Online Appendix for “Markets and Markup: A New Empirical Framework and Evidence on Exporters from China”

Giancarlo Corsetti
European University Institute and CEPR

Meredith Crowley
University of Cambridge and CEPR

Lu Han
Bank of Canada and CEPR

Huasheng Song
Zhejiang University

9 March 2023

OA1 The TPSFE Estimator

OA1.1 Key properties of the TPSFE estimator

As highlighted in section 6 of the paper, the fundamental reason for omitted variable and selection biases to arise is the missing information on key variables. Once the variation of these missing variables is properly controlled for, both omitted variable and selection biases will disappear. In large customs databases with four panel dimensions (i.e., firm, product, destination and time), fixed effects provide a natural tool to control for unobserved confounding variables.

However, due to endogenous market decisions of firms, correctly controlling for the desired variation of the unobserved variables that vary along multiple panel dimensions is a non-trivial task. The key difficulty is to design partition matrices that can account for the unbalanced panel structure and correctly eliminate the variation of unobserved confounding variables. The most relevant reference to our TPSFE demeaning procedure is Wansbeek and Kapteyn (1989), who consider an unbalanced panel with two panel dimensions and two fixed effects.

The econometrics contribution of our TPSFE estimator is to (a) improve the partition matrices proposed by Wansbeek and Kapteyn (1989), (b) generalize it into a four-dimension unbalanced panel and (c) apply the method to the estimation of markup elasticities in a large customs database. In particular for (c), thanks to the simplicity and transparency of our method, our TPSFE approach makes it easy to understand the underlying variation that is used to identify the markup elasticity to exchange rates. The approach points to the relevance of including trade patterns of firms' products to controlling for unobserved confounding variables.

Proposition 1. *In an unbalanced panel, our proposed TPSFE procedure eliminates all confounding variables that vary along the $fidD + fit$ panel dimensions.*

We start by introducing Proposition 1, which states that our TPSFE procedure can address all omitted variable and selection biases that are driven by variables varying along the $fidD+fit$ panel dimensions. For example, the unobserved marginal cost of a firm's product varies along fit panel dimension, while the differences in time-invariant demand conditions across markets facing a firm's product vary along fid panel dimension. The additional D in $fidD$ further allows for unobserved firm-product-destination-specific factors that co-move with the trade patterns of the firm-product. For example, a change in economic fundamentals $\mathcal{F}_t$ that has firm-product-destination specific effects and influences the set of destination markets of the firm-product will result in variation along the $fidD$ panel dimension, which can be controlled by our proposed estimator.

We proceed as follows. Subsections OA1.1.1 to OA1.1.3 discuss the key idea and mechanism behind our estimator and compare it to the partition matrices proposed by Wansbeek and Kapteyn (1989) in a two-dimensional panel. Subsection OA1.1.4 provides a numerical example to clarify our notation and discussions. Subsection OA1.1.5 generalizes the results to four-dimensional unbalanced panels.

OA1.1.1 Identifying the markup elasticity in a two-dimensional unbalanced panel

In this subsection, we discuss the identification of the markup elasticity in a two-dimensional unbalanced panel and introduce two useful lemmas that lay the foundation for the proof of Proposition 1. The idea is that identifying the markup elasticity and controlling for the unobserved confounding variables in a large customs database with four panel dimensions can be thought of as a collection of many smaller firm-product level problems that each have two panel dimensions, i.e., destination (d) and time (t). In those more refined two-dimensional problems, Lemma 1 shows the original partition methods of Wansbeek and Kapteyn (1989) can be decomposed into a two-step procedure with the second step implicitly applying a trade pattern related partition.

Lemma 1. *In a two-dimensional unbalanced panel, factors varying along the d+t panel dimensions can be eliminated using a two-step procedure by which, in the first step, all variables are demeaned across observed destinations within each period and, in the second step, destination (d) and trade pattern (D) fixed effects are applied additively, i.e., $d + D$.*

Building on the insights of Lemma 1, Lemma 2 shows a better estimator can be constructed to deal with more complicated cases, where the unobserved confounding variables vary along the $dD+t$ panel dimensions. The key idea is that, in the second step of the procedure, we can combine the d and D fixed effects interactively instead of additively.

Lemma 2. *In a two-dimensional unbalanced panel, factors varying along the $dD + t$ dimensions can be eliminated in a two-step procedure in which all variables are demeaned across observed destinations within each period in the first stage and destination (d) and trade pattern (D) fixed effects are applied multiplicatively, i.e., dD, in the second stage. This procedure also eliminates all confounding factors that the $d + t$ fixed effects can address.*

OA1.1.2 Proof of Lemma 1

The proof proceeds with two steps. In the first step, we construct a demeaned fixed effect estimator following Wansbeek and Kapteyn (1989). In the second step, we show that the constructed estimator implicitly applies trade pattern fixed effects.

Step 1: Let n_t^D $\left(n_t^D \leq n^D\right)$ be the number of observed destinations for year t. Let $n^{DT} \equiv \sum_t n_t^D$. Let A_t be the $\left(n_t^D \times n^D\right)$ matrix obtained from the $(n^D \times n^D)$ identity matrix from which the rows corresponding to the destinations not observed in year t have been omitted, and consider

$$Z \equiv \left(\begin{array}{cc} Z_1, & Z_2 \\ n^{DT} \times n^D & n^{DT} \times n^T \end{array} \right) \equiv \begin{bmatrix} A_1 & A_1 \iota_{n^D} & & \\ \vdots & & \ddots & \\ A_{n^T} & & & A_{n^T} \iota_{n^D} \end{bmatrix} \tag{OA1-1}$$

where ι_x is a vector of ones with length x, e.g., ι_{n^D} is a vector of ones with length n^D. The matrix Z gives the dummy-variable structure for the incomplete-data model. (For complete data, $Z_1 = \iota_{n^T} \otimes I_{n^D}, Z_2 = I_{n^T} \otimes \iota_{n^D}$.) Define

$$P_2 \equiv I_{n^{DT}} - Z_2 \left(Z_2' Z_2\right)^{-1} Z_2'$$
$$\bar{Z} \equiv P_2 Z_1.$$

Wansbeek and Kapteyn (1989) show P is a projection matrix onto the null-space of Z:

$$P \equiv P_2 - \bar{Z}(\bar{Z}'\bar{Z})^{-}\bar{Z}'$$

where '–' stands for a generalized inverse. It follows that, in an unbalanced panel with unobserved confounding variables varying along d and t panel dimensions, unbiased and consistent estimates can be obtained by running an OLS regression with the demeaned data obtained by pre-multiplying the data matrix (Y, X) by the projection matrix P.

Step 2: We now show the projection matrix P can be decomposed into two projection matrices with the second projection matrix applying destination and trade pattern fixed effects in additive terms. We begin by noting that the following relationship holds:

$$P \equiv P_2 - \bar{Z}(\bar{Z}'\bar{Z})^{-}\bar{Z}' = (I_{n^{DT}} - \bar{Z}(\bar{Z}'\bar{Z})^{-}\bar{Z}')P_2 \equiv P_1 P_2 \tag{OA1-2}$$

where $P_1 \equiv I_{n^{DT}} - \bar{Z}(\bar{Z}'\bar{Z})^{-}\bar{Z}'$ and the equality of (OA1-2) uses the fact that P_2 is idempotent (i.e., $P_2 Z_1 = P_2 P_2 Z_1 = P_2 \bar{Z}$). Therefore, applying the projection matrix P to the data matrix (Y, X) is equivalent to first pre-multiplying (Y, X) by the projection matrix P_2, and then pre-multiplying $(P_2 Y, P_2 X)$ by the projection matrix P_1. The projection P_2 applied in the first step is essentially a destination-demean process (the same first step as our TPSFE estimator).[1] The projection P_1 applied in the second step is, by definition, a "demeaning" process at the $\bar{Z}$ level. To see the exact dummy structure based on which the second "demeaning" process is applied, note that $\bar{Z}$ can be rewritten as

$$\bar{Z} = P_2 Z_1 = Z_1 - Z_2 \left(Z_2' Z_2\right)^{-1} Z_2' Z_1 \tag{OA1-3}$$

[1] See the numerical example in subsection OA1.1.4.

where Z_1 is a set of destination dummies as defined in (OA1-1) and $Z_2\left(Z_2'Z_2\right)^{-1}Z_2'Z_1$ is a set of trade pattern dummies.

To see that $Z_2\left(Z_2'Z_2\right)^{-1}Z_2'Z_1$ follows a trade pattern structure, note that $Z_2\left(Z_2'Z_2\right)^{-1}Z_2'$ is a block diagonal matrix with its diagonal blocks equal to a matrix of ones multiplied by (the inverse of) the number of destinations in each period, i.e.,

$$\begin{aligned}Z_2\left(Z_2'Z_2\right)^{-1}Z_2' &= \text{diag}\left(\frac{1}{n_1^D}A_1\iota_{n^D}\iota_{n^D}'A_1', ..., \frac{1}{n_{n^T}^D}A_{n^D}\iota_{n^D}\iota_{n^D}'A_{n^D}'\right)\\ &= \text{diag}\left(\frac{1}{n_1^D}\iota_{n_1^D}\iota_{n_1^D}', ..., \frac{1}{n_{n^T}^D}\iota_{n_{n^T}^D}\iota_{n_{n^T}^D}'\right)\end{aligned} \tag{OA1-4}$$

where the first equality holds by the definition of Z_2 in (OA1-1) and given the fact that $\left(Z_2'Z_2\right)^{-1}$ is a diagonal matrix, with its elements indicating (the inverse of) the number of observed destinations in each period, i.e.,

$$\left(Z_2'Z_2\right)^{-1} = \text{diag}\left(\frac{1}{n_1^D}, \frac{1}{n_2^D}, ..., \frac{1}{n_{n^T}^D}\right); \tag{OA1-5}$$

the second equality in (OA1-3) holds by the definition of the A matrices in (OA1-1). Pre-multiplying Z_1 by $Z_2\left(Z_2'Z_2\right)^{-1}Z_2'$ and using the definition of Z_1, we have

$$Z_2\left(Z_2'Z_2\right)^{-1}Z_2'Z_1 = \begin{bmatrix}\frac{1}{n_1^D}\iota_{n_1^D}\iota_{n_1^D}'A_1\\ \vdots\\ \frac{1}{n_{n^T}^D}\iota_{n_{n^T}^D}\iota_{n_{n^T}^D}'A_{n^D}\end{bmatrix} \tag{OA1-6}$$

where $\iota_{n_t^D}'A_t$ gives the trade pattern in year t and pre-multiplying it by $\iota_{n_t^D}$ repeats the same trade pattern n_t^D times—resulting in the trade pattern matrix for all destinations in period t.[2]

Therefore, the second "demeaning" projection matrix $P_1 \equiv I_{n^{DT}} - \bar{Z}(\bar{Z}'\bar{Z})^{-}\bar{Z}'$ is applied on $\bar{Z}$ that consists of two *additive* parts: (a) the destination dummies Z_1 and (b) the trade pattern dummies $Z_2\left(Z_2'Z_2\right)^{-1}Z_2'Z_1$.

