					M	M	M	34
DJI	M	M	M		M	M		M		M			M	M	M	M		M	M	M	M	M	M		M		M	39
DZA			M			M				M			M	M	M	M		M	M			M			M	M		24
EGY													M	M	M	M			M						M			19
ERI	M	M	M	M	M	M		M	M	M	M	M	M	M	M	M		M	M	M					M		M	42
ETH	M			M									M	M		M		M	M						M			22
GAB			M								M		M		M			M					M			M		16
GHA	M				M								M	M	M				M						M			21
GIN			M										M		M	M									M		M	14
GMB	M		M										M		M	M		M	M						M			23
GNB		M	M	M	M	M	M	M		M	M		M	M	M	M	M	M	M	M		M	M	M	M		M	44
GNQ		M	M	M		M		M		M	M		M	M	M	M		M	M	M	M	M	M		M	M	M	40
KEN													M			M												5
LBR		M	M	M	M	M		M		M	M		M	M	M	M		M	M			M	M		M	M	M	35
LBY	M			M	M	M	M						M	M	M	M	M	M	M			M	M	M	M	M	M	36
LSO																												0
MAR										M			M	M	M	M		M	M						M			20
MDG				M									M	M	M	M			M				M			M		16
MLI			M											M	M	M		M										15
MOZ		M		M				M				M	M	M		M			M	M	M							26
MRT	M		M			M				M	M		M	M	M	M		M	M				M		M		M	28
MUS				M				M					M		M	M			M									12
MWI										M	M		M			M			M						M			15
NAM																												0
NER			M							M			M		M	M		M			M				M			17
NGA													M	M	M			M							M			14
RWA	M	M		M	M			M			M	M	M	M	M	M		M	M	M	M						M	36
SDN		M	M	M		M				M		M	M	M	M	M		M							M			23
SEN													M		M	M		M										8
SHN	M	M	M	M	M	M		M		M	M			M	M	M		M	M	M	M				M			41
SLE		M	M	M							M		M		M	M		M	M	M					M		M	30
SOM	M	M	M	M	M	M		M		M		M	M	M		M		M	M	M			M		M	M	M	40
STP	M	M	M	M	M	M		M		M	M	M	M	M	M			M	M	M	M				M	M		37
SWZ																												0
SYC		M	M	M		M		M		M		M	M	M	M	M			M	M					M			35
TCD			M	M		M	M					M	M	M	M	M	M	M		M		M		M	M	M	M	38
TGO			M		M					M			M		M			M										14
TUN			M										M		M			M							M		M	7
TZA																M			M	M								11
UGA				M									M		M				M									15
ZAF																												0
ZAR		M		M	M			M			M		M	M	M	M		M	M									30
ZMB													M		M	M			M		M							17
ZWE		M		M				M					M	M	M	M			M									18
#	14	21	27	23	16	21	3	18	2	20	21	9	46	32	43	38	3	34	33	17	8	10	13	3	31	14	18	

Source: BACI (See Gaulier, et al. 2007 p. 25)

Note: This matrix shows the exporters in the columns and the importers in the rows. An "M" indicates that the flow is missing. Empty cells are either zero-valued or postive trade flows as explained in the text (Section 2). Full country names have been ommited for reducing clutter, instead only ISO codes are shown. Please consult Table 1 for full country names. The final row and colum (labeled #) is the count of missing transactions, out of 53 potential partners.

