ONLINE APPENDIX

THE DETERMINANTS OF INTRAFIRM TRADE: EVIDENCE FROM FRENCH FIRMS

The EIIG database

Intrafirm trade is defined in the primary EIIG data as trade with an affiliate controlled by a single French entity with at least fifty percent of its equity capital. The SESSI defines two types of trade with independent suppliers: 1) formal contractual relationships that refer to alliances, franchising, joint-ventures, and licensing agreements; 2) ‘informal’ relationships that involve less stringent contractual links. We consider both types of trade with independent suppliers as outsourcing.

In the primary EIIG data 20,952 out of the 81,217 import flows are ‘pure’ intrafirm (in the sense that 100% of imports of product \(p\) from country \(c\) come from a foreign affiliate), 50,021 are ‘pure’ outsourcing, and 10,244 are ‘mixed triples’. Out of the 10,244 mixed triples, 5,391 have 80% or more of the import value under a single sourcing mode (intrafirm or outsourcing). We choose to record these mixed triples according to the prevalent sourcing mode. Therefore, we end up with 76,364 triples from the EIIG database. As will be explained below, our final sample has 281,419 observations. Mixed triples represent 3.64% of the final sample, while the 4,853 mixed triples excluded from the analysis represent only 1.72% of all observations in the final sample. Further details on the EIIG database can be found in Guannel and Plateau (2003).

Construction of the estimation sample

In order to construct our estimation sample, we start by refining the population of interest. The EIIG survey was addressed to large traders, i.e. firms trading more than 1 million euros. There are 30,028 such firms in 1999 French Customs data, accounting for the bulk of imports (95.46%) in 1999. Out of these 30,028 large traders, 8,236 belong to the EIIG target population. We match the Customs and EIIG datasets under the assumption that import flows recorded in Customs data by firms other than the EIIG target population, occur with a third party. Put differently, we assume that SESSI successfully identified multinational firms among large traders.

Had all the 8,236 firms who received the EIIG questionnaire replied to the survey, a simple match of the EIIG data with imports by the remaining 21,792 non-multinational firms would provide us with full information on the population of large traders. However, about half of them (3,931) did not reply to the survey with these firms accounting for less than 20% of total exports and imports of the EIIG target population. Non-response to the EIIG survey thus seems to be
non-random with responding firms likely to be larger and possibly more productive than non-respondents.

To address potential biases, we construct a representative sample of the population of both multinational and non-multinational large importing French firms. To deal with sample selection due to non-response in the EIIG survey we use a two-stage Heckman procedure. In the first stage we estimate the probability that one of the importing firms in the EIIG target population responds to the survey, using firm total imports value, number of product categories imported, number origin countries involved, and NACE rev1 3-digit industry dummies. These variables reflect our presumption that a higher data collection effort was allocated by SESSI to large importers and/or certain sectors. This generates an inverse Mills ratio ($IM_1$) at the firm-level for all firms responding to the EIIG survey. We subsequently use $IM_1$, but none of the above first-stage variables, as an additional regressor throughout our firm-country-product analysis.

Finally, we construct a random sample of the population of non-multinational French large importers, i.e. importers that trade 1 million euros or more but do not belong to the EIIG target population. We do so by drawing a fraction of non-multinational importers that matches the response rate of the EIIG survey. By merging such a random sample with the EIIG data on respondent firms we get our final sample. In our analysis we will refer to this final sample as the ‘large sample’. It includes 281,419 observations spanning over 14,711 firms, 219 countries and 272 CPA96 4-digit products. Matching that sample with the EAE survey that documents manufacturing firm characteristics generates a ’small’ sample of 98,168 observations spanning over 5,175 firms, 185 countries and 270 products.

**Firm Variables**

**Productivity.** We estimate TFP as the residual (plus the constant) of a log-linearized three-factor Cobb-Douglas production function with labor, capital and material inputs. We use the value-added based Levinsohn and Petrin (2003) estimator (LP).

We estimate TFP based on the unbalanced EAE panel of 28,587 firms over 3 years (1998 to 2000) for a total of 74,120 observations. Observations with negative or missing values of value added, production, capital stock, material inputs and wages are eliminated. Outliers, identified as observations falling outside the 1st and 99th percentile of the distributions of value added per worker and capital stock per worker, are also not considered. This leaves us with TFP information on 22,673 firms for the core year 1999. TFP estimation has been carried out separately for each of the NACE 3-digit industries in the manufacturing sector.
Factor intensities. Our measure of capital intensity \( k_i \) is the log of the ratio between the capital stock and employment of firm \( i \). \( h_i \) is the log of the ratio between total wage expenses and employment of firm \( i \). This variable is meant to capture the average skills of workers of firm \( i \) with the underlying hypothesis being that more skilled workers are paid higher salaries.