OA1.1.3 Proof of Lemma 2

A key difference between our proposed TPSFE estimator and a conventional fixed effect estimator adding destination and time fixed effects lies in the way the trade patterns are applied in the second step. While the conventional approach applies the destination and trade pattern fixed effects additively (as can be seen from (OA1-3) and (OA1-6)), our estimator applies the trade pattern fixed effect multiplicatively.

[2]See Appendix OA1.1.4 for an numerical example of the matrices.

We start our proof by introducing notation and definitions. Denote the set of exporting destinations in year t as D_t.[3] Let $\mathcal{TP}$ be the set of unique trade patterns in all years, i.e.,

$$\mathcal{TP} \equiv \{D_1, ..., D_{n^T}\}_{\neq} \tag{OA1-7}$$

and $n^{\mathcal{TP}} \equiv |\mathcal{TP}|$ be the number of unique trade patterns. Let $\mathcal{TP}_x$ denote the x'th element of $\mathcal{TP}$. We create destination-specific trade patterns by combining the destinations in a trade pattern with the trade pattern itself, i.e., $\{(d, \mathcal{TP}_x) : d \in \mathcal{TP}_x\}$. Let $\mathcal{DTP}$ be the set of destination-specific trade patterns, i.e.,

$$\mathcal{DTP} \equiv \{(d, \mathcal{TP}_1) : d \in \mathcal{TP}_1, ..., (d, \mathcal{TP}_{n^{\mathcal{TP}}}) : d \in \mathcal{TP}_{n^{\mathcal{TP}}}\}.$$

Let $n^{\mathcal{DTP}} \equiv |\mathcal{DTP}|$ be the number of unique destination-trade pattern pairs observed in the data.

The dummy structure of destination-specific trade patterns is given by the following $(n^{DT} \times n^{\mathcal{DTP}})$ matrix:

$$Z_3 \equiv \begin{bmatrix} B_1 \\ \vdots \\ B_{n^T} \end{bmatrix} \equiv \begin{bmatrix} K_{11} & \cdots & K_{1n^{\mathcal{TP}}} \\ \vdots & \ddots & \vdots \\ K_{n^T 1} & \cdots & K_{n^T n^{\mathcal{TP}}} \end{bmatrix} \tag{OA1-8}$$

where B_t is an $n_t^D \times n^{\mathcal{DTP}}$ matrix indicating the destination-specific trade patterns in period t. Each B_t can be decomposed into $n^{\mathcal{TP}}$ block matrices with its y'th block being equal to an identity matrix if the trade pattern of period t, D_t, is the same as the y'th trade pattern, $\mathcal{TP}_y$, and a matrix of zeros otherwise. That is, $\forall x \in \{1, ..., n^T\}, y \in \{1, ..., n^{\mathcal{TP}}\}$,

$$K_{xy} \equiv \begin{cases} I_{n_x^D} & \text{if } D_x = \mathcal{TP}_y \\ \mathbf{0}_{n_x^D \times n_{\mathcal{TP}}^D(y)} & \text{if } D_x \neq \mathcal{TP}_y \end{cases} \tag{OA1-9}$$

where $I_{n_x^D}$ is an identity matrix of size n_x^D; $\mathbf{0}_{n_x^D \times n_{\mathcal{TP}}^D(y)}$ is a matrix of zeros of size $n_x^D \times n_{\mathcal{TP}}^D(y)$; and $n_{\mathcal{TP}}^D(y) \equiv |\{d : d \in \mathcal{TP}_y\}|$ is the number of destinations in the y'th unique trade pattern $\mathcal{TP}_y$.

Let the projection matrix be $P_3 P_2$, where $P_3 \equiv I_{n^{DT}} - Z_3 \left(Z_3' Z_3\right)^{-1} Z_3'$. The first projection P_2 is the same destination-demean process, whereas the second projection P_3 applies demeaning at the destination-trade pattern level. As discussed in previous sections, the interactive construction of trade pattern fixed effects enables us to handle interactive error terms and reduce the time variation of the unobserved confounding variables.

[3] In a vector form, $\iota'_{n_t^D} A_t$ indicates the set of destinations in year t.

To formally prove Lemma 2, we need to show that

$$P_3 P_2 Z_1 = \mathbf{0},$$
$$P_3 P_2 Z_2 = \mathbf{0},$$
$$P_3 P_2 Z_3 = \mathbf{0}.$$

We begin by noting that the second relationship holds by definition (of P_2):

$$P_3 P_2 Z_2 = [I_{n^{DT}} - Z_3 \left(Z_3' Z_3\right)^{-1} Z_3'][I_{n^{DT}} - Z_2 \left(Z_2' Z_2\right)^{-1} Z_2']Z_2 = \mathbf{0}.$$

We prove $P_3 P_2 Z_1 = \mathbf{0}$ and $P_3 P_2 Z_3 = \mathbf{0}$ by relying on two relationships that we state here and prove later in the text. First, the two projection matrices $T_3 \equiv Z_3 \left(Z_3' Z_3\right)^{-1} Z_3'$ and $T_2 \equiv Z_2 \left(Z_2' Z_2\right)^{-1} Z_2'$ commute:

$$T_3 T_2 = T_2 T_3. \tag{OA1-10}$$

Second, T_3 projects Z_1 to itself:

$$T_3 Z_1 = Z_1. \tag{OA1-11}$$

Given (OA1-10) and (OA1-11), it follows that

$$\begin{aligned} P_3 P_2 Z_1 &= [I_{n^{DT}} - T_3][I_{n^{DT}} - T_2]Z_1 \\ &= Z_1 - T_3 Z_1 + T_3 T_2 Z_1 - T_2 Z_1 \\ &= T_3 T_2 Z_1 - T_2 Z_1 \\ &= T_2 T_3 Z_1 - T_2 Z_1 \\ &= T_2 Z_1 - T_2 Z_1 \\ &= \mathbf{0} \end{aligned}$$

where the second equality is due to (OA1-11); the third equality holds due to the commutativity (OA1-10); the fourth equality applies (OA1-11) one more time. Following the same procedure, it can be shown that $P_3 P_2 Z_3 = \mathbf{0}$.

We complete our proofs showing that (OA1-10) and (OA1-11) hold.

Proof of (OA1-10):

Proof. We want to prove that the two projection matrices $Z_3 \left(Z_3' Z_3\right)^{-1} Z_3'$ and $Z_2 \left(Z_2' Z_2\right)^{-1} Z_2'$ commute. We do so by proving that the product of these two matrices $Z_3 \left(Z_3' Z_3\right)^{-1} Z_3' Z_2 \left(Z_2' Z_2\right)^{-1} Z_2'$

is symmetric.

$Z_3\left(Z_3'Z_3\right)^{-1}Z_3'$ can be written as:

$$Z_3\left(Z_3'Z_3\right)^{-1}Z_3' = \begin{bmatrix} B_1\left(Z_3'Z_3\right)^{-1}B_1' & \cdots & B_1\left(Z_3'Z_3\right)^{-1}B_{n^T}' \\ \vdots & \ddots & \vdots \\ B_1\left(Z_3'Z_3\right)^{-1}B_{n^T}' & \cdots & B_{n^T}\left(Z_3'Z_3\right)^{-1}B_{n^T}' \end{bmatrix} \tag{OA1-12}$$

The blocks of $Z_3\left(Z_3'Z_3\right)^{-1}Z_3'$ can be further simplified using the following two observations. First, $\left(Z_3'Z_3\right)^{-1}$ is an $n^{\mathcal{DTP}} \times n^{\mathcal{DTP}}$ diagonal matrix with its elements indicating (the reverse of) the number of repetitions for each destination-trade pattern pair, i.e.,

$$\begin{aligned}
\left(Z_3'Z_3\right)^{-1} &= \left(\sum_t B_t'B_t\right)^{-1} \\
&= \begin{bmatrix} \sum_t K_{t1}'K_{t1} & \cdots & \sum_t K_{t1}'K_{tn^{\mathcal{TP}}} \\ \vdots & \ddots & \vdots \\ \sum_t K_{tn^T}'K_{t1} & \cdots & \sum_t K_{tn^{\mathcal{TP}}}'K_{tn^{\mathcal{TP}}} \end{bmatrix}^{-1} \\
&= \begin{bmatrix} r_1^{\mathcal{TP}} I_{n^D_{\mathcal{TP}(1)}} & & \\ & \ddots & \\ & & r_{n^{\mathcal{TP}}}^{\mathcal{TP}} I_{n^D_{\mathcal{TP}(n^{\mathcal{TP}})}} \end{bmatrix}^{-1} \\
&= \text{diag}\left(\frac{1}{r_1^{\mathcal{TP}}} I_{n^D_{\mathcal{TP}(1)}}, ..., \frac{1}{r_{n^{\mathcal{TP}}}^{\mathcal{TP}}} I_{n^D_{\mathcal{TP}(n^{\mathcal{TP}})}}\right)
\end{aligned} \tag{OA1-13}$$

where $r_z^{\mathcal{TP}} \equiv |\{t : D_t = \mathcal{TP}_z\}|$ is the number of periods that the trade pattern $\mathcal{TP}_z$ is observed for $z \in \left\{1, ..., n^{\mathcal{TP}}\right\}$. The third equality holds as $K_{th}'K_{tj} = \mathbf{0}\ \forall h \neq j$ and $K_{th}'K_{tj} = I_{n_h^D}\ \forall h = j$ by definitions of (OA1-8) and (OA1-9).

