Online Appendix II:

Technical Documentation for the International Trade and Production Database for Simulation (ITPD-S)

Ingo Borchert* Mario Larch† Serge Shikher‡ Yoto V. Yotov§

December 2, 2024

This document provides the technical details of creating the International Trade and Production Database for Simulation (ITPD-S). It starts with a short motivation and background for the need for ITPD-S, followed by a description of the objectives and a summary of the applied methods. Sections 3 and 4 contain detailed descriptions of the simple and econometric methods, respectively, employed in the construction of the database, and Section 5 offers a summary of the estimation methods.

The Database is available on USITC's Gravity Portal
https://gravity.usitc.gov

*University of Sussex Business School, CITP, UKTPO.
†University of Bayreuth, CEPII, CESifo, WIFO, GEP, ifo.
‡United States International Trade Commission. The views expressed in this paper are strictly those of the authors and do not represent the opinions of the U.S. International Trade Commission or of any of its Commissioners.
§Drexel University, ifo, CESifo.

Contents

1 Motivation and Background

Modern, international quantitative trade policy analysis often requires ax-ante simulation analysis, which has demanding data needs. It requires availability and consistency across international and internal sales; across countries; across industries; and over time. Another issue is estimation of model parameters: to be consistent, it should be done using the same methodology and data as simulation. While new quantitative trade models are very capable, the availability of data that are suited to the needs of policy analysts and researchers is lacking. The datasets used for simulation are not suitable for estimation because they include estimated data. Existing datasets are also limited in terms of countries and industries.[1]

The recently developed International Trade and Production Database for Estimation (ITPD-E) is well suited for estimating key parameters of trade models, especially gravity models. This database is heavily used by researchers and policy analysts and is available on USITC's Gravity Portal, gravity.usitc.gov. It covers 265 countries, 170 industries, and in its second release over 30 years for most industries. Since it is solely based on reported, administrative data, it is suitable for estimation. However, ITPD-E is not suitable for simulation as it is highly unbalanced due to missing observations since some data points were not reported by national statistical agencies. Most importantly, ITPD-E has many missing observations for domestic trade.

The International Trade and Production Database for Simulation (ITPD-S) fills the needs of simulation analyses. Its data has the same level of detail as the ITPD-E, but with missing domestic trade observations filled in. This is done in steps relying on theory, and ensuring internal consistency. Hence, in combination, ITPD-E and ITPD-S provide researchers and policy analysts with mutually consistent databases for estimation and simulation.

[1] For example, GTAP data (https://www.gtap.agecon.purdue.edu/) and WIOD (https://www.rug.nl/ggdc/valuechain/wiod/?lang=en).

2 Summary of Methods

ITPD-S takes as a starting point the latest version of ITPD-E, which is ITPD-E-R02. It includes international and domestic trade data for all broad sectors (agriculture, mining and energy, manufacturing, and services), for 265 countries, and 170 industries. Following ITPD-E-R02, ITPD-S uses dynamic country codes that augment ISO3 country codes when the same ISO3 code is shared by multiple countries. The years covered are 1986-2019 for agriculture, 1988-2019 for mining and energy and manufacturing, and 2000-2019 for services.

ITPD-S fills in missing values for domestic trade observations. In order to do so, we use both simple and econometric methods:

1. **Simple methods:** Some data points can be filled in using linear interpolation, forward fill, and backward fill. Our analysis shows that these methods do a good job filling in missing values.

2. **Filling-in using a structural model:** Relying on the structural gravity model, we predict domestic trade flows. We only predict domestic trade flows because missing domestic trade flows are more prevalent than missing regular international trade flows due to a dearth of disaggregated gross output statistics, and because assuming that missing international trade flows are zero will be a plausible assumption in many cases.

3 Simple Methods

Taking advantage of the latest developments in the structural gravity literature, the ITPD-S construction is done in steps, as described below. Note that ITPD-E-R02 fills in missing values for international trade flows with zeros. It contains a flag, $flag\_zero$, which is equal to 'r' for observations with zeroes coming from original data sources, 'p' for observations with positive trade flows, and 'u' for observations filled with zeros. ITPD-E-R02 is still unbalanced because it only keeps observations that are used in a Poisson Pseudo-Maximum likelihood

(PPML) estimation with a demanding set of fixed effects (i.e., exporter-year, importer-year, and directional country-pair fixed effects). Specifically, this eliminates all zero observations not used for estimation because they are captured by fixed effects, e.g. if a country does not export in a given industry and year, then the corresponding zeros will be captured perfectly by this country's exporter-year fixed effect. For ITPD-S we start with the zeros coming from original data sources ($flag\_zero$ equal to 'r'), and with positive trade flow observations ($flag\_zero$ equal to 'p'), and mark them as actual, observed data. We flag them with a value of the flag variable flag_itpds equal to 1.

Keeping these observations, there are 265 exporters, 265 importers, 170 industries, and 34 years. We only keep the years for the broad industries that are also captured in ITPD-E-R02, i.e., 1986-2019 for agriculture, 1988-2019 for mining and energy and manufacturing, and 2000-2019 for services. In terms of countries, we fully balance at this stage. Hence, we end up with 352,256,688 observations ($(34 \times 28 + 32 \times 125 + 20 \times 17) \times (258 \times 258)$). Of those 352,256,688 observations, only 28,687,598 observations have trade values (of which 427,413 are zeros), that is, 8.1%. In terms of domestic sales, the fully balanced dataset has 1,365,336 observations ($(34 \times 28 + 32 \times 125 + 20 \times 17) \times 258$).

We set all missing international trade flows to zero. We believe that this is a plausible assumption, as reported international trade flows are quite comprehensive and it is well known that the international trade flow matrix is indeed sparse.

3.1 Assuming zero domestic trade in some cases

As a first step, we assume zero domestic sales when there are no yearly exports to any destination. If a country does not have any export within an industry over all years, and output data is missing, we set domestic sales to zero. This is not done in services industries.[2] This way, we are able to fill 314,316 domestic sales observations with zeroes. We flag these

[2]For services, it may very well be that output is positive even if we do not observe any international sales. Note that, using our prediction procedures below, it may very well be that we get fillings of domestic sales in cases where we do not observe any actual domestic sales. Such values should be used with great care. We flag all values that are filled so that they can easily be identified.

observations with flag_itpds=2.

3.2 Interpolation and forward and backward fill

Second, we fill in missing values by **interpolation**. Therefore, we use the time structure to fill in-between missing years by interpolation. With linear interpolation, we can add an additional 221,241 domestic trade observations. These observations are indicated with flag_itpds=3.

Besides interpolation, which fills in missing data between two years for which data are available, we also use **forward and backward fill**. Forward fill carries the value for the latest available year to more recent missing years while backward fill carries the value of the first available year to earlier missing years. Forward filled values are indicated by flag_itpds=4 while backward filled values are indicated by flag_itpds=5.

While we investigated linear and cubic extrapolation, we obtained a substantial share of negative values. Hence, we decided to not use extrapolated values to fill in missing.

4 Econometric Methods Using Gravity Theory

Next, we use **state-of-the-art structural gravity models** for each ITPD-E-R02 industry to predict missing domestic trade. To describe the theory-consistent filling up of data, we start with the standard gravity system (see Anderson and van Wincoop, 2003; Yotov et al., 2016, for examples):

$$X_{ij,t}^k = \frac{Y_{i,t}^k E_{j,t}^k}{Y_t^k} \left(\frac{t_{ij,t}^k}{\Pi_{i,t}^k P_{j,t}^k} \right)^{1-\sigma^k} \qquad \forall i, j, t, k; \tag{1}$$

$$\left(\Pi_{i,t}^k\right)^{1-\sigma^k} = \sum_{j=1}^{N} \left(\frac{t_{ij,t}^k}{P_{j,t}^k} \right)^{1-\sigma^k} \frac{E_{j,t}^k}{Y_t^k} \qquad \forall i, t, k; \tag{2}$$

$$\left(P_{j,t}^k\right)^{1-\sigma^k} = \sum_{i=1}^{N} \left(\frac{t_{ij,t}^k}{\Pi_{i,t}^k} \right)^{1-\sigma^k} \frac{Y_{i,t}^k}{Y_t^k} \qquad \forall j, t, k; \tag{3}$$

where $X^k_{ij,t}$ denote the value of shipments at destination prices from origin i to destination j in industry k at time t. $Y^k_{i,t}$ are total outputs in country i and industry k at time t, while Y^k_t is the world output of industry k at time t, i.e., $Y^k_t = \sum_{i=1}^N Y^k_{i,t}$. $E^k_{j,t}$ denote the total expenditures of country j on industry k from all origins at time t. $t^k_{ij,t}$ are iceberg trade costs on shipments of industry k from country i to country j at time t. σ^k denotes the elasticity of substitution across varieties from different origin countries, assuming product differentiation by place of origin (Armington, 1969). $\Pi^k_{i,t}$ and $P^k_{j,t}$ are the country-industry-time-specific outward and inward multilateral resistance terms, respectively (see Anderson and van Wincoop, 2003).

The trade flow equation (1) can be translated into an estimating equation as follows:

$$X^k_{ij,t} = \chi^k_t x^k_{i,t} m^k_{j,t} \tau^k_{ij,t} + \varepsilon^k_{ij,t}, \qquad \forall i, j, t, k, \tag{4}$$

where χ^k_t is an industry-time constant, $x^k_{i,t}$ collects all exporter-industry-year-specific terms which are captured by exporter-industry-year fixed effects, $m^k_{j,t}$ collects all terms on the importer side controlled for by importer-industry-year fixed effects, $\tau^k_{ij,t} \equiv t^{1-\sigma^k}_{ij,t} \leq 1$, and $\varepsilon^k_{ij,t}$ is an additive remainder error term with conditional expectation equal to zero. Hence, the conditional expectation of $X^k_{ij,t}$ is given by:

$$E[X^k_{ij,t}|.] = \chi^k_t x^k_{i,t} m^k_{j,t} \tau^k_{ij,t}. \tag{5}$$

This can be re-written as:

$$E[X^k_{ij,t}|.] = \exp\left(\pi^k_{i,t} + \chi^k_{j,t} + \ln\left(\tau^k_{ij,t}\right)\right). \tag{6}$$

This conditional expectation function can be consistently estimated with Poisson Pseudo-Maximum likelihood (PPML), as advocated by Santos Silva and Tenreyro (2006), where $x^k_{i,t}$ and χ^k_t are controlled for by exporter-industry-time fixed effects, $\pi^k_{i,t}$, and $m^k_{j,t}$ is controlled for by importer-industry-time fixed effects, $\chi^k_{j,t}$, respectively.