APPENDIX 2. Variables employed in constructing gravity regressors

Country	Closest Partner		Remotest Partner		Neighbors	Language	Landlocked	Colonizer Country	Agreement	Religion	% Religion
	ISO	km	ISO	km							
Angola (AGO)	ZAR	553.11	MAR	5,221.76	3	Portuguese		Portugal	SADC	Catholic	38
Burundi (BDI)	RWA	180.01	CPV	6,165.72	3	French	x	Belgium	COMESA	Catholic	62
Benin (BEN)	NGA	105.18	MUS	6,701.09	4	French		France	WAEMU	Muslim	15
BurkiNA Faso (BFA)	NER	428.14	MUS	7,417.87	6	French	x	France	WAEMU	Protestant	
BotswaNA (BWA)	LSO	531.45	MAR	7,402.95	4	English	x	United Kingdom	SADC	Protestant	25
Central African Republic (CAF)	CMR	790.41	MUS	5,052.41	5	French	x	France	CAEMC	Protestant	25
Côte d'Ivoire (CIV)	GHA	418.15	MUS	7,296.39	5	French		France	WAEMU	Muslim	25
Cameroon (CMR)	GNQ	301.90	MUS	5,689.97	6	French		France	CAEMC	Protestant	17
Congo (COG)	ZAR	10.48	MUS	4,907.54	4	French		France	CAEMC	Protestant	25
Comoros (COM)	TZA	690.75	CPV	7,931.31	0	Arabic		France	COMESA	Muslim	86
Cape Verde (CPV)	SEN	651.84	MUS	9,677.82	0	Portuguese		Portugal	ECOWAS	Protestant	
Djibouti (DJI)	ETH	551.29	CPV	7,204.09	3	French		France	COMESA	Protestant	
Algeria (DZA)	TUN	642.72	MUS	8,539.50	6	Arabic		France		Muslim	99
Egypt (EGY)	SDN	1,619.16	ZAF	7,247.01	2	Arabic		United Kingdom		Muslim	94
Eritrea (ERI)	DJI	621.16	CPV	6,677.53	3	English		United Kingdom	COMESA	Protestant	
Ethiopia (ETH)	DJI	551.29	CPV	6,798.18	5	Amharic	x		COMESA	Muslim	42
Gabon (GAB)	STP	301.20	MUS	5,713.49	3	French		France	CAEMC	Protestant	33
GhaNA (GHA)	TGO	189.98	MUS	6,931.29	3	English		United Kingdom	ECOWAS	Muslim	30
Guinea (GIN)	GNB	261.65	MUS	8,665.62	6	French		France	ECOWAS	Muslim	85
Gambia (GMB)	SEN	155.95	MUS	8,933.76	1	English		United Kingdom	ECOWAS	Muslim	90
Guinea-Bissau (GNB)	GMB	208.57	MUS	8,762.87	2	Portuguese		Portugal	WAEMU	Muslim	30
Equatorial Guinea (GNQ)	CMR	301.90	MUS	5,947.82	2	Spanish		Spain	CAEMC	Protestant	
Kenya (KEN)	UGA	506.06	CPV	6,879.46	5	English		United Kingdom	COMESA	Protestant	38
Liberia (LBR)	SLE	367.82	MUS	8,024.15	3	English		France	COMESA	Muslim	20
Libyan Arab Jamahiriya (LBY)	TUN	527.67	MUS	7,545.62	6	Arabic		Turkey		Protestant	
Lesotho (LSO)	SWZ	489.15	MAR	7,921.95	1	English	x	United Kingdom	SADC	Protestant	40
Morocco (MAR)	DZA	945.49	MUS	9,096.82	1	Arabic		France		Muslim	99
Madagascar (MDG)	COM	921.11	CPV	8,638.64	0	French		France	COMESADC	Protestant	21
Mali (MLI)	BFA	687.36	MUS	8,044.20	7	French	x	France	WAEMU	Muslim	90
Mozambique (MOZ)	SWZ	150.50	MAR	7,869.63	6	Portuguese		Portugal	SADC	Protestant	15
Mauritania (MRT)	SEN	422.22	MUS	9,079.69	3	Arabic		France	ECOWAS	Muslim	100
Mauritius (MUS)	MDG	1,058.80	CPV	9,677.82	0	English		United Kingdom	COMESADC	Hinduism	52
Malawi (MWI)	ZWE	522.37	CPV	7,081.42	3	English	x	United Kingdom	COMESADC	Protestant	55
NAmibia (NAM)	BWA	930.86	MAR	6,790.84	4	English		Germany	SADC	Protestant	
Niger (NER)	BFA	428.14	MUS	7,130.48	7	French	x	France	WAEMU	Muslim	80
Nigeria (NGA)	BEN	105.18	MUS	6,610.81	4	English		United Kingdom	ECOWAS	Muslim	50
Rwanda (RWA)	BDI	180.01	CPV	6,189.04	4	French	x	Belgium	COMESA	Catholic	65
Sudan (SDN)	ERI	681.64	CPV	5,999.56	9	Arabic		United Kingdom	COMESA	Muslim	70
Senegal (SEN)	GMB	155.95	MUS	9,064.88	5	French		France	WAEMU	Muslim	92
Saint HeleNA (SHN)	AGO	2,204.00	SYC	6,792.15	0	English		United Kingdom	COMESA	Protestant	
Sierra Leone (SLE)	GIN	290.43	MUS	8,375.20	2	English		United Kingdom	ECOWAS	Muslim	30
Somalia (SOM)	KEN	1,017.10	CPV	7,685.12	3	Somali		United Kingdom	COMESA	Muslim	100
Sao Tome and Principe (STP)	GAB	301.20	MUS	5,974.60	0	Portuguese		Portugal	COMESA	Protestant	
Swaziland (SWZ)	MOZ	150.50	MAR	7,824.08	2	Swati	x	United Kingdom	SADC	Protestant	30
Seychelles (SYC)	SOM	1,347.70	CPV	8,969.38	0	English		United Kingdom	SADC	Catholic	90
Chad (TCD)	CAF	954.11	MUS	5,883.84	6	French	x	France	CAEMC	Muslim	44
Togo (TGO)	BEN	131.69	MUS	6,803.80	3	French		France	WAEMU	Protestant	10
Tunisia (TUN)	LBY	527.67	MUS	8,053.87	2	Arabic		France		Muslim	98
Tanzania, United Rep. of (TZA)	KEN	677.34	CPV	7,338.05	8	English		United Kingdom	SADC	Muslim	35
Uganda (UGA)	RWA	376.93	CPV	6,374.18	5	English	x	United Kingdom	COMESA	Protestant	33
South Africa (ZAF)	LSO	996.48	DZA	8,038.00	6	English		United Kingdom	SADC	Protestant	34
Congo (Democratic Republic of the) (ZAR)	COG	10.48	MUS	4,897.71	9	French		Belgium	COMESADC	Catholic	50
Zambia (ZMB)	ZWE	396.80	MAR	6,647.94	8	English	x	United Kingdom	COMESADC	Protestant	31
Zimbabwe (ZWE)	ZMB	396.80	MAR	7,032.50	4	English	x	United Kingdom	COMESADC	Protestant	25