Second, the (h, j) block of $Z_3\left(Z_3'Z_3\right)^{-1}Z_3'$, i.e., $B_h\left(Z_3'Z_3\right)^{-1}B_j'$, is equal to a matrix of zeros if the trade pattern of period h is different from that of period j and is equal to an identity matrix multiplied by a scalar if the trade pattern of the two periods is the same:

$$B_h\left(Z_3'Z_3\right)^{-1}B_j' = \sum_{z\in\{1,..,n^{\mathcal{TP}}\}} \frac{1}{r_z^{\mathcal{TP}}} K_{hz} I_{n^D_{\mathcal{TP}(z)}} K_{jz}' = \begin{cases} \frac{1}{r_h^{\mathcal{D}}} I_{n_h^D} & \text{if } D_h = D_j \\ \mathbf{0}_{n_h^D \times n_j^D} & \text{if } D_h \neq D_j \end{cases} \tag{OA1-14}$$

where $r_z^{\mathcal{D}} \equiv |\{t : D_t = D_z\}|$ is the number of periods that the trade pattern D_z is observed.

Finally, from (OA1-12) and (OA1-4), $Z_3\left(Z_3'Z_3\right)^{-1}Z_3'Z_2\left(Z_2'Z_2\right)^{-1}Z_2'$ can be decomposed into

$n^T \times n^T$ blocks:

$$
\begin{aligned}
T &\equiv Z_3\left(Z_3'Z_3\right)^{-1}Z_3'Z_2\left(Z_2'Z_2\right)^{-1}Z_2' \\
&= \begin{bmatrix}
B_1\left(Z_3'Z_3\right)^{-1}B_1'\frac{1}{n_1^D}\iota_{n_1^D}\iota_{n_1^D}' & \cdots & B_1\left(Z_3'Z_3\right)^{-1}B_{n^T}'\frac{1}{n_{n^T}^D}\iota_{n_{n^T}^D}\iota_{n_{n^T}^D}' \\
\vdots & \ddots & \vdots \\
B_1\left(Z_3'Z_3\right)^{-1}B_{n^T}'\frac{1}{n_1^D}\iota_{n_1^D}\iota_{n_1^D}' & \cdots & B_{n^T}\left(Z_3'Z_3\right)^{-1}B_{n^T}'\frac{1}{n_{n^T}^D}\iota_{n_{n^T}^D}\iota_{n_{n^T}^D}'
\end{bmatrix}
\end{aligned}
$$

where block (x, y) of T is given by

$$T(x,y) = B_x\left(Z_3'Z_3\right)^{-1}B_y'\frac{1}{n_y^D}\iota_{n_y^D}\iota_{n_y^D}'.$$

From (OA1-14), it is straightforward to see that $T(x,y) = T(y,x)'$. That is, if the trade pattern of period x is the same as that of period y, then $T(x,y) = T(y,x)' = \frac{1}{r_x^D n_x^D}\iota_{n_x^D}\iota_{n_x^D}' = \frac{1}{r_y^D n_y^D}\iota_{n_y^D}\iota_{n_y^D}'$; if the trade pattern of period x is different from that of period y, then $T(x,y) = T(y,x)' = \mathbf{0}_{n_x^D \times n_y^D}$.

Now, given that $Z_3\left(Z_3'Z_3\right)^{-1}Z_3'$, $Z_2\left(Z_2'Z_2\right)^{-1}Z_2'$, and T are all symmetric, it follows that

$$T = Z_3\left(Z_3'Z_3\right)^{-1}Z_3'Z_2\left(Z_2'Z_2\right)^{-1}Z_2' = T' = Z_2\left(Z_2'Z_2\right)^{-1}Z_2'Z_3\left(Z_3'Z_3\right)^{-1}Z_3'.$$

□

Proof of (OA1-11):

Proof. From (OA1-12) and the definition of Z_1 in (OA1-1), we can write T_3Z_1 as

$$T_3Z_1 = \begin{bmatrix} \sum_t B_1\left(Z_3'Z_3\right)^{-1}B_t'A_t \\ \vdots \\ \sum_t B_{n^T}\left(Z_3'Z_3\right)^{-1}B_t'A_t \end{bmatrix}.$$

Using (OA1-14), we have

$$B_x\left(Z_3'Z_3\right)^{-1}B_y'A_y = \begin{cases} \frac{1}{r_x^D}A_x = \frac{1}{r_y^D}A_y & \text{if } D_x = D_y \\ \mathbf{0}_{n_x^D \times n^D} & \text{if } D_x \neq D_y \end{cases} \tag{OA1-15}$$

With (OA1-15), it follows that

$$T_3 Z_1 = \begin{bmatrix} \sum_{t:D_t=D_1} \frac{1}{r_1^{\mathcal{D}}} A_1 \\ \vdots \\ \sum_{t:D_t=D_{n^T}} \frac{1}{r_{n^T}^{\mathcal{D}}} A_{n^T} \end{bmatrix} = \begin{bmatrix} A_1 \\ \vdots \\ A_{n^T} \end{bmatrix} = Z_1.$$

□

OA1.1.4 A numerical example with projection matrices to visualize differences across estimators

To clarify how the estimator works, we now spell out all the key matrices from the above discussions and provide a numerical example. For illustrative purposes, we use a much simpler data generating process:

$$\begin{aligned} p_{dt} &= \beta_0 + \beta_1 e_{dt} + \beta_2 m_{dt} \\ e_{dt} &= \sigma_e(m_{dt} + u_{dt}) \\ m_{dt} &= \vartheta_d + \epsilon_t + \psi_d * \upsilon_t \end{aligned}$$

with the following reduced form selection rule:

$$p_{dt} = \begin{cases} \text{observed} & \text{if} \quad \gamma_0 + \gamma_1 e_{dt} + \gamma_2 m_{dt} < 0 \\ \text{missing} & \text{if} \quad \gamma_0 + \gamma_1 e_{dt} + \gamma_2 m_{dt} \geq 0 \end{cases}$$

where ϑ_d, ϵ_t, ψ_t, υ_t and u_{dt} are simulated from a standard normal distribution. We set σ_e to be 0.5 such that the bilateral exchange rate shocks are slightly less volatile than the idiosyncratic marginal cost shocks. We set $\beta_1 = \beta_2 = 1$ such that an exchange rate appreciation of the home currency and a positive marginal cost shock increase the border price denominated in the home currency. This also implies a positive omitted variable bias. We set $\gamma_1 = -0.1$ and $\gamma_2 = 1$ such that the selection bias is also positive. The magnitude of γ_1 is set to be smaller than that of γ_2 to reflect the fact that the aggregate shocks (such as bilateral exchange rates) is less detrimental for the firm's entry decisions compared to idiosyncratic factors (such as the unobserved marginal cost). We reduce the number of destinations to 5 and the number of years to 4 to keep the size of the matrices tractable. To keep the example clean, we only allow for two distinct values of the factors affecting the time variation of the unobserved marginal cost (i.e., ϵ_t and υ_t). We set γ_0 such that half of the observations (destination-year pairs) are dropped.

Table OA1-1 shows one particular realization of such a data generating process. The firm exports in all four periods, and its decisions generate two unique trade patterns. In the first two years, the firm exports to destinations 2, 4 and 5. In the last two years, the firm exports only to destinations 4 and 5.

Table OA1-1: Simulated Data

Year	Destination	Trade Pattern	p_{dt}	e_{dt}	m_{dt}	ϵ_t	υ_t
1	2	2_4_5	-0.072	0.155	-0.227	0.843	0.277
1	4	2_4_5	0.178	-0.092	0.270	0.843	0.277
1	5	2_4_5	-1.138	-1.252	0.114	0.843	0.277
2	2	2_4_5	0.455	0.682	-0.227	0.843	0.277
2	4	2_4_5	0.636	0.366	0.270	0.843	0.277
2	5	2_4_5	0.068	-0.046	0.114	0.843	0.277
3	4	4_5	-0.313	0.689	-1.002	-0.191	1.117
3	5	4_5	-0.315	0.071	-0.387	-0.191	1.117
4	4	4_5	-1.099	-0.097	-1.002	-0.191	1.117
4	5	4_5	-0.747	-0.360	-0.387	-0.191	1.117

Z_1 is the matrix that contains the destination dummies. To economize on the matrix size, we only create dummies for destinations that are observed, i.e., we do not create dummies for destinations 1 and 3. For example, the first column of Z_1 reports the observations in which the firm sells to destination 2. From the matrix, we can see that the firm sells to destination 2 two times. Z_2 is the matrix that contains the year dummies. Z_3 gives our proposed destination-specific trade pattern dummies. As defined in (OA1-8) and (OA1-9), it is constructed by interacting the destination dummies with the trade pattern dummies. For example, the first three columns represent the dummy structure for the destinations related to the 2_4_5 trade pattern, i.e., $2-2\_4\_5$, $4-2\_4\_5$ and $5-2\_4\_5$. Similarly, the last two columns represent the dummy structure for the destinations related to the 4_5 trade pattern, i.e., $4-4\_5$ and $5-4\_5$.

$$Z_1 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad Z_2 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad Z_3 = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \tag{OA1-16}$$

From these, we can see clearly that P_2 is a destination demean process.

$$P_2 = \begin{bmatrix} 0.67 & -0.33 & -0.33 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ -0.33 & 0.67 & -0.33 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ -0.33 & -0.33 & 0.67 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.67 & -0.33 & -0.33 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -0.33 & 0.67 & -0.33 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -0.33 & -0.33 & 0.67 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0.50 & -0.50 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & -0.50 & 0.50 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.50 & -0.50 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -0.50 & 0.50 \end{bmatrix}$$

By way of example, for the first observation, $2/3p_{11} - 1/3p_{21} - 1/3p_{31} = p_{11} - \frac{1}{3}(p_{11} + p_{21} + p_{31})$.