The fixed effects structure for international trade flows follows theory-grounded best practice as in equation (4) above. There is considerable flexibility in modeling the trade cost function $\tau^k_{ij,t}$. We follow two approaches: i) proxy for the trade costs using observables, ii) relying on the panel structure to proxy for the trade costs using bilateral fixed effects in combination with utilizing different aggregations and common fixed effects for these aggregates. We will describe each step in turn.

4.1 Proxy for Trade Costs Using Observables

We approximate trade costs by a rich set of 10 bilateral time-varying observables (in addition to the fixed effects structure and border effects):

$$\begin{aligned} Z_{ij,t} \;=\; & \Big\{ DIST_{ij};\ CNTG_{ij};\ LANG_{ij};\ CLEG_{ij};\ CREL_{ij};\ COLY_{ij}; \\ & EU_{ij,t};\ WTO_{ij,t};\ CSTU_{ij,t};\ PTA_{ij,t} \Big\}, \end{aligned} \tag{7}$$

where $DIST_{ij}$ denotes the log of bilateral distance, $CNTG_{ij}$ denotes contiguity between countries i and j, $LANG_{ij}$ denotes common language, $CLEG_{ij}$ common legal origin, $CREL_{ij}$ common religion, $COLY_{ij}$ a common colonial past, $EU_{ij,t}$ joint EU membership, $WTO_{ij,t}$ joint membership of the WTO, $CSTU_{ij,t}$ is an indicator variable denoting the existence of a customs union, and similarly, $PTA_{ij,t}$ whether or not countries i and j are signatories to any kind of preferential trade agreement at time t. Notice that some elements in this trade cost vector $Z_{ij,t}$ are time-varying whereas others are not. The data come partly from USITC's Dynamic Gravity Dataset and partly from CEPII.

These trade cost proxies for international trade costs are not directly relevant for projecting domestic trade, as the value for most of these (indicator) variables are set to zero for domestic trade flows. Yet, border effects for domestic trade, the modeling of which will be described in detail below, are identified relative to international trade and, as such, the accuracy of describing international trade costs has a beneficial indirect effect on border effects as well.

Domestic trade costs are usually captured by an indicator variable $BRDER_{ij}$, a dummy variable which is equal to one for international trade, i.e., whenever $i \neq j$, and zero else, i.e., whenever $i = j$. The definition of $BRDER_{ij}$ plays a key role in guiding our empirical analysis because it allows us to estimate (and hence predict) domestic trade flows in different, alternative ways. Accordingly, we use a range of options for specifying international border dummy variables (e.g. country- and time-specific or country-time-specific), see Anderson et al. (2018).

Taking all these considerations into account, trade costs are proxied as follows:

$$\ln\left(\tau_{ij,t}^{k}\right) = \beta^{k} Z_{ij,t} + DTC_{ii,t}, \tag{8}$$

where $Z_{ij,t}$ denotes the array of gravity variables as defined above with its corresponding parameter vector β^k. $DTC_{ii,t}$ is a proxy for domestic trade costs, which we parameterize in five different alternative ways. We explore the following increasingly flexible specifications for modeling domestic trade costs:

1. One common border effect for all countries at all points in time ($\alpha^k BRDER_{ij}$);

2. Time-varying border effects constructed by interaction terms of the border dummy and dummies for each year ($\alpha_t^k \left(BRDER_{ij} \times D_t\right)$). The associated coefficients α_t^k capture general globalization trends (see Bergstrand, Larch and Yotov, 2015);

3. Country-specific border effects constructed by interaction terms of border dummy and country dummies ($\alpha_i^k \left(BRDER_{ij} \times D_i\right)$). These variables capture heterogeneity across countries in the difference between international trade and domestic sales;

4. One common border effect that is allowed to vary with observable country characteristics, $W_{i,t}$, as defined below ($\alpha^k \left(BRDER_{ij} \times W_{i,t}\right)$);

5. A border effect interacted with observable country characteristics that are allowed to vary over time ($\alpha_t^k\left(BRDER_{ij} \times W_{i,t} \times D_t\right)$).

Let $W_{i,t}$ denote a vector of country characteristics that we use in the projection of domestic trade costs when domestic trade flows are missing for a particular country in a given year and industry. These country characteristics are assumed to have explanatory power for describing the variability of domestic trade costs across countries; for instance, the relative border barrier (BRDER) may be systematically higher in poorer or smaller economies. With this rationale in mind, we define $W_{i,t}$ as follows:

$$W_{i,t} = \Big\{ \ln DIST_i;\ CREL_i;\ \ln GDP_{i,t};\ \ln GDPPC_{i,t} \Big\}, \tag{9}$$

where $\ln DIST_i$ denotes the log of a country's internal (CES-weighted) distance, $CREL_i$ is a continuous variable denoting the degree of religious homogeneity within a country, $\ln GDP_{i,t}$ is the log of GDP used as a proxy for market size and therefore home bias, and $\ln GDPPC_{i,t}$ denotes the log of a country's GDP per capita in year t as a proxy for the stage of development.

Gravity equation specifications as described above deliver very high explanatory power for bilateral trade flows. Hence, we use it to fill in missing values by estimating equation (4) with five different alternative specifications for domestic trade costs $\mathbf{DTC}_{ii,t}$ in equation (8). Then we calculate projected trade flows as follows:

$$\begin{aligned}
\widehat{X}_{ij,t}^k &= \exp\left(\hat{\beta}^k Z_{ij,t} + \hat{\pi}_{i,t}^k + \hat{\chi}_{j,t}^k + \widehat{\alpha}^k BRDER_{ij}\right) && (10)\\
&= \exp\left(\hat{\beta}^k Z_{ij,t} + \hat{\pi}_{i,t}^k + \hat{\chi}_{j,t}^k + \widehat{\alpha}_t^k\left(BRDER_{ij} \times D_t\right)\right) && (11)\\
&= \exp\left(\hat{\beta}^k Z_{ij,t} + \hat{\pi}_{i,t}^k + \hat{\chi}_{j,t}^k + \widehat{\alpha}_i^k\left(BRDER_{ij} \times D_i\right)\right) && (12)\\
&= \exp\left(\hat{\beta}^k Z_{ij,t} + \hat{\pi}_{i,t}^k + \hat{\chi}_{j,t}^k + \widehat{\alpha}^k\left(BRDER_{ij} \times W_{i,t}\right)\right) && (13)\\
&= \exp\left(\hat{\beta}^k Z_{ij,t} + \hat{\pi}_{i,t}^k + \hat{\chi}_{j,t}^k + \widehat{\alpha}_t^k\left(BRDER_{ij} \times W_{i,t} \times D_t\right)\right), && (14)
\end{aligned}$$

where $\hat{\pi}_{i,t}^{k}$ and $\hat{\chi}_{j,t}^{k}$ are exporter-industry-time and importer-industry-time fixed effects controlling for $\ln\left(x_{i,t}^{k}\right)$ and $\ln\left(m_{j,t}^{k}\right)$, respectively, and hats denote estimates. When estimating equation (4) with PPML, the fixed effects have an exact structural interpretation if the true data-generating process follows the theoretical structure. Standard errors are clustered at the country-pair level.

Based on our results, it appears that the third model for domestic trade costs, i.e. that with country-specific domestic trade costs ($BRDER_{ij} \times D_i$), exhibits the highest degree of in-sample accuracy for predicting domestic sales when that information is available. This does not necessarily mean that this model is always superior, as we cannot really gauge the accuracy of other models out-of-sample because no actual domestic trade data are available for comparison. Thus it might well be that one of the other specifications is better suited to instances in which hardly any or no domestic sales are available.

Notice that this section's gravity-based approach to filling domestic sales breaks down when no domestic trade flow data at all are available in a given industry, as the coefficient on domestic trade costs $DTC_{ij,t}$ is then not identified. Our aggregation procedure described in Section 4.2 helps in these cases.

It also happens that for some country pairs or domestic sales in some industries, we cannot obtain trade flow predictions because some corresponding fixed effects cannot be estimated, or because some of the explanatory variables used to proxy trade costs are missing. In particular, time series for GDP and income per capita, although being fundamental macroeconomic variables, exhibit more gaps among developing countries than one might have anticipated.[3] Again, in these cases, our aggregation procedure described in Section 4.2 will help.

[3]For instance, we had to abandon the use of variables from the World Bank's Doing Business database, such as "time it takes to export" or "number of documents required for exporting", because of insufficient coverage, although such information would arguably be highly relevant for explaining variation in border effects across countries.

4.2 Using Panel Structure and Aggregation

The idea behind this approach is to obtain estimates of the key variables by exploiting the panel data at a bit more aggregated level (e.g., at the sectoral level). This approach is implemented in 3 steps, one for each of the 3 key variables. And, within each of the three steps, we take two sub-steps.

4.2.1 Predict Missing Domestic Trade Costs

The first step delivers estimates of some missing values for the domestic trade costs. It includes two sub-steps.