Notes: Distances, colonizers, languages and number of neigbors are sourced from CEPII (see Mayer and Zignago, 2006). The variable religion comes from Sala-i-Martin (1997).

The column "closest partner" indicates the ISO code of the partner and the number of kilometers that separate the economic centers of both countries. Similarly, the column “remotest partner” shows which partner in the sample is located farthest away. The column "neighbors" is just a count of the number of different countries that a given nation shares borders with. Language is the official language (or the language spoken by more than 20% of the population). The Colonizer Country is shown next. The Agreement is the main regional economic agreement to which each country belongs. Finally, we have the dominant religion, with the percentage of people practicing that faith next to it.

.

Appendix 3. Predicted Intervals for intra-African Missing Trade

Country	iso	Lower Bound	Point-Estimate	Upper Bound	Country	iso	Lower Bound	Point-Estimate	Upper Bound
Angola	AGO	507	21,273	148,554	Mali	MLI	838	1,562	11,578
Burundi	BDI	361	1,393	10,256	Mozambique	MOZ	599	1,513	10,953
Benin	BEN	727	1,449	10,553	Mauritania	MRT	558	3,468	25,662
BurkiNA Faso	BFA	618	2,078	15,045	Mauritius	MUS	904	2,212	15,973
Central African Republic	CAF	534	1,375	10,155	Malawi	MWI	839	3,373	23,652
Côte d'Ivoire	CIV	1,063	1,323	10,045	Niger	NER	796	5,338	36,919
Cameroon	CMR	995	1,962	14,244	Nigeria	NGA	874	6,119	42,338
Congo	COG	572	10,896	76,091	Rwanda	RWA	374	494	3,826
Comoros	COM	288	625	4,855	Sudan	SDN	667	2,034	14,604
Cape Verde	CPV	421	909	7,208	Senegal	SEN	992	1,538	11,425
Djibouti	DJI	327	24,441	176,148	Saint HeleNA	SHN	282	6,961	50,851
Algeria	DZA	643	1,071	8,050	Sierra Leone	SLE	522	6,866	48,395
Egypt	EGY	900	29,486	199,943	Somalia	SOM	287	1,385	10,854
Eritrea	ERI	248	5,366	44,373	Sao Tome and Principe	STP	355	1,872	14,266
Ethiopia	ETH	685	1,954	14,154	Seychelles	SYC	399	1,721	12,481
Gabon	GAB	823	2,403	18,352	Chad	TCD	337	4,443	32,820
GhaNA	GHA	762	9,928	67,671	Togo	TGO	868	2,570	18,268
Guinea	GIN	859	2,083	17,608	Tunisia	TUN	1,022	12,278	83,165
Gambia	GMB	661	1,370	10,163	Tanzania	TZA	925	976	7,565
Guinea-Bissau	GNB	202	2,814	21,042	Uganda	UGA	840	3,594	25,308
Equatorial Guinea	GNQ	287	1,047	8,486	Congo (Democratic Republic of the)	ZAR	533	14,454	98,805
Kenya	KEN	1,060	4,350	30,217	Zambia	ZMB	793	1,527	11,239
Liberia	LBR	428	14,700	103,618	Zimbabwe	ZWE	773	3,656	25,674
Libyan Arab Jamahiriya	LBY	398	13,483	94,214	**TOTAL**		**31,431**	**297,066**	**2,091,874**
Morocco	MAR	866	46,864	316,323					
Madagascar	MDG	818	2,470	17,885					

Notes: For each country we show the lower and upper bounds of the predicted interval constructed around the point-estimates. The intervals were built using the individual's prediction's standard errors contained in the accompanying HAR file.