As discussed in subsection OA1.1.2, $Z_2 (Z_2'Z_2)^{-1} Z_2'Z_1$ follows a trade pattern structure and $\bar{Z}$ suggests an additive relationship between the destination dummies Z_1 and the trade pattern dummies $Z_2 (Z_2'Z_2)^{-1} Z_2'Z_1$.

$$Z_2 (Z_2'Z_2)^{-1} Z_2'Z_1 = \begin{bmatrix} 0.33 & 0.33 & 0.33 \\ 0.33 & 0.33 & 0.33 \\ 0.33 & 0.33 & 0.33 \\ 0.33 & 0.33 & 0.33 \\ 0.33 & 0.33 & 0.33 \\ 0.33 & 0.33 & 0.33 \\ 0 & 0.50 & 0.50 \\ 0 & 0.50 & 0.50 \\ 0 & 0.50 & 0.50 \\ 0 & 0.50 & 0.50 \end{bmatrix} \quad \bar{Z} = Z_1 - Z_2 (Z_2'Z_2)^{-1} Z_2'Z_1 = \begin{bmatrix} 0.67 & -0.33 & -0.33 \\ -0.33 & 0.67 & -0.33 \\ -0.33 & -0.33 & 0.67 \\ 0.67 & -0.33 & -0.33 \\ -0.33 & 0.67 & -0.33 \\ -0.33 & -0.33 & 0.67 \\ 0 & 0.50 & -0.50 \\ 0 & -0.50 & 0.50 \\ 0 & 0.50 & -0.50 \\ 0 & -0.50 & 0.50 \end{bmatrix}$$

As we can see from (OA1-17), the projection P does not follow a particular structure. Therefore, our two-step decomposition $P = P_1P_2$ discussed in subsection OA1.1.2 helps to unveil the key economic mechanisms behind the statistical projection.

$$P = \begin{bmatrix} 0.46 & -0.29 & -0.17 & -0.21 & 0.04 & 0.17 & -0.13 & 0.13 & -0.13 & 0.13 \\ -0.29 & 0.46 & -0.17 & 0.04 & -0.21 & 0.17 & 0.13 & -0.13 & 0.13 & -0.13 \\ -0.17 & -0.17 & 0.33 & 0.17 & 0.17 & -0.33 & 0 & 0 & 0 & 0 \\ -0.21 & 0.04 & 0.17 & 0.46 & -0.29 & -0.17 & -0.13 & 0.13 & -0.13 & 0.13 \\ 0.04 & -0.21 & 0.17 & -0.29 & 0.46 & -0.17 & 0.13 & -0.13 & 0.13 & -0.13 \\ 0.17 & 0.17 & -0.33 & -0.17 & -0.17 & 0.33 & 0 & 0 & 0 & 0 \\ -0.13 & 0.13 & 0 & -0.13 & 0.13 & 0 & 0.38 & -0.38 & -0.13 & 0.13 \\ 0.13 & -0.13 & 0 & 0.13 & -0.13 & 0 & -0.38 & 0.38 & 0.13 & -0.13 \\ -0.13 & 0.13 & 0 & -0.13 & 0.13 & 0 & -0.13 & 0.13 & 0.38 & -0.38 \\ 0.13 & -0.13 & 0 & 0.13 & -0.13 & 0 & 0.13 & -0.13 & -0.38 & 0.38 \end{bmatrix} \qquad \text{(OA1-17)}$$

Let $Y = [-0.072, 0.178, -1.138, 0.455, 0.636, 0.068, -0.313, -0.315, -1.099, -0.747]'$ and $X = [0.155, -0.092, -1.252, 0.682, 0.366, -0.046, 0.689, 0.071, -0.097, -0.360]'$. The OLS estimator is given by $(X'X)^{-1}X'Y$, which gives an estimate of $\widehat{\beta}_1 = 0.745$. The estimator applying d and t fixed effects is given by $(X'P'PX)^{-1}X'P'Y$, which gives $\widehat{\beta}_1 = 1.508$. The estimator applying dD

and t fixed effects is given by $(X'P_2'P_3'P_3P_2X)^{-1}X'P_2'P_3'P_3P_2Y$, which gives the calibrated value of $\widehat{\beta}_1 = 1.000$.

OA1.1.5 Identifying markup elasticities in unbalanced panels: adding firm and product dimensions

In this subsection, we introduce firm and product panel dimensions and prove Proposition 1. The key idea is that the data structure of a more complicated customs dataset with four panel dimensions can be viewed as a collection of two dimensional problems presented in (OA1-1).

Let n_{fi}^D denote the total number of export destinations by the firm-product and n_{fit}^D $\left(n_{fit}^D \leq n_{fi}^D\right)$ be the number of observed destinations in year t. Let n_{fi}^T denote the maximum number of exporting years and the $n_{fi}^{DT} \equiv \sum_t n_{fit}^D$ be the number of observed transactions by firm-product fi. Let A_{fit} be the $\left(n_{fit}^D \times n_{fi}^D\right)$ matrix obtained from the $(n_{fi}^D \times n_{fi}^D)$ identity matrix from which, for each firm-product fi, the rows corresponding to the destinations not observed in year t have been omitted. For each firm-product fi, the destination and time fixed effects of the firm-product can be defined analogously to (OA1-1) as

$$Z_{fi,1} \equiv \begin{bmatrix} A_{fi1} \\ \vdots \\ A_{fin_{fi}^T} \end{bmatrix}, \quad Z_{fi,2} \equiv \begin{bmatrix} A_{fi1}\iota_{n_{fi}^D} & & \\ & \ddots & \\ & & A_{fin_{fi}^T}\iota_{n_{fi}^D} \end{bmatrix}$$

where $Z_{fi,1}$ is an $n_{fi}^{DT} \times n_{fi}^D$ matrix that gives the dummy structure for the destination fixed effects of firm-product fi and $Z_{fi,2}$ is an $n_{fi}^{DT} \times n_{fi}^T$ matrix that gives the dummy structure for the year fixed effects of firm-product fi. Similarly, the destination-specific trade pattern dummies of the firm-product, $Z_{fi,3}$, can be defined as in (OA1-8) and (OA1-9).

Let n^{FIDT} be the total number of (non-missing) observations in the dataset; n^{FI} be the total number of distinct firm-products in the dataset; $n^{FID} \equiv \sum_{fi} n_{fi}^D$ be the sum of distinct destinations over all firm-products; $n^{FIT} \equiv \sum_{fi} n_{fi}^T$ be the sum of distinct time periods over all firm-products; and $n^{FI\mathcal{DTP}} \equiv \sum_{fi} n_{fi}^{\mathcal{DTP}}$ be the sum of distinct destination-specific trade patterns over all firm-products. The dummy structure for the full dataset including all firm-products can be constructed as:

$$Z_1 \equiv \begin{bmatrix} Z_{1,1} & & \\ & \ddots & \\ & & Z_{n^{FI},1} \end{bmatrix}, \quad Z_2 \equiv \begin{bmatrix} Z_{1,2} & & \\ & \ddots & \\ & & Z_{n^{FI},2} \end{bmatrix}, \quad Z_3 \equiv \begin{bmatrix} Z_{1,3} & & \\ & \ddots & \\ & & Z_{n^{FI},3} \end{bmatrix}$$

where Z_1 is an $n^{FIDT} \times n^{FID}$ block diagonal matrix representing the dummy structure of firm-product-destination fixed effects; Z_2 is an $n^{FIDT} \times n^{FIT}$ block diagonal matrix representing the dummy structure of firm-product-time fixed effects; and Z_3 is an $n^{FIDT} \times n^{FI\mathcal{DTP}}$ block diagonal matrix representing the dummy structure of firm-product-destination-trade pattern fixed effects. The matrices inside Z_1, Z_2 and Z_3 represent the dummy structure of the corresponding firm-product. For example, the $Z_{1,1}$ and $Z_{n^{FI},1}$ inside Z_1 give the dummy structure of destination fixed effects for the first and the last firm-product in the dataset respectively. Matrices Z_1, Z_2 and Z_3 are block diagonal because all the fixed effects we consider are firm-product specific, under which the elements of $Z_{fi,1}$, $Z_{fi,2}$ and $Z_{fi,3}$ must be zero for the observations associated with the firm-products other than fi.

Proof of Proposition 1:

Proof. Define the two demeaning processes of the TPSFE as

$$P_2 \equiv I_{n^{FIDT}} - Z_2 \left(Z_2' Z_2\right)^{-1} Z_2' \qquad \text{(step 1 of TPSFE)}$$
$$P_3 \equiv I_{n^{FIDT}} - Z_3 \left(Z_3' Z_3\right)^{-} Z_3' \qquad \text{(step 2 of TPSFE)}$$

where $I_{n^{FIDT}}$ is an $n^{FIDT} \times n^{FIDT}$ identity matrix.

We want to show

$$P_3 P_2 Z_1 = \mathbf{0},$$
$$P_3 P_2 Z_2 = \mathbf{0},$$
$$P_3 P_2 Z_3 = \mathbf{0}.$$

First of all, similar to the two-dimensional case, the second equality holds trivially by the design of P_2 (since $[I_{n^{FIDT}} - Z_2 \left(Z_2' Z_2\right)^{-1} Z_2'] Z_2 = \mathbf{0}$). Secondly, block diagonal matrices have a nice property that the multiplication of two conformable block diagonal matrices is equal to the multiplication of the corresponding diagonal blocks of the two matrices. This allows us to apply the key relationships in the two-dimensional panel case to each of the block matrices in Z_1, Z_2 and Z_3. Specifically, we have

$$\begin{aligned} Z_3 \left(Z_3' Z_3\right)^{-} Z_3' Z_1 &= \begin{bmatrix} Z_{1,3}\left(Z_{1,3}' Z_{1,3}\right)^{-} Z_{1,3}' Z_{1,1} & & \\ & \ddots & \\ & & Z_{n^{FI},3}\left(Z_{n^{FI},3}' Z_{n^{FI},3}\right)^{-} Z_{n^{FI},3}' Z_{n^{FI},1} \end{bmatrix} \\ &= \begin{bmatrix} Z_{1,1} & & \\ & \ddots & \\ & & Z_{n^{FI},1} \end{bmatrix} = Z_1 \end{aligned} \qquad \text{(OA1-18)}$$

where the first equality uses the property of block diagonal matrices and the the second equality uses the relationship of (OA1-11). Similarly, using the property of block diagonal matrices and the firm-product level relationship (OA1-10), it is straightforward to show the following equations hold:[4]

$$Z_3\left(Z_3'Z_3\right)^{-}Z_3'Z_2\left(Z_2'Z_2\right)^{-1}Z_2' = Z_2\left(Z_2'Z_2\right)^{-1}Z_2'Z_3\left(Z_3'Z_3\right)^{-}Z_3' \tag{OA1-19}$$

$$Z_3\left(Z_3'Z_3\right)^{-}Z_3'Z_2\left(Z_2'Z_2\right)^{-1}Z_2'Z_1 = Z_2\left(Z_2'Z_2\right)^{-1}Z_2'Z_1 \tag{OA1-20}$$