4.2.1.1 Estimate a Pooled Gravity Model at the Desired Level of Aggregation

First, we estimate the following gravity models at the desired level of aggregation:

$$X_{ij,t}^{k} = \exp[\alpha_t^k \left(BRDER_{ij} \times D_t\right) + \pi_{i,t}^k + \chi_{j,t}^k + \overrightarrow{\mu}_{ij}^k] \times \epsilon_{ij,t}^k, \forall i, j, \tag{15}$$

$$X_{ij,t}^{k} = \exp[\alpha_t^k \left(BRDER_{ij} \times D_t\right) + \pi_{i,t}^k + \chi_{j,t}^k + \overrightarrow{\mu}_{ij}^s] \times \epsilon_{ij,t}^k, \forall i, j. \tag{16}$$

The difference between equations (15) and (16) is the industry (superscript) index on asymmetric, bilateral fixed effects $\overrightarrow{\mu}_{ij}$. In equation (15) these fixed effects are for industry k. Hence, instead of using several proxies for trade costs, we exploit the panel structure and control for all asymmetric, bilateral, and non-time-varying influences by these fixed effects.

Equation (16) exploits different levels of aggregation. Hence, equation (16) is estimated by pooling several industries into an industry group s. The choice of which and how many industries to pool together is flexible. We want (i) many industries and (ii) industries that are as close as possible in terms of the type of products that they include. Since there are trade-offs, we start with a few close industries and then broaden the selection by moving to a more aggregated level. In sum, what this procedure does is impose a common industry group estimate for domestic trade costs within an otherwise most disaggregated structural gravity specification. Note that, as long as we have data for some of the industries in this industry group, we should be able to obtain a corresponding estimate for domestic trade costs in the industry group.

We utilize six different levels of aggregation. The lowest level is based on estimates for each of the 170 industries, corresponding to the industry description "ID 1" in Table 1. Table 1 provides the details for four more levels, which go up to "ID 5", where we aggregate to the four broad sectors "Agriculture", "Mining and Energy", "Manufacturing", and "Services". The sixth level of aggregation pools all 170 industries together.

4.2.1.2 Predict Missing Domestic Trade Costs from a Second-stage Regression

Second, we predict missing domestic trade costs from the following second-stage regression:

$$\widehat{\overrightarrow{\mu}}_i^k = \widehat{\overrightarrow{\mu}}_i^s + \psi_i + \phi^k + \epsilon_i^k, \tag{17}$$

where, $\widehat{\overrightarrow{\mu}}_i^k$ are the estimates of domestic trade costs from equation (15), $\widehat{\overrightarrow{\mu}}_i^s$ are the corresponding domestic trade cost estimates from equation (16), and ψ_i and ϕ^k are country and industry fixed effects, respectively, which will control for any country- and industry-specific characteristics.

In principle, we can add any additional country-industry covariates for which data are available. Note also that when we move to more aggregate analysis, we can add more such covariates. Finally, note that, instead of using the second-stage regression analysis, we can simply replace the missing industry-domestic trade costs with the corresponding sectoral estimates. This, of course, is even simpler. However, comparisons between the two approaches show that, as expected, the regression analysis delivers better results. Hence, this is the approach that we use for ITPD-S.

4.2.2 Predict Missing Exporter-time Fixed Effects

This step is very similar to Step 4.2.1. The data used is the same. However, instead of imposing a common sectoral estimate for the pair fixed effects, we are imposing a common sectoral estimate for the exporter-time fixed effects. As before, all other variables are at the most disaggregated level. The corresponding estimating equation becomes:

$$X_{ij,t}^{k} = \exp[\alpha_t^k \left(BRDER_{ij} \times D_t\right) + \pi_{i,t}^{s} + \chi_{j,t}^{k} + \overrightarrow{\mu}_{ij}^{k}] \times \epsilon_{ij,t}^{k}, \forall i, j. \quad (18)$$

Then, we use the resulting sectoral estimates in a second-stage regression, where the dependent variable is the exporter-product-time estimates from (15):

$$\widehat{\pi}_{i,t}^{k} = \widehat{\pi}_{i,t}^{S} + \psi_i + \phi^k + \gamma_t + \epsilon_{i,t}^{k}. \quad (19)$$

4.2.3 Predict Missing Importer-time Fixed Effects

This step is identical to Step 4.2.2. However, this time we implement it for the importer-time fixed effects. Armed with the predicted values of the 3 key variables, we can predict more missing domestic trade values.

4.2.4 Repeat Steps At Increasingly Aggregate Levels

We repeat the steps described in Sections 4.2.1, 4.2.2, and 4.2.3 at different levels of aggregation. Specifically, we apply this procedure for all the five aggregates described in Table 1, as well as for all the 170 industries pooled. Note that this includes the natural step to aggregate the broad sectoral levels, i.e., Agriculture, Manufacturing, Mining and Energy, and Services. The caveats of this approach are that with each step at a more aggregate level, we are moving further and further away from the industry level. Thus, the sectoral estimates of the three key variables are weaker and weaker predictors of the missing product-level variables that we need.

Table 1: Levels of Aggregation

ID 1	Description 1	ID 2	Description 2	ID 3	Description 3	ID 4	Description 4	ID 5	Description 5
1	Wheat	1	Cereals	1	Cereals	1	Cereals	1	Agriculture
2	Rice (raw)	1	Cereals	1	Cereals	1	Cereals	1	Agriculture
3	Corn	1	Cereals	1	Cereals	1	Cereals	1	Agriculture
4	Other cereals	1	Cereals	1	Cereals	1	Cereals	1	Agriculture
5	Cereal products	1	Cereals	1	Cereals	1	Cereals	1	Agriculture
6	Soybeans	2	Oilseeds	2	Oilseeds	2	Plant products	1	Agriculture
7	Other oilseeds (excluding peanuts)	2	Oilseeds	2	Oilseeds	2	Plant products	1	Agriculture
8	Animal feed ingredients and pet foods	3	Animal feed ingredients and pet foods	3	Animal products	3	Animal products	1	Agriculture
9	Raw and refined sugar and sugar crops	4	Sugars	4	Sugars	2	Plant products	1	Agriculture
10	Other sweeteners	4	Sugars	4	Sugars	2	Plant products	1	Agriculture
11	Pulses and legumes, dried, preserved	5	Plant products	5	Plant products	2	Plant products	1	Agriculture
12	Fresh fruit	5	Plant products	5	Plant products	2	Plant products	1	Agriculture
13	Fresh vegetables	5	Plant products	5	Plant products	2	Plant products	1	Agriculture
14	Prepared fruits and fruit juices	5	Plant products	5	Plant products	2	Plant products	1	Agriculture
15	Prepared vegetables	5	Plant products	5	Plant products	2	Plant products	1	Agriculture
16	Nuts	5	Plant products	5	Plant products	2	Plant products	1	Agriculture
17	Live Cattle	6	Meat	3	Animal products	3	Animal products	1	Agriculture
18	Live Swine	6	Meat	3	Animal products	3	Animal products	1	Agriculture
19	Eggs	7	Animal products	3	Animal products	3	Animal products	1	Agriculture
20	Other meats, livestock products, and live animals	6	Meat	3	Animal products	3	Animal products	1	Agriculture
21	Cocoa and cocoa products	5	Plant products	5	Plant products	2	Plant products	1	Agriculture
22	Beverages, nec	5	Plant products	5	Plant products	2	Plant products	1	Agriculture
23	Cotton	5	Plant products	5	Plant products	2	Plant products	1	Agriculture
24	Tobacco leaves and cigarettes	5	Plant products	5	Plant products	2	Plant products	1	Agriculture
25	Spices	5	Plant products	5	Plant products	2	Plant products	1	Agriculture
26	Other agricultural products, nec	5	Plant products	5	Plant products	2	Plant products	1	Agriculture
27	Forestry	5	Plant products	5	Plant products	2	Plant products	1	Agriculture
28	Fishing	7	Animal products	3	Animal products	3	Animal products	1	Agriculture
29	Mining of hard coal	1	Mining	1	Mining and Energy	1	Mining and Energy	2	Mining and Energy

Continued on next page

Continues from previous page

ID 1	Description 1	ID 2	Description 2	ID 3	Description 3	ID 4	Description 4	ID 5	Description 5
30	Mining of lignite	1	Mining	1	Mining and Energy	1	Mining and Energy	2	Mining and Energy
31	Extraction crude petroleum and natural gas	1	Mining	1	Mining and Energy	1	Mining and Energy	2	Mining and Energy
32	Mining of iron ores	1	Mining	1	Mining and Energy	1	Mining and Energy	2	Mining and Energy
33	Other mining and quarring	1	Mining	1	Mining and Energy	1	Mining and Energy	2	Mining and Energy
34	Electricity production, collection, and distribution	2	Eenegy	1	Mining and Energy	1	Mining and Energy	2	Mining and Energy
35	Gas production and distribution	2	Eenegy	1	Mining and Energy	1	Mining and Energy	2	Mining and Energy
36	Processing/preserving of meat	1	Food	1	Food, Drinks, and Tobacco	1	Food, Drinks, and Tobacco	3	Manufacturing
37	Processing/preserving of fish	1	Food	1	Food, Drinks, and Tobacco	1	Food, Drinks, and Tobacco	3	Manufacturing
38	Processing/preserving of fruit & vegetables	1	Food	1	Food, Drinks, and Tobacco	1	Food, Drinks, and Tobacco	3	Manufacturing
39	Vegetable and animal oils and fats	1	Food	1	Food, Drinks, and Tobacco	1	Food, Drinks, and Tobacco	3	Manufacturing
40	Dairy products	1	Food	1	Food, Drinks, and Tobacco	1	Food, Drinks, and Tobacco	3	Manufacturing
41	Grain mill products	1	Food	1	Food, Drinks, and Tobacco	1	Food, Drinks, and Tobacco	3	Manufacturing
42	Starches and starch products	1	Food	1	Food, Drinks, and Tobacco	1	Food, Drinks, and Tobacco	3	Manufacturing
43	Prepared animal feeds	1	Food	1	Food, Drinks, and Tobacco	1	Food, Drinks, and Tobacco	3	Manufacturing
44	Bakery products	1	Food	1	Food, Drinks, and Tobacco	1	Food, Drinks, and Tobacco	3	Manufacturing
45	Sugar	1	Food	1	Food, Drinks, and Tobacco	1	Food, Drinks, and Tobacco	3	Manufacturing
46	Cocoa chocolate and sugar confectionery	1	Food	1	Food, Drinks, and Tobacco	1	Food, Drinks, and Tobacco	3	Manufacturing
47	Macaroni noodles & similar products	1	Food	1	Food, Drinks, and Tobacco	1	Food, Drinks, and Tobacco	3	Manufacturing
48	Other food products n.e.c.	1	Food	1	Food, Drinks, and Tobacco	1	Food, Drinks, and Tobacco	3	Manufacturing
49	Distilling rectifying & blending of spirits	2	Drinks and Tobacco	1	Food, Drinks, and Tobacco	1	Food, Drinks, and Tobacco	3	Manufacturing
50	Wines	2	Drinks and Tobacco	1	Food, Drinks, and Tobacco	1	Food, Drinks, and Tobacco	3	Manufacturing
51	Malt liquors and malt	2	Drinks and Tobacco	1	Food, Drinks, and Tobacco	1	Food, Drinks, and Tobacco	3	Manufacturing
52	Soft drinks; mineral waters	2	Drinks and Tobacco	1	Food, Drinks, and Tobacco	1	Food, Drinks, and Tobacco	3	Manufacturing
53	Tobacco products	2	Drinks and Tobacco	1	Food, Drinks, and Tobacco	1	Food, Drinks, and Tobacco	3	Manufacturing
54	Textile fibre preparation; textile weaving	3	Textile	2	Textile, Apparel, and Footwear	2	Textile, Apparel, and Footwear	3	Manufacturing
55	Made-up textile articles except apparel	3	Textile	2	Textile, Apparel, and Footwear	2	Textile, Apparel, and Footwear	3	Manufacturing
56	Carpets and rugs	3	Textile	2	Textile, Apparel, and Footwear	2	Textile, Apparel, and Footwear	3	Manufacturing