Imported Product Variables

Contractibility. Our contractibility measure is based on the same idea as Nunn (2007): inputs sold on an organized exchange or reference priced are likely to be less relationship-specific, and therefore that sales contracts of these inputs are less incomplete. We also use the Rauch (1999) ‘Liberal’ product classification.

However we apply the relationship-specificity index to imported products directly. In Section 2 we contended that a proper test of Antrás and Helpman (2008) predictions would need to distinguish between contractibility of the inputs provided by the foreign supplier and those provided by the final producer. A second advantage of having product-level measures is that a final producer typically imports several products, with potentially different organizational decisions.

Our approach contrasts with Nunn (2007) and Nunn and Trefler (2008) who compute a weighted average of that index by final industry, using input-output matrix coefficients as weights. In their approach an emphasis on institutional comparative advantage makes it logical to measure how much exporting industries rely on complex inputs. Our approach focuses on organizational decisions by importers, so that a measure at the imported product level is more appropriate.

Denoting by \( R_{j}^{\text{neither}} \) a dummy variable that takes value 1 if the HS6 product \( j \) is neither sold on an organized exchange nor reference priced, and by \( \theta_{p,j} \) the share of the HS6 product \( j \) in the French imports of CPA96 4digit product \( p \) in 1999 we have:

\[
\mu_{p} = 1 - (\sum_{j} \theta_{p,j} R_{j}^{\text{neither}})
\]

The basic data needed to construct contractibility measures comes from Rauch (1999) and are organized on the basis of the SITC rev2 4 digit. Our import data are at the CPA96 4digit classification. To aggregate the Rauch data at the imported goods level, we proceed in two steps. First we establish (using data from the RAMON project) a correspondence between HS6 and SITC rev2 4 digit and a correspondence between HS6 and CPA96 4digit. Then we use import trade data in 1999 for France at the HS6 level (provided by EUROSTAT) as weights to aggregate the original SITC rev2 4 digit information to the CPA96 4digit.
Embodied capital and skill intensity. Embodied capital $k_p$ and skill intensity $h_p$ of the imported product $p$ are constructed using French technology. We introduce these variables because in Antràs (2003) factor intensities of the imported product play a key role. In the absence of cross-country product-level data on technology these variables should be seen as reasonable proxies.

To build $k_p$ and $h_p$, we start by using a correspondence table between the industry classification NACE rev1 4digit (available in our EAE firm dataset) and the product classification CPA96 4digit. We then compute the average capital intensity (log of capital/labor ratio) and skill intensity (log of total wage expenses/number of full time equivalent workers) of French firms associated to a given CPA96 4digit product.

Final Product Variables

Contractibility. Contractibility of a final good $f$ is measured in the same way as that of an imported product. Defining $R_{j \text{neither}}$ as dummy variable that takes value 1 if the PRODCOM2002 8 digit product $j$ is neither sold on an organized exchange nor reference priced, and by $\theta_{f,j}$ the share of the PRODCOM2002 8 digit product $j$ in the French production of CPA96 4digit product $f$ in 1999 we have:

$$\mu_f = 1 - \left( \sum_j \theta_{f,j} R_{j \text{neither}} \right)$$

Origin Country Variables

Key covariates. $k_c$ and $h_c$ stand (respectively) for the capital and skill abundance of country $c$. They are respectively measured by the log of the capital/labor and human capital/labor ratios provided by Hall and Jones (1999). $Q_c$ is a measure of the quality of institutions based on the “Rule of Law” index from Kaufmann, Kraay, and Mastruzzi (2003). This is a weighted average of a number of variables that measure individuals’ perceptions of the effectiveness and predictability of the judiciary and the enforcement of contracts in each country between 1997 and 1998.

Controls. $Tax_c$ is the top corporate tax rate prevailing in a given country in 1999 taken from the World Tax Database (University of Michigan). $\text{Fin} - \text{Dev}_c$ is a proxy for the degree of development of financial markets which we borrow from Beck (2002). $\text{OECD}_c$ is a dummy indicating membership to the OECD in 1999. $\text{Same} - \text{leg} - \text{orig}_c$ is a dummy indicating whether country $c$ has a French civil law system (Djankov et al., 2003). $\text{Dist}_c$ is the log of distance of country $c$ to France. $\text{Colony}_c$ is dummy indicating whether country $c$ is a former French colony and $\text{Language}_c$
is a dummy indicating whether French is spoken in country $c$. Data on $Distw_{c}$, $Colony_{c}$, and $Language_{c}$ come from CEPII (Centre d’Etude Prospectives et d’Informations Internationales).