Using (OA1-18), (OA1-19) and (OA1-20), it follows that

$$\begin{aligned} P_3P_2Z_1 &= [I_{n^{FIDT}} - Z_3\left(Z_3'Z_3\right)^{-}Z_3'][I_{n^{FIDT}} - Z_2\left(Z_2'Z_2\right)^{-1}Z_2']Z_1 \\ &= [I_{n^{FIDT}} - Z_2\left(Z_2'Z_2\right)^{-1}Z_2']Z_1 - Z_3\left(Z_3'Z_3\right)^{-}Z_3'[I_{n^{FIDT}} - Z_2\left(Z_2'Z_2\right)^{-1}Z_2']Z_1 \\ &= [I_{n^{FIDT}} - Z_2\left(Z_2'Z_2\right)^{-1}Z_2']Z_1 - [I_{n^{FIDT}} - Z_2\left(Z_2'Z_2\right)^{-1}Z_2']Z_1 = \mathbf{0} \end{aligned}$$

and

$$\begin{aligned} P_3P_2Z_3 &= [I_{n^{FIDT}} - Z_3\left(Z_3'Z_3\right)^{-}Z_3'][I_{n^{FIDT}} - Z_2\left(Z_2'Z_2\right)^{-1}Z_2']Z_3 \\ &= [I_{n^{FIDT}} - Z_3\left(Z_3'Z_3\right)^{-}Z_3']Z_3 - [I_{n^{FIDT}} - Z_3\left(Z_3'Z_3\right)^{-}Z_3']Z_2\left(Z_2'Z_2\right)^{-1}Z_2'Z_3 \\ &= \mathbf{0} - [Z_2\left(Z_2'Z_2\right)^{-1}Z_2'Z_3 - Z_3\left(Z_3'Z_3\right)^{-}Z_3'Z_2\left(Z_2'Z_2\right)^{-1}Z_2'Z_3] = \mathbf{0} \end{aligned}$$

□

OA1.2 The TPSFE estimator in view of the control function approach

In this subsection, we discuss how our approach relates to the classical control function approach (e.g., Heckman (1979)) and the first difference approach pursued by Kyriazidou (1997).[5] We start by rewriting the problem addressed by Heckman (1979) in his seminal work on selection in cross-sectional data. In what follows, think of p_t as the price of a product, and as a function of a set of

[4]It is worth noting that the modification of the projection matrix in an unbalanced panel needs to be done with extreme caution. A seemingly more general setting can, in lots of cases, result in more (rather than less) bias. Alternative demeaning or partition methods do not necessarily satisfy (OA1-19) and (OA1-20) and can potentially result in substantial biases.

[5]Our estimation approach is related to three strands of the panel data literature. The first strand focuses on estimating the parameter of interest in a panel data model with selection. Existing discussions are restricted to selection equations with one dimensional fixed effects or those that can be combined into one dimensional fixed effects (see recent handbook chapters by Verbeek and Nijman (1996), Honoré et al. (2008) and Matyas (2017) for a complete literature review). The second strand constructs methods of estimating selection equations with unobserved heterogeneity along two dimensions (e.g., Fernández-Val and Weidner (2016) and Charbonneau (2017)). Our approach differs from theirs in that we do not need to estimate the selection equation, but instead, we rely on the realized patterns to formulate a new panel dimension to address the selection problem. A few papers have examined multi-dimensional fixed effects in unbalanced panels (e.g., Wansbeek and Kapteyn (1989) and Balazsi et al. (2018)).

controls $\boldsymbol{x}_t'$, observed if the firm decides to enter the market:

$$\begin{aligned} p_t &= \boldsymbol{x}_t'\boldsymbol{\beta} + \varepsilon_t \\ &= \boldsymbol{x}_t'\boldsymbol{\beta} + E(\varepsilon_t|\boldsymbol{x}_t, s_t) + \nu_t \\ s_t &= \mathbb{1}\{\boldsymbol{w}_t'\boldsymbol{\gamma} + u_t\} \end{aligned}$$

where s_t is an indicator variable that equals one if p_t is observed; $E(\varepsilon_t|\boldsymbol{x}_t, s_t)$ is the selection bias and $\nu_t \equiv [\varepsilon_t - E(\varepsilon_t|\boldsymbol{x}_t, s_t)]$ is an error term that is uncorrelated with the vector of observed variables $\boldsymbol{x_t}$ and the selection bias. $\boldsymbol{w}_t$ is a vector of observed variables in the selection equation which can overlap with the elements in $\boldsymbol{x}_t$. As is well known, selection bias is a problem if $E(\varepsilon_t|\boldsymbol{x}_t, s_t) \neq 0$. The solution of Heckman (1979) is to estimate the function of $E(\varepsilon_t|\boldsymbol{x}_t, s_t)$ under some parametric assumptions and then add the predicted value $E(\widehat{\varepsilon_t|\boldsymbol{x}_t, s_t})$ as a control variable in the main estimating equation. The essence of this approach is to estimate the parameter of interest conditional on the probability of an observation being observed.

Closer to our problem, where the firm chooses among potential export destination markets, Kyriazidou (1997) studies selection in a two dimensional panel with one fixed effect:

$$\begin{aligned} p_{dt} &= \boldsymbol{x}_{dt}'\boldsymbol{\beta} + \mathcal{M}_d + \varepsilon_{dt} && \text{(OA1-21)} \\ &= \boldsymbol{x}_{dt}'\boldsymbol{\beta} + \mathcal{M}_d + E(\mathcal{M}_d|\boldsymbol{x}_{dt}, s_{dt}) + E(\varepsilon_{dt}|\boldsymbol{x}_{dt}, s_{dt}) + \nu_{dt} \\ s_{dt} &= \mathbb{1}\{\boldsymbol{w}_{dt}'\boldsymbol{\gamma} + \mathcal{W}_d + u_{dt}\} && \text{(OA1-22)} \end{aligned}$$

where $\mathcal{M}_d$ and $\mathcal{W}_d$ are unobserved variables varying along the destination d dimension (i.e. destination fixed effects). $E(\mathcal{M}_d|\boldsymbol{x}_{dt}, s_{dt})$ and $E(\varepsilon_{dt}|\boldsymbol{x}_{dt}, s_{dt})$ represent the selection biases caused by the unobserved destination-specific heterogeneity and other omitted variables, respectively. $\nu_{dt} \equiv [\varepsilon_{dt} - E(\varepsilon_{dt}|\boldsymbol{x}_{dt}, s_{dt}) - E(\mathcal{M}_d|\boldsymbol{x}_{dt}, s_{dt})]$ is an error term that is uncorrelated with the observed explanatory variables and the selection biases. p_{dt} denotes the price and s_{dt} is an indicator variable that takes a value of one if the firm exports to destination d in period t and zero otherwise.[6] Kyriazidou (1997) notes that $E(\mathcal{M}_d|\boldsymbol{x}_{dt}, s_{dt})$ and $E(\varepsilon_{dt}|\boldsymbol{x}_{dt}, s_{dt})$ no longer vary along the time dimension when $\boldsymbol{w}_{d1}'\boldsymbol{\gamma} = \boldsymbol{w}_{d2}'\boldsymbol{\gamma}$, i.e., under the following *conditional exchangeability* condition:

$$F(\varepsilon_{d1}, \varepsilon_{d2}, u_{d1}, u_{d2}|\boldsymbol{\vartheta}_d) = F(\varepsilon_{d2}, \varepsilon_{d1}, u_{d2}, u_{d1}|\boldsymbol{\vartheta}_d) \qquad \text{(OA1-23)}$$

where $\boldsymbol{\vartheta}_d \equiv (\boldsymbol{x}_{d1}, \boldsymbol{x}_{d2}, \boldsymbol{w}_{d1}, \boldsymbol{w}_{d2}, \mathcal{W}_d, \mathcal{M}_d)$ is a destination specific vector containing information on observed and unobserved variables. Condition (OA1-23) states that $(\varepsilon_{d1}, \varepsilon_{d2}, u_{d1}, u_{d2})$ and

[6] Kyriazidou (1997) discusses a case in which the number of time periods is small ($n^T = 2$). Therefore, a Heckman (1979) style estimator cannot be applied as it will suffer from the incidental parameters problem due to the limited time dimension.

$(\varepsilon_{d2}, \varepsilon_{d1}, u_{d2}, u_{d1})$ are identically distributed conditional on $\boldsymbol{\vartheta}_d$. As noted by Kyriazidou (1997), the main term causing the selection bias, $E(\varepsilon_{dt}|\boldsymbol{x}_{dt}, s_{dt})$, is no longer time-varying when $\boldsymbol{w}'_{d1}\boldsymbol{\gamma} = \boldsymbol{w}'_{d2}\boldsymbol{\gamma}$ under condition (OA1-23):

$$
\begin{aligned}
& E(\varepsilon_{d1}|s_{d1} = 1, s_{d2} = 1|\boldsymbol{\vartheta}_d) \\
\equiv\ & E(\varepsilon_{d1}|u_{d1} < \boldsymbol{w}'_{d1}\boldsymbol{\gamma} + \mathcal{W}_d, u_{d2} < \boldsymbol{w}'_{d2}\boldsymbol{\gamma} + \mathcal{W}_d, \boldsymbol{\vartheta}_d) \\
=\ & E(\varepsilon_{d1}|u_{d1} < \boldsymbol{w}'_{d2}\boldsymbol{\gamma} + \mathcal{W}_d, u_{d2} < \boldsymbol{w}'_{d1}\boldsymbol{\gamma} + \mathcal{W}_d, \boldsymbol{\vartheta}_d) && \text{(OA1-24)} \\
=\ & E(\varepsilon_{d2}|u_{d2} < \boldsymbol{w}'_{d2}\boldsymbol{\gamma} + \mathcal{W}_d, u_{d1} < \boldsymbol{w}'_{d1}\boldsymbol{\gamma} + \mathcal{W}_d, \boldsymbol{\vartheta}_d) && \text{(OA1-25)} \\
\equiv\ & E(\varepsilon_{d2}|s_{d2} = 1, s_{d1} = 1|\boldsymbol{\vartheta}_d)
\end{aligned}
$$

where the first equality (OA1-24) holds because $\boldsymbol{w}'_{d1}\boldsymbol{\gamma} = \boldsymbol{w}'_{d2}\boldsymbol{\gamma}$ and the second equality (OA1-25) holds because of the *conditional exchangeability* condition (OA1-23). Since the selection bias is no longer time varying, i.e., $E(\varepsilon_{d1}|s_{d1} = 1, s_{d2} = 1|\boldsymbol{\vartheta}_d) = E(\varepsilon_{d2}|s_{d2} = 1, s_{d1} = 1|\boldsymbol{\vartheta}_d)$, it can be absorbed by destination fixed effects. Kyriazidou (1997) proposes a two-step estimator: the first step consistently estimates $\widehat{\boldsymbol{\gamma}}$ and the second step differences out the fixed effect and the selection terms conditional on destinations for which $\boldsymbol{w}'_{d1}\widehat{\boldsymbol{\gamma}} = \boldsymbol{w}'_{d2}\widehat{\boldsymbol{\gamma}}$.