Continued on next page

Continues from previous page

ID 1	Description 1	ID 2	Description 2	ID 3	Description 3	ID 4	Description 4	ID 5	Description 5
57	Cordage rope twine and netting	3	Textile	2	Textile, Apparel, and Footwear	2	Textile, Apparel, and Footwear	3	Manufacturing
58	Other textiles n.e.c.	3	Textile	2	Textile, Apparel, and Footwear	2	Textile, Apparel, and Footwear	3	Manufacturing
59	Knitted and crocheted fabrics and articles	4	Apparel and footwear	2	Textile, Apparel, and Footwear	2	Textile, Apparel, and Footwear	3	Manufacturing
60	Wearing apparel except fur apparel	4	Apparel and footwear	2	Textile, Apparel, and Footwear	2	Textile, Apparel, and Footwear	3	Manufacturing
61	Dressing & dyeing of fur; processing of fur	4	Apparel and footwear	2	Textile, Apparel, and Footwear	2	Textile, Apparel, and Footwear	3	Manufacturing
62	Tanning and dressing of leather	4	Apparel and footwear	2	Textile, Apparel, and Footwear	2	Textile, Apparel, and Footwear	3	Manufacturing
63	Luggage handbags etc.; saddlery & harness	4	Apparel and footwear	2	Textile, Apparel, and Footwear	2	Textile, Apparel, and Footwear	3	Manufacturing
64	Footwear	4	Apparel and footwear	2	Textile, Apparel, and Footwear	2	Textile, Apparel, and Footwear	3	Manufacturing
65	Sawmilling and planing of wood	5	Wood	3	Wood and Paper	3	Wood and Paper	3	Manufacturing
66	Veneer sheets plywood particle board etc.	5	Wood	3	Wood and Paper	3	Wood and Paper	3	Manufacturing
67	Builders' carpentry and joinery	5	Wood	3	Wood and Paper	3	Wood and Paper	3	Manufacturing
68	Wooden containers	5	Wood	3	Wood and Paper	3	Wood and Paper	3	Manufacturing
69	Other wood products; articles of cork/straw	5	Wood	3	Wood and Paper	3	Wood and Paper	3	Manufacturing
70	Pulp paper and paperboard	6	Paper	3	Wood and Paper	3	Wood and Paper	3	Manufacturing
71	Corrugated paper and paperboard	6	Paper	3	Wood and Paper	3	Wood and Paper	3	Manufacturing
72	Other articles of paper and paperboard	6	Paper	3	Wood and Paper	3	Wood and Paper	3	Manufacturing
73	Publishing of books and other publications	6	Paper	3	Wood and Paper	3	Wood and Paper	3	Manufacturing
74	Publishing of newspapers journals etc.	6	Paper	3	Wood and Paper	3	Wood and Paper	3	Manufacturing
75	Publishing of recorded media	6	Paper	3	Wood and Paper	3	Wood and Paper	3	Manufacturing
76	Other publishing	6	Paper	3	Wood and Paper	3	Wood and Paper	3	Manufacturing
77	Printing	6	Paper	3	Wood and Paper	3	Wood and Paper	3	Manufacturing
78	Service activities related to printing	6	Paper	3	Wood and Paper	3	Wood and Paper	3	Manufacturing
79	Coke oven products	7	Chemicals	4	Chemicals, Rubber, and Plastic products	4	Chemicals, Rubber, Plastic, Mineral and Metal products	3	Manufacturing

Continued on next page

Continues from previous page

ID 1	Description 1	ID 2	Description 2	ID 3	Description 3	ID 4	Description 4	ID 5	Description 5
80	Refined petroleum products	7	Chemicals	4	Chemicals, Rubber, and Plastic products	4	Chemicals, Rubber, Plastic, Mineral and Metal products	3	Manufacturing
81	Processing of nuclear fuel	7	Chemicals	4	Chemicals, Rubber, and Plastic products	4	Chemicals, Rubber, Plastic, Mineral and Metal products	3	Manufacturing
82	Basic chemicals except fertilizers	7	Chemicals	4	Chemicals, Rubber, and Plastic products	4	Chemicals, Rubber, Plastic, Mineral and Metal products	3	Manufacturing
83	Fertilizers and nitrogen compounds	7	Chemicals	4	Chemicals, Rubber, and Plastic products	4	Chemicals, Rubber, Plastic, Mineral and Metal products	3	Manufacturing
84	Plastics in primary forms; synthetic rubber	7	Chemicals	4	Chemicals, Rubber, and Plastic products	4	Chemicals, Rubber, Plastic, Mineral and Metal products	3	Manufacturing
85	Pesticides and other agrochemical products	7	Chemicals	4	Chemicals, Rubber, and Plastic products	4	Chemicals, Rubber, Plastic, Mineral and Metal products	3	Manufacturing
86	Paints varnishes printing ink and mastics	7	Chemicals	4	Chemicals, Rubber, and Plastic products	4	Chemicals, Rubber, Plastic, Mineral and Metal products	3	Manufacturing
87	Pharmaceuticals medicinal chemicals etc.	7	Chemicals	4	Chemicals, Rubber, and Plastic products	4	Chemicals, Rubber, Plastic, Mineral and Metal products	3	Manufacturing
88	Soap cleaning & cosmetic preparations	7	Chemicals	4	Chemicals, Rubber, and Plastic products	4	Chemicals, Rubber, Plastic, Mineral and Metal products	3	Manufacturing
89	Other chemical products n.e.c.	7	Chemicals	4	Chemicals, Rubber, and Plastic products	4	Chemicals, Rubber, Plastic, Mineral and Metal products	3	Manufacturing
90	Man-made fibres	8	Rubber and Plastic	4	Chemicals, Rubber, and Plastic products	4	Chemicals, Rubber, Plastic, Mineral and Metal products	3	Manufacturing
91	Rubber tyres and tubes	8	Rubber and Plastic	4	Chemicals, Rubber, and Plastic products	4	Chemicals, Rubber, Plastic, Mineral and Metal products	3	Manufacturing
92	Other rubber products	8	Rubber and Plastic	4	Chemicals, Rubber, and Plastic products	4	Chemicals, Rubber, Plastic, Mineral and Metal products	3	Manufacturing
93	Plastic products	8	Rubber and Plastic	4	Chemicals, Rubber, and Plastic products	4	Chemicals, Rubber, Plastic, Mineral and Metal products	3	Manufacturing
94	Glass and glass products	9	Mineral products	5	Mineral and Metal products	4	Chemicals, Rubber, Plastic, Mineral and Metal products	3	Manufacturing
95	Pottery china and earthenware	9	Mineral products	5	Mineral and Metal products	4	Chemicals, Rubber, Plastic, Mineral and Metal products	3	Manufacturing
96	Refractory ceramic products	9	Mineral products	5	Mineral and Metal products	4	Chemicals, Rubber, Plastic, Mineral and Metal products	3	Manufacturing
97	Struct.non-refractory clay; ceramic products	9	Mineral products	5	Mineral and Metal products	4	Chemicals, Rubber, Plastic, Mineral and Metal products	3	Manufacturing

Continued on next page

Continues from previous page

ID 1	Description 1	ID 2	Description 2	ID 3	Description 3	ID 4	Description 4	ID 5	Description 5
98	Cement lime and plaster	9	Mineral products	5	Mineral and Metal products	4	Chemicals, Rubber, Plastic, Mineral and Metal products	3	Manufacturing
99	Articles of concrete cement and plaster	9	Mineral products	5	Mineral and Metal products	4	Chemicals, Rubber, Plastic, Mineral and Metal products	3	Manufacturing
100	Cutting shaping & finishing of stone	9	Mineral products	5	Mineral and Metal products	4	Chemicals, Rubber, Plastic, Mineral and Metal products	3	Manufacturing
101	Other non-metallic mineral products n.e.c.	9	Mineral products	5	Mineral and Metal products	4	Chemicals, Rubber, Plastic, Mineral and Metal products	3	Manufacturing
102	Basic iron and steel	10	Metal products	5	Mineral and Metal products	4	Chemicals, Rubber, Plastic, Mineral and Metal products	3	Manufacturing
103	Basic precious and non-ferrous metals	10	Metal products	5	Mineral and Metal products	4	Chemicals, Rubber, Plastic, Mineral and Metal products	3	Manufacturing
104	Structural metal products	10	Metal products	5	Mineral and Metal products	4	Chemicals, Rubber, Plastic, Mineral and Metal products	3	Manufacturing
105	Tanks reservoirs and containers of metal	10	Metal products	5	Mineral and Metal products	4	Chemicals, Rubber, Plastic, Mineral and Metal products	3	Manufacturing
106	Steam generators	10	Metal products	5	Mineral and Metal products	4	Chemicals, Rubber, Plastic, Mineral and Metal products	3	Manufacturing
107	Cutlery hand tools and general hardware	10	Metal products	5	Mineral and Metal products	4	Chemicals, Rubber, Plastic, Mineral and Metal products	3	Manufacturing
108	Other fabricated metal products n.e.c.	10	Metal products	5	Mineral and Metal products	4	Chemicals, Rubber, Plastic, Mineral and Metal products	3	Manufacturing
109	Engines & turbines (not for transport equipment)	11	Machines General	6	Machinery	5	Machinery, Electronics, and Transportation	3	Manufacturing
110	Pumps compressors taps and valves	11	Machines General	6	Machinery	5	Machinery, Electronics, and Transportation	3	Manufacturing
111	Bearings gears gearing & driving elements	11	Machines General	6	Machinery	5	Machinery, Electronics, and Transportation	3	Manufacturing
112	Ovens furnaces and furnace burners	11	Machines General	6	Machinery	5	Machinery, Electronics, and Transportation	3	Manufacturing
113	Lifting and handling equipment	11	Machines General	6	Machinery	5	Machinery, Electronics, and Transportation	3	Manufacturing
114	Other general purpose machinery	11	Machines General	6	Machinery	5	Machinery, Electronics, and Transportation	3	Manufacturing
115	Agricultural and forestry machinery	12	Machines Specific	6	Machinery	5	Machinery, Electronics, and Transportation	3	Manufacturing

Continued on next page

Continues from previous page

ID 1	Description 1	ID 2	Description 2	ID 3	Description 3	ID 4	Description 4	ID 5	Description 5
116	Machine tools	12	Machines Specific	6	Machinery	5	Machinery, Electronics, and Transportation	3	Manufacturing
117	Machinery for metallurgy	12	Machines Specific	6	Machinery	5	Machinery, Electronics, and Transportation	3	Manufacturing
118	Machinery for mining & construction	12	Machines Specific	6	Machinery	5	Machinery, Electronics, and Transportation	3	Manufacturing
119	Food/beverage/tobacco processing machinery	12	Machines Specific	6	Machinery	5	Machinery, Electronics, and Transportation	3	Manufacturing
120	Machinery for textile apparel and leather	12	Machines Specific	6	Machinery	5	Machinery, Electronics, and Transportation	3	Manufacturing
121	Weapons and ammunition	12	Machines Specific	6	Machinery	5	Machinery, Electronics, and Transportation	3	Manufacturing
122	Other special purpose machinery	12	Machines Specific	6	Machinery	5	Machinery, Electronics, and Transportation	3	Manufacturing
123	Domestic appliances n.e.c.	13	Electronics	7	Electronics and Transportation	5	Machinery, Electronics, and Transportation	3	Manufacturing
124	Office accounting and computing machinery	13	Electronics	7	Electronics and Transportation	5	Machinery, Electronics, and Transportation	3	Manufacturing
125	Electric motors generators and transformers	13	Electronics	7	Electronics and Transportation	5	Machinery, Electronics, and Transportation	3	Manufacturing
126	Electricity distribution & control apparatus	13	Electronics	7	Electronics and Transportation	5	Machinery, Electronics, and Transportation	3	Manufacturing
127	Insulated wire and cable	13	Electronics	7	Electronics and Transportation	5	Machinery, Electronics, and Transportation	3	Manufacturing
128	Accumulators primary cells and batteries	13	Electronics	7	Electronics and Transportation	5	Machinery, Electronics, and Transportation	3	Manufacturing
129	Lighting equipment and electric lamps	13	Electronics	7	Electronics and Transportation	5	Machinery, Electronics, and Transportation	3	Manufacturing
130	Other electrical equipment n.e.c.	13	Electronics	7	Electronics and Transportation	5	Machinery, Electronics, and Transportation	3	Manufacturing
131	Electronic valves tubes etc.	13	Electronics	7	Electronics and Transportation	5	Machinery, Electronics, and Transportation	3	Manufacturing
132	TV/radio transmitters; line comm. apparatus	13	Electronics	7	Electronics and Transportation	5	Machinery, Electronics, and Transportation	3	Manufacturing
133	TV and radio receivers and associated goods	13	Electronics	7	Electronics and Transportation	5	Machinery, Electronics, and Transportation	3	Manufacturing

Continued on next page

Continues from previous page

ID 1	Description 1	ID 2	Description 2	ID 3	Description 3	ID 4	Description 4	ID 5	Description 5
134	Medical surgical and orthopaedic equipment	13	Electronics	7	Electronics and Transportation	5	Machinery, Electronics, and Transportation	3	Manufacturing
135	Measuring/testing/navigating appliances etc.	13	Electronics	7	Electronics and Transportation	5	Machinery, Electronics, and Transportation	3	Manufacturing
136	Optical instruments & photographic equipment	13	Electronics	7	Electronics and Transportation	5	Machinery, Electronics, and Transportation	3	Manufacturing
137	Watches and clocks	13	Electronics	7	Electronics and Transportation	5	Machinery, Electronics, and Transportation	3	Manufacturing
138	Motor vehicles	14	Transportation	7	Electronics and Transportation	5	Machinery, Electronics, and Transportation	3	Manufacturing
139	Automobile bodies trailers & semi-trailers	14	Transportation	7	Electronics and Transportation	5	Machinery, Electronics, and Transportation	3	Manufacturing
140	Parts/accessories for automobiles	14	Transportation	7	Electronics and Transportation	5	Machinery, Electronics, and Transportation	3	Manufacturing
141	Building and repairing of ships	14	Transportation	7	Electronics and Transportation	5	Machinery, Electronics, and Transportation	3	Manufacturing
142	Building/repairing of pleasure/sport. boats	14	Transportation	7	Electronics and Transportation	5	Machinery, Electronics, and Transportation	3	Manufacturing
143	Railway/tramway locomotives & rolling stock	14	Transportation	7	Electronics and Transportation	5	Machinery, Electronics, and Transportation	3	Manufacturing
144	Aircraft and spacecraft	14	Transportation	7	Electronics and Transportation	5	Machinery, Electronics, and Transportation	3	Manufacturing
145	Motorcycles	14	Transportation	7	Electronics and Transportation	5	Machinery, Electronics, and Transportation	3	Manufacturing
146	Bicycles and invalid carriages	14	Transportation	7	Electronics and Transportation	5	Machinery, Electronics, and Transportation	3	Manufacturing
147	Other transport equipment n.e.c.	14	Transportation	7	Electronics and Transportation	5	Machinery, Electronics, and Transportation	3	Manufacturing
148	Furniture	15	Other Manufacturing	8	Other Manufacturing	6	Other Manufacturing	3	Manufacturing
149	Jewellery and related articles	15	Other Manufacturing	8	Other Manufacturing	6	Other Manufacturing	3	Manufacturing
150	Musical instruments	15	Other Manufacturing	8	Other Manufacturing	6	Other Manufacturing	3	Manufacturing
151	Sports goods	15	Other Manufacturing	8	Other Manufacturing	6	Other Manufacturing	3	Manufacturing
152	Games and toys	15	Other Manufacturing	8	Other Manufacturing	6	Other Manufacturing	3	Manufacturing
153	Other manufacturing n.e.c.	15	Other Manufacturing	8	Other Manufacturing	6	Other Manufacturing	3	Manufacturing
154	Manufacturing services on physical inputs	1	Services	1	Services	1	Services	4	Services

Continued on next page

Continues from previous page

ID 1	Description 1	ID 2	Description 2	ID 3	Description 3	ID 4	Description 4	ID 5	Description 5
155	Maintenance and repair services n.i.e.	1	Services	1	Services	1	Services	4	Services
156	Transport	1	Services	1	Services	1	Services	4	Services
157	Travel	1	Services	1	Services	1	Services	4	Services
158	Construction	1	Services	1	Services	1	Services	4	Services
159	Insurance and pension services	1	Services	1	Services	1	Services	4	Services
160	Financial services	1	Services	1	Services	1	Services	4	Services
161	Charges for use of intellectual property	1	Services	1	Services	1	Services	4	Services
162	Telecom, computer, information services	1	Services	1	Services	1	Services	4	Services
163	Other business services	1	Services	1	Services	1	Services	4	Services
164	Heritage and recreational services	1	Services	1	Services	1	Services	4	Services
165	Health services	1	Services	1	Services	1	Services	4	Services
166	Education services	1	Services	1	Services	1	Services	4	Services
167	Government goods and services n.i.e.	1	Services	1	Services	1	Services	4	Services
168	Services not allocated	1	Services	1	Services	1	Services	4	Services
169	Trade-related services	1	Services	1	Services	1	Services	4	Services
170	Other personal services	1	Services	1	Services	1	Services	4	Services

5 Summary of the Estimation Procedure Methods and Flags

As explained earlier, the methods used to estimate missing domestic trade observations are divided into simple and econometric methods. Simple methods do not involve statistical estimation. The method used to obtain the value of domestic trade is denoted by a flag variable $flag\_itpds$. The values and this flag and the corresponding methods are listed below. Flag values 13-15 are explained in the subsequent sections.