Our problem can be specified in (OA1-26) and (OA1-27) as follows:

$$
p_{fidt} = \boldsymbol{x}'_{dt}\boldsymbol{\beta} + \mathcal{M}_{fid} + \mathcal{C}_{fit} + \varepsilon_{fidt} \tag{OA1-26}
$$

$$
s_{fidt} = \mathbb{1}\{\boldsymbol{w}'_{dt}\boldsymbol{\gamma} + \mathcal{W}_{fid} + \mathcal{Q}_{fit} + u_{fidt}\} \tag{OA1-27}
$$

This problem differs from Kyriazidou (1997)'s in two crucial respects. On the one hand, our problem adds unobserved firm-product-time-varying variables $\mathcal{C}_{fit}$ to equation (OA1-21) and $\mathcal{Q}_{fit}$ to equation (OA1-22). In the presence of these time-varying unobserved factors, the *conditional exchangeablitiy* condition no longer holds. On the other hand, many aggregate-level economic indicators of interest in our study—e.g., exchange rates—vary along the destination and time dimensions, but not at the firm or product dimensions. This is actually helpful. As discussed below, the fact that key variables vary along dimensions that are a subset of the dimensions of the dependent variable facilitates the control of selection biases.

While the method we propose to address the above problem is conceptually close to Kyriazidou (1997), the approach we take is fundamentally different. Specifically, if we were to follow Kyriazidou (1997)'s approach, we would require all variables driving $\mathcal{Q}_{fit}$ to be observed and controlled for. For our purposes, however, this condition cannot be satisfied—if only because the marginal cost is unobserved and cannot be generally estimated at product-firm level. Rather, we need to rely on a method that avoids direct estimation of the selection equation and works in a multi-dimensional panel where more than one fixed effect is present in both the structural equation and the selection equation. Our main innovation is to use the realized selection pattern in a panel dimension, instead of the observed variables in the selection equation, to control for selection biases.

Before analyzing how our method addresses the general problem characterized in equations (OA1-26) and (OA1-27), we find it useful to provide insight by focusing on a two-dimensional panel, tracking the choices of a single firm selling one product across a set of endogenous destinations.

OA1.2.1 A two dimensional panel case

Consider the following for a firms' destination choices with two panel dimensions, destination d and time t:

$$p_{dt} = \boldsymbol{x}'_{dt}\boldsymbol{\beta} + \mathcal{M}_d + \mathcal{C}_t + \varepsilon_{dt} \tag{OA1-28}$$

$$s_{dt} = \mathbb{1}\{u_{dt}\} \tag{OA1-29}$$

where $\mathcal{M}_d$ and $\mathcal{C}_t$ are unobserved destination and time specific factors, respectively, which are potentially correlated with the explanatory variables contained in the vector $\boldsymbol{x}_{dt}$. The price p_{dt} is observed only if s_{dt} equals one or equivalently, if $u_{dt} > 0$.

The first two steps in our approach involve transforming the variables in (OA1-28) to eliminate the unobserved destination and time specific factors. Specifically, in the first step, we demean variables at the time (t) dimension. In the second step, we demean variables at the destination-trade pattern (dD) dimension. After applying these two transformations,

$$\ddot{p}_{dt} = \ddot{\boldsymbol{x}}'_{dt}\boldsymbol{\beta} + \ddot{\varepsilon}_{dt}$$

where

$$\ddot{\boldsymbol{x}}_{dt} = \boldsymbol{x}_{dt} - \frac{1}{n_t^D}\sum_{d\in D_t}\boldsymbol{x}_{dt} - \frac{1}{n_{dD}^T}\sum_{t\in T_{dD}}\boldsymbol{x}_{dt} + \frac{1}{n_{dD}^T}\sum_{t\in T_{dD}}\frac{1}{n_t^D}\sum_{d\in D_t}\boldsymbol{x}_{dt} \tag{OA1-30}$$

$$\ddot{\varepsilon}_{dt} = \varepsilon_{dt} - \frac{1}{n_t^D}\sum_{d\in D_t}\varepsilon_{dt} - \frac{1}{n_{dD}^T}\sum_{t\in T_{dD}}\varepsilon_{dt} + \frac{1}{n_{dD}^T}\sum_{t\in T_{dD}}\frac{1}{n_t^D}\sum_{d\in D_t}\varepsilon_{dt}, \tag{OA1-31}$$

D_t is the set of destinations the firm serves at time t; and $n_t^D \equiv |D_t|$ the number of export destinations at time t. Similarly, T_{dD} denotes the set of time periods in which a destination-specific trade pattern dD is observed, and n_{dD}^T represents the corresponding number of time periods in which the destination-specific trade pattern emerges. For our proposed approach to work in a two dimensional panel, we need[7]

$$F(\varepsilon_{dD1},\varepsilon_{dD2},u_{dD1},u_{dD2}|\boldsymbol{\vartheta}_{dD})=F(\varepsilon_{dD2},\varepsilon_{dD1},u_{dD2},u_{dD1}|\boldsymbol{\vartheta}_{dD}), \tag{OA1-33}$$

where we use ε_{dD1} to indicate the first error within the destination-specific trade pattern dD. Given (OA1-33), it is straightforward to see that the selection bias can be differenced out over two time periods within a destination-specific trade pattern dD, since the following relationship holds:

$$E\left(\varepsilon_{dDt}|u_{dD1}>0,u_{dD2}>0,\boldsymbol{\vartheta}_{dD}\right)=E\left(\varepsilon_{dD\tau}|u_{dD1}>0,u_{dD2}>0,\boldsymbol{\vartheta}_{dD}\right)\quad \forall\tau\in T_{dD} \tag{OA1-34}$$

Condition (OA1-33) can be viewed as the analog of the *conditional exchangeability* assumption imposed by Kyriazidou (1997). Instead of controlling for the relationship among the observed variables in the selection process (i.e., $\boldsymbol{w}'_{d1}\boldsymbol{\gamma}=\boldsymbol{w}'_{d2}\boldsymbol{\gamma}$), we control for the realised patterns of selection in a panel dimension (i.e., the pattern of d conditional on t). That is, as long as the distribution of errors is the same for all time periods satisfying a destination-specific trade pattern dD, our approach produces unbiased and consistent estimates.[8]

OA1.2.2 General setting

We now discuss the general multi-dimensional setting specified in (OA1-26) and (OA1-27). With an additional dimension,[9] we can write the condition for identification as follows:

$$E\left[E\left(\varepsilon_{fidDt}|\mathbf{s}_{fidD},\boldsymbol{\vartheta}_{fidD}\right)\Big|dt\right]=E\left[E\left(\varepsilon_{fidD\tau}|\mathbf{s}_{fidD},\boldsymbol{\vartheta}_{fidD}\right)\Big|dt\right]\quad \forall\,\tau\in T_{fidD} \tag{OA1-35}$$

where $\mathbf{s}_{fidD}\equiv(\boldsymbol{w}'_{d1}\boldsymbol{\gamma}+\mathcal{W}_{fid}+\mathcal{Q}_{if1}+u_{fidD1}>0,...,\boldsymbol{w}'_{dn^T_{fidD}}\boldsymbol{\gamma}+\mathcal{W}_{fid}+\mathcal{Q}_{ifn^T_{fidD}}+u_{fidDn^T_{fidD}}>0)$, $\boldsymbol{\vartheta}_{fidD}\equiv(\boldsymbol{x}_{dD1},...,\boldsymbol{x}_{dDn^T_{fidD}},\boldsymbol{w}_{dD1},...,\boldsymbol{w}_{dDn^T_{fidD}},\mathcal{W}_{fid},\mathcal{M}_{fid})$ and $E(.|dt)$ means taking the expectation over the firm (f) and product (i) panel dimensions while keeping the destination and time panel dimensions fixed.

[7]Note that Kyriazidou (1997)'s original conditions (and proofs) only cover the case when the number of time periods is equal to two. For a more general case with more than two time periods, we impose a condition:

$$E\left(\varepsilon_{dDt}|u_{dD1}>0,...,u_{dDn^T_{dD}}>0,\vartheta_{dD}\right)=E\left(\varepsilon_{dD\tau}|u_{dD1}>0,...,u_{dDn^T_{dD}}>0,\vartheta_{dD}\right)\quad\forall\tau\in T_{dD} \tag{OA1-32}$$

As will be discussed later, our estimator works under a much weaker condition than (OA1-32) if another panel dimension is available.

[8]The condition for consistency, i.e., $E(s_{dt}\ddot{\boldsymbol{x}}_{dt}\ddot{\varepsilon}_{dt})=0$, is satisfied under (OA1-32). First, note that $\frac{1}{n^D_t}\sum_{d\in D_t}\varepsilon_{dt}-\frac{1}{n^T_{dD}}\sum_{t\in T_{dD}}\frac{1}{n^D_t}\sum_{d\in D_t}\varepsilon_{dt}=0$. This is because the expression $\frac{1}{n^D_t}\sum_{d\in D_t}\varepsilon_{dt}$ is moving at the dD dimension only. As there is no variation left after conditioning on the dD dimension, the demeaning process naturally gives zero. Second, demeaning conditional on the same trade pattern is zero under assumption (OA1-32), i.e., $E\left(\varepsilon_{dt}-\frac{1}{n^T_{dD}}\sum_{t\in T_{dD}}\varepsilon_{dt}\Big|s_{dD1},s_{dD2},s_{dD3},...,\vartheta_{dD}\right)=0$.

[9]In the following discussions, we consider firm and product as one combined panel dimension fi.

As can be seen from (OA1-35), we no longer need the error to be zero conditional on the observed pattern ($E\left(\varepsilon_{fidDt} - \varepsilon_{fidD\tau} | \mathbf{s}_{fidD}, \boldsymbol{\vartheta}_{fidD}\right) = 0$) as in the two dimensional case. Instead, it is sufficient to have the expectation of $E\left(\varepsilon_{fidDt} - \varepsilon_{fidD\tau} | \mathbf{s}_{fidD}, \boldsymbol{\vartheta}_{fidD}\right)$ be zero, once it is aggregated at the firm and product dimension. For example, if $E\left(\varepsilon_{fidDt} - \varepsilon_{fidD\tau} | \mathbf{s}_{fidD}, \boldsymbol{\vartheta}_{fidD}\right)$ consists of random errors for each firm and product, the mean of these random errors converges to zero when the number of firm-product pairs increases.