1. Simple estimation methods.

    - Flag=1: Trade values from the data, not estimated.
    - Flag=2: Domestic trade flows are set to zero when there are no exports to any destination in the given industry and year.
    - Flag=3: Using data from step 2 as the starting point, domestic and international trade flows are estimated by interpolation.
    - Flag=4: Using data from step 3 as the starting point, domestic and international trade flows are estimated by forward fill up to the maximum of 7 years.[4]
    - Flag=5: Using data from step 4 as the starting point, domestic and international trade flows are estimated by backward fill up to the maximum of 7 years.[5]
    - Flag=13: Extends the final data by filling in the remaining missing observations by interpolation.
    - Flag=14: Extends the final data by filling in remaining missing observations by forward fill.
    - Flag=15: Extends the final data by filling in the remaining missing observations by backward fill.

[4] The results are the same whether this step is done using data from step 2 or step 3 as the starting point.
[5] See footnote 4.

2. Cross-sectional estimation methods.

    - Flag=21: Time-unvarying common border effect for all countries (model 1)
    - Flag=22: Time-varying common border effect for all countries (model 2)
    - Flag=23: Time-unvarying country-specific border effect (model 3)
    - Flag=24: Border effect proxied by country characteristics (model 4)
    - Flag=25: Border effect proxied by country characteristics interacted with year fixed effects (model 5)

3. Panel estimation methods.

    - Flag=31: Industry (level 1)
    - Flag=32: 26 industry groups (level 2)
    - Flag=33: 15 industry groups (level 3)
    - Flag=34: 11 industry groups (level 4)
    - Flag=35: 4 broad sectors (level 5)
    - Flag=36: All industries combined (level 6)

6 Comparison and Evaluation of Estimation Methods

To evaluate our methods for estimating missing domestic trade observations we randomly drop 10% of domestic trade observations in each industry. These dropped observations are then estimated together with all other missing trade observations using our estimation methodology described in the previous section. Once the dropped trade flows are estimated, we compare the estimated values to the values in the original data.[6]

[6]Dropped observations are not necessarily representative of all missing observations in ITPD-S because missing observations are not randomly distributed. They are concentrated in some years and some countries. Therefore, the mix of methods used to estimate dropped observations would be different from the mix of methods used to estimate all missing observations in ITPD-S. The procedure explained in this section makes it possible to evaluate and compare different methods used to estimate missing values in ITPD-S.

We do not use dropped trade values that were filled by simple estimation methods as data for statistical estimation. This means that all dropped trade values stay missing when statistical estimation is performed. This approach allows us to make apples-to-apples comparisons across various estimation methods that we use, simple and statistical.

We compare actual and estimated values of dropped domestic trade observations using several statistics. We calculate the mean absolute deviation and the mean deviation of estimated from actual trade. We also calculate the mean absolute log point deviation and the mean log point deviation of estimated from actual trade. Another statistic is the correlation between the estimated and actual trade values. We also show plots of predicted vs. actual trade values.

A total of 18,096 observations are dropped. Each dropped trade value can be estimated by multiple methods. Since simple methods in step 1 are used sequentially, only one simple method is used for each observation. Cross-sectional models in step 2 are all estimated at the same time, so multiple models in step 2 can produce estimates for the same missing value. Panel estimation methods (levels) in step 3 are used sequentially, so only one aggregation level is used to estimate each missing value.

The results from step 1 are shown in Table 2. The second column shows how many dropped observations were filled by each simple method in column 1. The majority were filled by interpolation, which has flag 3. The total of 17,953 out of 18,096 dropped observations could be filled by simple methods and 143 observations could not be filled by any simple method.

Table 2: Summary statistics for each simple method

flag	count	mean abs dev	mean dropped	mean estimated	mean dev	mean abs log points	mean log points	corr
2	841	284.2	284.2	0.0	-284.2	na	na	na
3	15,623	499.7	6,609.6	6,599.4	-10.1	27.9	3.8	0.998
4	1,181	720.2	8,548.1	8,341.3	-206.8	46.8	5.3	0.999
5	308	744.1	7,478.8	7,766.8	288	38.1	5.1	0.995

Columns 3-8 show various measures of quality of the estimates. The third column shows the mean absolute deviation of estimated from actual domestic trade observation for each of the simple estimation methods (in millions of dollars). This is our preferred measure of quality. The fourth column shows the mean actual dropped trade value while the fifth column shows the mean estimated dropped trade value. Comparing these two columns allows us to see whether the estimation methods overestimate or underestimate trade. The sixth column shows the mean deviation of estimated from dropped observations. The seventh and eighth columns show the mean absolute log point deviation and the mean log point deviation.[7]

The mean absolute deviation is the smallest for flag 2, which means that the actual domestic trade is small when estimated domestic trade is set to zero by assumption. This simple method underestimates trade by construction since it can never produce an estimate that is greater than the actual value. Interpolation, which has flag 3, has the next smallest mean absolute deviation. This method has a small mean deviation from estimated from actual trade values. Overall, interpolation produces good estimates.

Forward filling, which has flag 4 and produced 1,181 estimates, results in higher mean absolute deviation than interpolation. Not surprisingly, this method tends to underestimate trade since trade tends to grow over time. Backward filling, which has flag 5, tends to overestimate trade for the same reason. The mean absolute deviations for forward and backward filling are similar. The seventh column shows that in log point terms all simple methods produce estimates that are close to the actual trade. The last column of Table 2 shows that the correlations between the estimated and actual trade are very high for all simple methods.

Figure 1 shows graphs of estimated vs. actual trade values that were dropped for testing purposes. They also show a 45-degree line, where the ideal estimates would lie. Note that the graphs have different scales. The graphs show that predicted values are generally close to the 45-degree line.

Step 2 in our estimation procedure uses 5 different cross-sectional models to estimate

[7]Infinite values are ignored when calculating log point deviations.

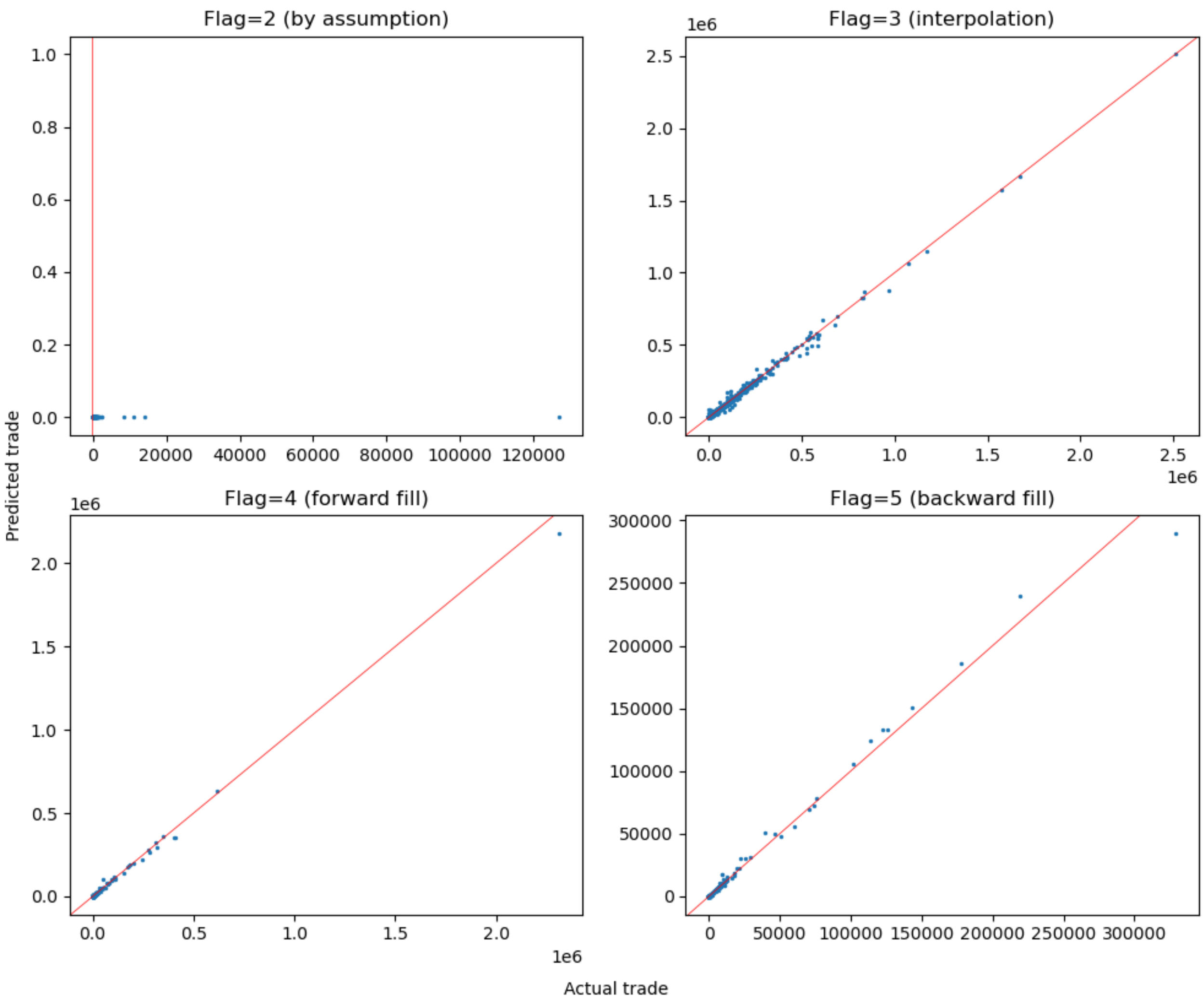

Figure 1: Predicted vs. actual domestic trade, simple methods

missing domestic trade. Table 3 below shows the number of models that can be used to fill dropped observations. There are 3,342 dropped trade observations that cannot be filled by step 2 methodology, which means that 14,754 dropped observations can be estimated by at least one model. There are 14,550 dropped observations that can be estimated by all 5 models.