We now show that our proposed approach gives unbiased estimates under condition (OA1-35). Let $v_{fidt} \equiv \mathcal{M}_{fid} + \mathcal{C}_{fit} + \varepsilon_{fidt}$. The underlying independent variables and the error term under our estimation approach can be written as

$$\ddot{\boldsymbol{x}}_{fidt} = \boldsymbol{x}_{dt} - \frac{1}{n_{fit}^{D}} \sum_{d \in D_{fit}} \boldsymbol{x}_{dt} - \frac{1}{n_{fidD}^{T}} \sum_{t \in T_{fidD}} \boldsymbol{x}_{dt} + \frac{1}{n_{fidD}^{T}} \sum_{t \in T_{fidD}} \frac{1}{n_{fit}^{D}} \sum_{d \in D_{fit}} \boldsymbol{x}_{dt} \tag{OA1-36}$$

$$\ddot{v}_{fidt} = v_{fidt} - \frac{1}{n_{fit}^{D}} \sum_{d \in D_{fit}} v_{fidt} - \frac{1}{n_{fidD}^{T}} \sum_{t \in T_{fidD}} v_{fidt} + \frac{1}{n_{fidD}^{T}} \sum_{t \in T_{fidD}} \frac{1}{n_{fit}^{D}} \sum_{d \in D_{fit}} v_{fidt}. \tag{OA1-37}$$

The independent variable of interest now varies along four dimensions because it embodies selection that varies across firms and products, even if the variable is specified for only two dimensions, i.e., $\boldsymbol{x}_{dt}$ or e_{dt}.

Note that the exchange rate depends on the firm and product dimensions only through trade and time patterns. To see this, it is useful to rewrite the variables in expressions (OA1-36) and (OA1-37) in terms of their corresponding variability:

$$\begin{aligned}
\ddot{\boldsymbol{x}}_{fidt} &= \boldsymbol{x}_{dt} - \boldsymbol{x}_{Dt} - \boldsymbol{x}_{dT} + \boldsymbol{x}_{DT} \\
\ddot{v}_{fidt} &= v_{fidt} - v_{fiDt} - v_{fidT} + v_{fiDT} \\
&= \varepsilon_{fidt} - \varepsilon_{fiDt} - \varepsilon_{fidT} + \varepsilon_{fiDT} \\
&= \ddot{\varepsilon}_{fidt}.
\end{aligned}$$

Rearranging these expressions, we can show that our main variables of interest $\boldsymbol{x}$ (including exchange rates) in the following expression no longer depend on firm and product dimensions:

$$\frac{1}{n^{FIDT}} \sum_{fidt} \ddot{\varepsilon}_{fidt} \ddot{\boldsymbol{x}}_{fidt} = \frac{1}{n^{FIDT}} \sum_{fidt} (\varepsilon_{fidt} - \varepsilon_{fiDt} - \varepsilon_{fidT} + \varepsilon_{fiDT}) \boldsymbol{x}_{dt} \tag{OA1-38}$$

$$= \frac{1}{n^{FIDT}} \sum_{fidt} (\varepsilon_{fidt} - \varepsilon_{fidT}) \boldsymbol{x}_{dt}. \tag{OA1-39}$$

As a result, the identification condition, $E(\ddot{\varepsilon}_{fidt}\ddot{\boldsymbol{x}}_{fidt}s_{fidt}) = 0$, can be rewritten as

$$\begin{aligned}
& E(\ddot{\varepsilon}_{fidt}\ddot{\boldsymbol{x}}_{fidt}s_{fidt}) \\
&= E\left[(\varepsilon_{fidt} - \varepsilon_{fidT})\boldsymbol{x}_{dt}s_{fidt}\right] \\
&= E\left\{\boldsymbol{x}_{dt}E\left[E\left(\varepsilon_{fidt} - \varepsilon_{fidT}|\mathbf{s}_{fidD}, \boldsymbol{\vartheta}_{fidD}\right)\Big|dt\right]\right\} \\
&= E\left\{\boldsymbol{x}_{dt}E\left[E\left(\varepsilon_{fidDt} - \frac{1}{n^T_{fidD}}\sum_{\tau\in T_{fidD}}\varepsilon_{fidD\tau}|\mathbf{s}_{fidD}, \boldsymbol{\vartheta}_{fidD}\right)\Big|dt\right]\right\} \\
&= 0
\end{aligned} \quad \text{(OA1-40)}$$

where the first equality follows from using (OA1-39) under our proposed "within transformation"; the second equality from applying the law of iterated expectations; and the last equality from using condition (OA1-35).

Two remarks are in order to clarify the implications of our identification condition and place our approach in the literature. First, note that the condition (OA1-35) is trivially satisfied if ε is always zero. For example, if goods sold to different destinations by the same firm under the same product category are identical, the marginal cost is only firm-product-time specific and therefore absorbed by $\mathcal{C}_{fit}$, leaving no additional residual term. It is worth stressing that the maintained assumption that marginal costs are non-destination-specific is implicit in studies aimed at estimating productivity (as these do not try to distinguish the marginal cost at the destination level)—see, e.g., Olley and Pakes (1996), Levinsohn and Petrin (2003), Wooldridge (2009) and De Loecker et al. (2016).

Second, an important instance in which condition (OA1-35) is satisfied is when the distribution of the destination-specific component does not change over time, e.g., when the composition of shipments is such that high quality varieties of a product are consistently sold to high-income destinations. From this perspective, the condition clarifies that the existence of destination-specific marginal cost components in ε does not automatically lead to a violation of identification.

OA1.3 The TPSFE estimator relative to De Loecker et al. (2016)

In this subsection, we extend the framework of De Loecker et al. (2016) to add a destination dimension, and discuss the structural assumptions that would be required for our main identification condition (OA1-35) to be satisfied in this new framework.

OA1.3.1 Structural interpretation of assumptions required by our estimator

We start by writing the production function as follows:

$$Q_{fidt} = F_{fi}(\boldsymbol{V}_{fidt}, \boldsymbol{K}_{fidt})\Omega_{fit}\vartheta_{fid} \tag{OA1-41}$$

where Q_{fidt} represents the quantity of exports for product i from firm f to destination d at time t; $\boldsymbol{V}_{fidt}$ denotes a vector of variable inputs, $\{V^1_{fidt}, V^2_{fidt}, ..., V^v_{fidt}\}$; $\boldsymbol{K}_{fidt}$ denotes a vector of dynamic inputs; a firm-product pair make decisions on allocating its dynamic inputs across destinations in each time period, $\{K^1_{fidt}, K^2_{fidt}, ..., K^k_{fidt}\}$. We stress that the above function allows for destination-specific inputs $\{\boldsymbol{V}_{fidt}, \boldsymbol{K}_{fidt}\}$ as well as destination-specific productivity differences, ϑ_{fid}, at the firm and product level. In addition, we allow for the production function and Hicks-neutral productivity to be firm-product specific.

Specifically, we posit the following:

1. The production technology is firm-product-specific.

2. $F_{fi}(.)$ is continuous and twice differentiable w.r.t. at least one element of V_{fidt}, and this element of V_{fidt} is a static (i.e., freely adjustable or variable) input in the production of product i.

3. $F_{fi}(.)$ is constant return to scale.

4. Hicks-neutral productivity Ω_{fit} is log-additive.

5. The destination specific technology advantage ϑ_{fid} takes a log-additive form and is not time varying.

6. Input prices $\boldsymbol{W}_{fit}$ are firm-product-time specific.

7. The state variables of the firm are

$$\boldsymbol{s}_{fit} = \{D_{fit}, \boldsymbol{K}_{fit}, \Omega_{fit}, \vartheta_{fid}, \boldsymbol{G}_{fi}, \boldsymbol{r}_{fidt}\} \tag{OA1-42}$$

   where $\boldsymbol{G}_{fi}$ includes variables indicating firm and product properties, e.g., firm registration types, product differentiation indicators. $\boldsymbol{r}_{fidt}$ collects other observables including variables that track the destination market conditions, such as the bilateral exchange rate and destination CPI.

8. Firms minimize short-run costs taking output quantity, Q_{fidt}, and input prices, $\boldsymbol{W}_{fit}$, at time t as given.

The assumptions 1, 2, 4, 8 are standard in the literature. De Loecker et al. (2016) also posit them, but in our version we allow the production function to be firm specific and the Hicks-neutral productivity to be product-specific. Compared to the conditions assumed in the literature, assumption 5 is a relaxation: it allows for the possibility that (log-additive) productivity be destination-specific.

Assumptions 6 and 7 allow prices of inputs to be firm and product specific. These two conditions indicate that firms source inputs at the product level, and then allocate these inputs into production for different destinations. Note that the firm can arrange different quantities of inputs and have different marginal costs across destinations for the same product.

The assumption that is crucial to our identification is that the production technology is constant returns to scale (condition 3). This condition implies that the marginal cost at the firm-product-destination level does not depend on the quantity produced. If changes in relative demand and exports across destinations were systematically associated to changes in relative marginal costs, condition (OA1-35) would be violated. As discussed in the next subsection, looking at the solution to the firms' cost minimization problem, condition 3 ensures that the difference in the marginal costs across destinations only reflects technology differences varying at the destination dimension.