Table 3: Number of models from step 2 that can be used to fill dropped observations

number of models	count of obs
0	3,342
2	6
3	104
4	94
5	14,550

Table 4 shows summary statistics for each model in step 2. The second column shows the number of dropped trade observations that can be estimated with each model. The third column shows the mean absolute deviation between the predicted and actual trade values. The deviations produced by the models in step 2 are larger than the deviations produced by simple methods in step 1, in both dollar and log point terms. In dollar terms, the models in step 2 on average overestimate the actual trade values. The correlations between the estimated and actual domestic trade vary between 0.37 and 0.75. Figure 2 shows the graphs of actual vs. predicted trade values for the five models used in step 2. Each graph also shows the 45-degree line where the ideal estimates would lie. Note that the graphs have different vertical scales. Both the table and the figure show that model 3 performs the best of all 5 models. It has the lowest deviation of estimated from actual values, in both dollar and log point terms.

Step 3 in our procedure uses panel estimation at one of 6 different industry aggregation levels to produce estimates of missing domestic trade. Only the lowest aggregation level for which an estimate is available is used to fill in dropped domestic trade. Column two in Table 5 shows the number of dropped domestic trade observations that were filled by each aggregation level in step 3. Of 18,096 dropped observation, 17,910 can be filled by

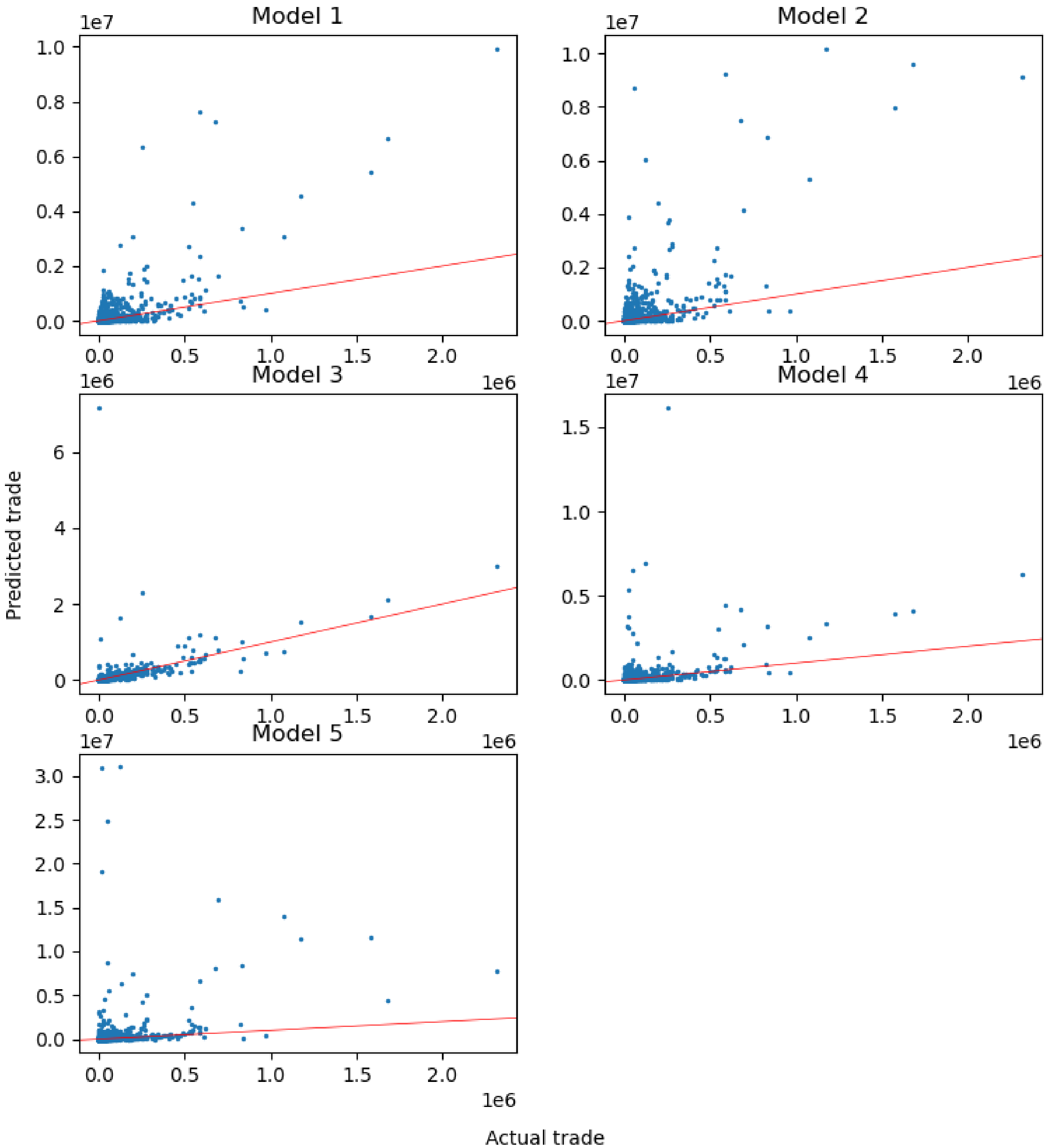

Figure 2: Predicted vs. actual domestic trade, step 2

Table 4: Summary statistics for each model in step 2

model	count	mean abs dev	mean dropped	mean estimated	mean dev	mean abs log points	mean log points	corr
1	14,754	12,197.8	6,703.7	15,092.6	8,389.0	213.6	-153.6	0.75
2	14,754	17,386.4	6,703.7	20,400.5	13,696.8	209.7	-142	0.708
3	14,654	3,688.2	6,747.8	7,896.2	1,148.4	95.6	-20.5	0.604
4	14,644	11,816.2	6,746.8	15,210.0	8,463.2	180.1	-91.2	0.501
5	14,644	28,785.1	6,746.8	32,052.9	25,306.1	192.9	-86.4	0.372

methodology of step 3. The vast majority of dropped observations can be filled using the first level of aggregation.

The mean absolute deviations for different levels of aggregation in step 3 are much lower than the mean absolute deviations for models 1, 2, 4, and 5 in step 2. The estimates at the levels of aggregation 1 and 5 on average overestimate actual trade in dollar terms while the estimates at levels of aggregation 2, 3, 4, and 6 on average underestimate actual trade in dollar terms. Mean log point deviations are the lowest for the first level of aggregation and then increase for higher levels of aggregation. The last column shows that the correlations for the first two levels of aggregation are about 0.6, but lower for the third and especially fourth and fifth levels of aggregation.

Table 5: Summary statistics by aggregation level in step 3

agg level	count	mean abs dev	mean dropped	mean estimated	mean dev	mean abs log points	mean log points	corr
1	16,737	3,806.5	6,701.8	8,265.7	1,563.9	88.3	-12.6	0.605
2	1,008	3,150.0	3,526.1	1,157.1	-2,369.0	272.1	-176.9	0.606
3	45	316.7	331.5	23.0	-308.5	350.3	-318.5	0.487
4	120	149.8	143.5	13.4	-130.1	468.8	-390.9	-0.046
5	96	174.7	55.3	131.1	75.8	516.6	-442.1	-0.039
6	64	1,429.9	1,437.0	7.1	-1,429.8	641.7	-641.7	0.707

Figure 3 shows the the graphs of actual vs. predicted trade values for the six levels of aggregations in step 3. Each graph also shows the 45-degree line where the ideal estimates would lie. Note that the graphs have different scales. The graphs show that the estimates produced in level 1 of step 3 are generally close to the 45 degree line while the estimates

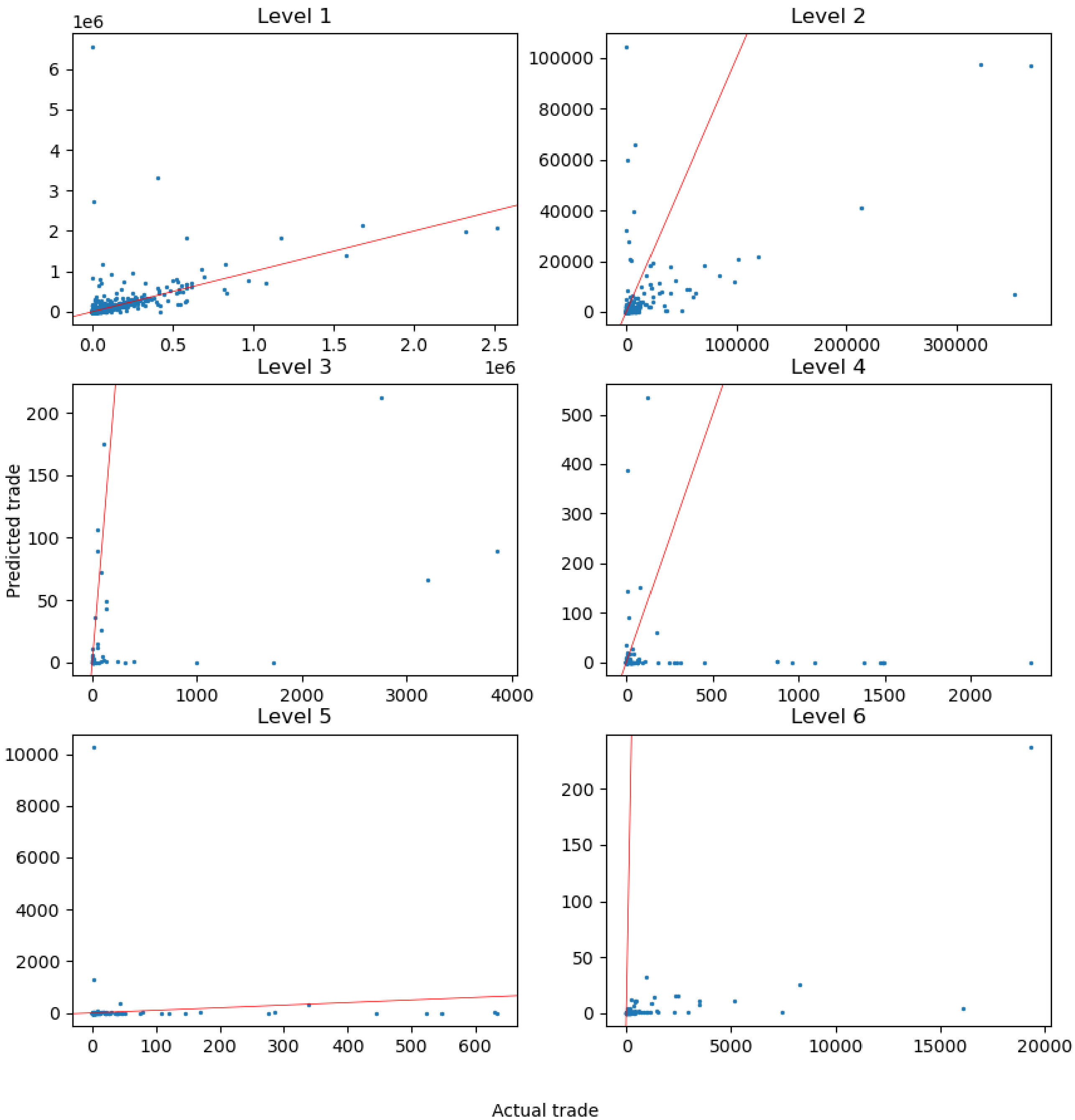

Figure 3: Predicted vs. actual domestic trade, step 3

from level 6 of step 3 are much lower than the actual data. The information in Table 5, especially the log point deviation, and Figure 3 suggests that level 1 aggregation in step 3 produces the most accurate results in step 3.

7 Procedure to Fill Missing Observations

The results in the previous section show that the simple methods, which have flags 2-5 produce the most accurate estimates of the dropped trade values in terms of all measures of quality. Therefore, these methods are used first to fill in missing trade observations. Then, we use the estimates obtained by model estimation, in the order suggested by the results of the previous section. Once the missing observations are filled in using model estimates, we check for outliers, which are defined as estimated observations that are greater than the maximum domestic trade value observed in the data (flag 1). This maximum value is observed in the United States. Outlier estimates are set to missing. Then, simple methods are applied again to fill in missing observations. Here is the summary of the sequence used to fill in missing observations:

1. Simple methods with flags 2-5
2. Panel method level 1, flag 31
3. Cross-sectional method model 3, flag 23
4. Panel method levels 2, 3, 4, 5, 6, flags 32-36
5. Cross-sectional methods 4, 1, 2, 5, flags 24, 21, 22, 25
6. Set outliers to missing
7. Simple methods with flags 13-15

8 Evaluation of Estimated Dropped Trade Values

All 18,096 dropped observations are estimated, vast majority by interpolation, which is marked by flag 3. Of all dropped observations, 143 observations are filled using estimates from panel methods and none using estimations from cross-sectional methods.

The results are shown in Table 6. The last row of the table shows the statistics for all dropped observations. The estimated trade values are on average 2.3 log points greater than the actual trade values. Model estimates on average fill smaller dropped trade observations than interpolation, backward fill, or forward fill (column 4). The smaller trade observations likely come from smaller countries, which are less likely to have enough domestic trade data to use simple methods.

Table 6: Summary statistics by method

flag	count	mean abs dev	mean dropped	mean estimated	mean dev	mean abs log points	mean log points	corr
2	841	284.2	284.2	0.0	-284.2	na	na	na
3	15,623	499.7	6,609.6	6,599.4	-10.1	27.9	3.8	0.998
4	1,181	720.2	8,548.1	8,341.3	-206.8	46.8	5.3	0.999
5	308	744.1	7,478.8	7,766.8	288	38.1	5.1	0.995
31	60	1,053.7	940.3	1,426.0	485.7	152.2	34.1	0.807
32	80	286.2	286.3	34.0	-252.2	399.3	-356.9	0.029
33	2	157.5	156.5	1.0	-155.5	636	-314.1	na
35	1	2.2	0.4	2.6	2.2	178.5	178.5	na
all	18,096	509.1	6,409.1	6,379.0	-30.1	31.6	2.3	0.998

9 Results of Filling in Missing Domestic Trade Observations

This section shows how all missing domestic trade observations in ITPD-S are estimated. The methods used to fill in missing domestic trade data in ITPD-S cannot estimate all missing observations. Some missing observations cannot be estimated because not enough information is available. However, there are 203 countries for which all missing observations are estimated.

Table 7 shows the results of filling in missing domestic trade observations in all countries, industries, and years of ITPD-E. There are 1,216,679 domestic trade observations in ITPD-S overall, as shown in column 1 of Table 7, of which 159,351 have data from ITPD-E and 1,057,328 are missing in ITPD-E. Of these missing observations, 1,037,821 are then estimated and 19,514 cannot be estimated. Simple methods with flags 2-5 provide 688,751 estimates, gravity models provide 331,160 estimates, and post-estimation simple methods with flags 13-15 provide another 17,903 estimates.

The 19,514 observations that cannot be estimated are in 62 countries. Table 8 shows the list of these countries. On the other hand, 203 countries have a complete set of domestic trade observations.

Table 7: Summary of all domestic trade observations

category or flag	count
all observations	1,216,679
missing	1,057,328
estimated	1,037,821
not estimated	19,514
1	159,351
2	314,316
3	221,241
4	138,130
5	15,064
31	22,920
23	29
32	100,907
33	20,622
34	9,460
35	34,906
36	128,337
24	7,294
21	6,685
22	0
25	0
13	1,009
14	16,733
15	161

Table 8: List of countries with missing observations

country code (dynamic)	country name
ABW	Aruba
AIA	Anguilla
AND	Andorra
ANT	Netherlands Antilles
ASM	American Samoa
ATA	Antarctica
ATF	French Southern Territories
BES	Bonaire, Sint Eustatius and Saba
BLM	Saint Barthelemy
BLX	Belgium-Luxembourg[a]
BVT	Bouvet Island
CCK	Cocos (Keeling) Islands
COM	Comoros
CSK	Czechoslovakia
CUW	Curacao
CXR	Christmas Island
DJI	Djibouti
ESH	Western Sahara
FLK	Falkland Islands
FRE	Free Zones
FRO	Faeroe Islands
FSM	Micronesia, Federated States of
GAZ	Gaza Strip
GIB	Gibraltar
GLP	Guadeloupe
GRL	Greenland
GUF	French Guiana
GUM	Guam
HMD	Heard Island and McDonald Islands

. . . continued

country code (dynamic)	country name
IMN	Isle of Man
IOT	British Indian Ocean Ter.
KIR	Kiribati
LIE	Liechtenstein
MCO	Monaco
MHL	Marshall Islands
MNP	Northern Marianas
MSR	Montserrat
MTQ	Martinique
MYT	Mayotte
NFK	Norfolk Island
NIU	Niue
NRU	Nauru
PCN	Pitcairn
PLW	Palau
PRI	Puerto Rico
PRK	Korea, North
REU	Reunion
SGS	South Georgia and South Sandwich Islands
SHN	Saint Helena, Ascension, and Tristan da Cunha
SLB	Solomon Islands
SOM	Somalia
SPM	Saint Pierre and Miquelon
SSD	South Sudan
SVU	Soviet Union
SXM	Sint Maarten
TCA	Turks and Caicos Islands
TKL	Tokelau
TUV	Tuvalu
UMI	U.S. Minor Outlying Islands

. . . continued

country code (dynamic)	country name
VAT	Vatican City[b]
VIR	U.S. Virgin Islands
WLF	Wallis and Futuna Islands

[a] Data for Belgium and Luxembourg exist under BEL and LUX

[b] Vatican City 1986-1995, Holy See 1996-present

<—

10 List of Variables

The variables included in ITPD-S are shown in Table 9. Most of the variables are carried over from ITPD-E. The only addition is $flag\_itpds$ which shows the provenance of domestic trade values.

Table 9: Variables in ITPD-S-R01 – Data File Columns

Column name	Column description
exporter_iso3	ISO 3-letter alpha code of the exporter
exporter_name	Name of the exporter
importer_iso3	ISO 3-letter alpha code of the importer
importer_name	Name of the importer
exporter_dynamic_code	Dynamic alpha code of the exporter based on DGD
importer_dynamic_code	Dynamic alpha code of the importer based on DGD
year	Year
industry_id	ITPD industry code
industry_descr	ITPD industry description
broad_sector	Broad sector description
trade	Trade flows in million of current US dollars
flag_mirror	Flag indicator, 1 if trade mirror value is used
flag_zero	Flag indicator: ‘p’ if positive trade ‘r’ if the raw data contained zero ‘u’ missing (unknown, assigned zero)
flag_itpds	Flag showing how domestic trade value was obtained

References

Anderson, James E., and Eric van Wincoop. 2003. "Gravity with Gravitas: A Solution to the Border Puzzle." *American Economic Review*, 93(1): 170–192.

Anderson, James E., Ingo Borchert, Aaditya Mattoo, and Yoto V. Yotov. 2018. "Dark Costs, Missing Data: Shedding Some Light on Services Trade." *European Economic Review*, 105: 193–214.

Armington, Paul S. 1969. "A Theory of Demand for Products Distinguished by Place of Production." *IMF Staff Papers*, 16: 159–176.

Bergstrand, Jeffrey H., Mario Larch, and Yoto V. Yotov. 2015. "Economic Integration Agreements, Border Effects, and Distance Elasticities in the Gravity Equation." *European Economic Review*, 78: 307–327.

Santos Silva, João M.C., and Silvana Tenreyro. 2006. "The Log of Gravity." *Review of Economics and Statistics*, 88(4): 641–658.

Yotov, Yoto V., Roberta Piermartini, José-Antonio Monteiro, and Mario Larch. 2016. *An Advanced Guide to Trade Policy Analysis: The Structural Gravity Model.* Geneva, Switzerland, available for download at https://unctad.org/publication/advanced-guide-trade-policy-analysis-structural-gravity-model-volume-2:United Nations and World Trade Organization.