OA1.3.2 The cost minimization problem by firm-product pair

Write the cost function

$$\mathcal{L}(\boldsymbol{V}_{fidt}, \boldsymbol{K}_{fidt}, \lambda_{fidt}) = \sum_{v=1}^{V} W_{fit}^{v} \sum_{d \in D_{fit}} V_{fidt}^{v} + \sum_{k=1}^{K} R_{fit}^{k} \left(\sum_{d \in D_{fit}} K_{fidt}^{k} - K_{fit}^{k} \right) + \sum_{d \in D_{fit}} \lambda_{fidt} [Q_{fidt} - F_{fi}(\boldsymbol{V}_{fidt}, \boldsymbol{K}_{fidt}) \Omega_{fit} \vartheta_{fid}]$$

where K_{fit}^{k} is the accumulated capital input k in the previous period; K_{fidt}^{k} stands for the corresponding allocation for destination d; R_{fit}^{k} is the implied cost of capital.[10]

The F.O.C.s of the cost minimization problem are

$$\frac{\partial \mathcal{L}_{fit}}{\partial V_{fidt}^{v}} = W_{fit}^{v} - \lambda_{fidt} \Omega_{fit} \vartheta_{fid} \frac{\partial F_{fi}(.)}{\partial V_{fidt}^{v}} = 0, \tag{OA1-43}$$

$$\frac{\partial \mathcal{L}_{fit}}{\partial K_{fidt}^{k}} = R_{fit}^{k} - \lambda_{fidt} \Omega_{fit} \vartheta_{fid} \frac{\partial F_{fi}(.)}{\partial K_{fidt}^{k}} = 0. \tag{OA1-44}$$

Conditions (OA1-43) and (OA1-44) need to hold across inputs and across destinations, which implies the following:

[10]The assumption that the production function $F_{fi}(.)$ is firm-product-specific ensures the implied cost of capital R_{fit}^{k} is not destination-specific.

$$\frac{W^1_{fit}}{W^v_{fit}} = \frac{\frac{\partial F_{fi}(.)}{\partial V^1_{fi1t}}}{\frac{\partial F_{fi}(.)}{\partial V^v_{fi1t}}} = \frac{\frac{\partial F_{fi}(.)}{\partial V^1_{fi2t}}}{\frac{\partial F_{fi}(.)}{\partial V^v_{fi2t}}} = ... = \frac{\frac{\partial F_{fi}(.)}{\partial V^1_{fidt}}}{\frac{\partial F_{fi}(.)}{\partial V^v_{fidt}}} \qquad \forall v = 1, ..., V; \quad d \in D_{fit}, \tag{OA1-45}$$

$$\frac{W^v_{fit}}{R^k_{fit}} = \frac{\frac{\partial F_{fi}(.)}{\partial V^v_{f,i,1,t}}}{\frac{\partial F_{fi}(.)}{\partial K^k_{fi1t}}} = \frac{\frac{\partial F_{fi}(.)}{\partial V^v_{fi2t}}}{\frac{\partial F_{fi}(.)}{\partial K^k_{fi2t}}} = ... = \frac{\frac{\partial F_{fi}(.)}{\partial V^v_{fidt}}}{\frac{\partial F_{fi}(.)}{\partial K^k_{fidt}}} \qquad \forall v, k; \quad d \in D_{fit}. \tag{OA1-46}$$

Note that the production function is assumed to be firm-product specific and constant return to scale. Together with equations (OA1-45) and (OA1-46), these assumptions imply that the allocation of variable inputs is inversely proportional to the ratio of the productivity deflated outputs across destinations, i.e.,

$$\frac{Q_{fidt}}{\Omega_{fit}\vartheta_{fid}} = c \cdot \frac{Q_{fid't}}{\Omega_{fit}\vartheta_{fid'}} \quad \rightarrow \quad c\boldsymbol{V}^*_{fidt} = \boldsymbol{V}^*_{fid't} \quad \text{and} \quad c\boldsymbol{K}^*_{fidt} = \boldsymbol{K}^*_{fid't}. \tag{OA1-47}$$

Utilizing the relationship of (OA1-47) and the assumption that $F_{fi}(.)$ is constant return to scale, it is straightforward to see

$$\frac{\partial F_{fi}(\boldsymbol{V}^*_{fidt}, \boldsymbol{K}^*_{fidt})}{\partial V^v_{fidt}} = \frac{\partial F_{fi}(c\boldsymbol{V}^*_{fidt}, c\boldsymbol{K}^*_{fidt})}{\partial (cV^v_{fidt})} = \frac{\partial F_{fi}(\boldsymbol{V}^*_{fid't}, \boldsymbol{K}^*_{fid't})}{\partial V^v_{fid't}}. \tag{OA1-48}$$

Rearranging (OA1-43) and (OA1-48) yields:

$$\begin{aligned} \lambda_{fidt} &= \left(\frac{\Omega_{fit}\vartheta_{fid}}{W^v_{fit}} \frac{\partial F_{fi}(\boldsymbol{V}^*_{fidt}, \boldsymbol{K}^*_{fidt})}{\partial V^v_{fidt}} \right)^{-1} \\ &= \left(\frac{\Omega_{fit}\vartheta_{fid}}{W^v_{fit}} \frac{\partial F_{fi}(\boldsymbol{V}^*_{fid't}, \boldsymbol{K}^*_{fid't})}{\partial V^v_{fid't}} \right)^{-1}. \end{aligned} \tag{OA1-49}$$

Therefore, the relative marginal cost across destinations is static, depending on the relative productivity difference across destinations, i.e.,

$$\frac{\lambda_{fidt}}{\lambda_{fid't}} = \frac{\vartheta_{fid'}}{\vartheta_{fid}} \tag{OA1-50}$$

Although the marginal cost is firm-product-destination specific and time-varying, the relative marginal cost is not. Therefore, condition (OA1-35) is satisfied.

OA1.3.3 An alternative approach

An alternative approach to reconcile our work with De Loecker et al. (2016) consists of directly redefining what a product variety is in their model. Namely, if one redefines a product-destination pair as a variety, i.e., $j = \{i, d\}$, then the original setting and assumptions will go through without any change.

We argue that this approach is not very useful, for two reasons. The first one is practical. De Loecker et al. (2016) define a product variety as a two-digit industry. The need to define a product at the industry level is mainly due to data limitations. If one adopts a more refined product definition, for instance, the estimator by De Loecker et al. (2016) would suffer from a small sample problem—there would not be enough power to estimate. The small sample problem will be much more severe if one defines a product-destination pair as a variety. This is due not only to the smaller number of observations in each cell, but also to the frequent changes in the set of destinations a firm exports a product to.

The second reason is related to conceptual assumptions regarding production functions. De Loecker et al. (2016) rely on the assumption that the production function is the same for single- and multi-product firms. When redefining a product-destination pair as a variety, the identification condition would require the production function to be product-destination specific and invariant along the firm dimension. In the context of our problem, controlling for firm-product level marginal cost is the primary concern. We think that keeping the flexibility of the production function at the product level is extremely valuable.

OA2 Supplementary Model and Simulation Results

In this appendix, we examine markup elasticities estimated using data generated from an alternative model developed by Corsetti and Dedola (2005) and used in Berman et al. (2012), where variable markups arise due to the existence of local production or distribution costs. Compared to the model with Kimball (1995) preference, the key advantage of the Corsetti and Dedola (2005) setting is that it allows us to derive analytical solutions and thus make a more transparent statement about the variables that affect firms' markup and exporting decisions.

The firm's problem is given as follows:

$$\max_{P_{fidt}, \phi_{fidt} \in \{0,1\}} \quad \phi_{fidt} \left[(P_{fidt} - \mathcal{MC}_{fidt})\, \psi_i(\alpha_{fidt}, P_{fidt}, \mathcal{E}_{dt}) - \zeta_i \right]$$

$$\psi_i(\alpha_{fidt}, P_{fidt}, \mathcal{E}_{dt}) \equiv \alpha_{fidt} \left(\frac{P_{fidt}}{\mathcal{E}_{dt}} + \chi_i \right)^{-\rho_i}$$

where $\chi_i > 0$ is the local distribution cost denominated in the destination country's currency; $\rho_i > 1$ is the elasticity of substitution across varieties of product i; $\phi_{fidt} \in \{0, 1\}$ is an indicator that equals one if firm f decides to export its product i to destination d at time t; P_{fidt} is the border price denominated in the exporter's currency; $\mathcal{MC}_{fidt}$ denotes the marginal cost; α_{fidt} is a markup-irrelevant demand shifter; $\mathcal{E}_{dt}$ is the bilateral exchange rate with an increase in $\mathcal{E}_{dt}$ meaning a depreciation of the exporting country's currency; and $\psi_i(.)$ gives the demand facing firm f selling product i in destination d in time t.

The firm's optimal price denominated in the exporter's currency is given by:

$$P^*_{fidt} = \frac{\rho_i}{\rho_i - 1} \left(\mathcal{MC}_{fidt} + \frac{\chi_i}{\rho_i} \mathcal{E}_{dt} \right) \tag{OA2-1}$$

Defining the markup as $\mu_{fidt} \equiv P^*_{fidt} / \mathcal{MC}_{fidt}$, the optimal markup adjustment can be written as a function of changes in the exchange rate $\widehat{\mathcal{E}}_{dt}$ and the marginal cost $\widehat{\mathcal{MC}}_{fidt}$ (up to a first-order approximation):

$$\widehat{\mu}_{fidt} = \Gamma_{fidt} \left(\widehat{\mathcal{E}}_{dt} - \widehat{\mathcal{MC}}_{fidt} \right) \tag{OA2-2}$$

with the markup elasticity to exchange rates given by:

$$\Gamma_{fidt} \equiv \frac{\chi_i \mathcal{E}_{dt}}{\rho_i \mathcal{MC}_{fidt} + \chi_i \mathcal{E}_{dt}} \tag{OA2-3}$$

Equations (OA2-2) and (OA2-3) highlight the two key theoretical predictions of the model: (a) the markup elasticity to the exchange rate is *decreasing* in ρ_i, suggesting high differentiation goods tend to have *higher* markup adjustments relative to low differentiation goods; and (b) the markup elasticity is *increasing* in the retail cost ratio, suggesting that more productive firms—with lower marginal costs and larger market shares—tend to make *higher* markup adjustments.

The entry and exit decisions of a firm's product depend crucially on the changes in the operational profit of the firm-product in a destination market:

$$\widehat{\pi}_{fidt} = \widehat{\alpha}_{fidt} + \left(1 + \frac{\rho_i - 1}{1 + \omega_{fidt}} \right) \widehat{\mathcal{E}}_{dt} - \frac{\rho_i - 1}{1 + \omega_{fidt}} \widehat{\mathcal{MC}}_{fidt} \tag{OA2-4}$$

where $\omega_{fidt} \equiv \chi_i \mathcal{E}_{dt} / \mathcal{MC}_{fidt} > 0$ is the retail cost ratio defined as the distribution cost expressed in the producer's currency divided by the marginal cost.

Direction of potential biases. As we discussed in section 6 of the paper, the direction of the selection bias depends on how the variable of interest (i.e., $\mathcal{E}_{dt}$) and the unobserved variable (e.g., $\mathcal{MC}_{fidt}$) enter the pricing and the selection equations. First of all, equations (OA2-1) and (OA2-4) show that the exchange rate $\mathcal{E}_{dt}$ has positive impacts on the optimal price P^*_{fidt} and the operational profit π_{fidt}. Second, we can see from these two equations that a higher marginal cost increases the optimal price of the firm but reduces the operating profit, making the firm less likely to enter a market. These relationships suggest that the unobserved marginal cost will result in an upward selection bias in the estimated markup elasticity to exchange rates. Intuitively, this is because when the exchange rate is unfavourable (i.e., when $\mathcal{E}_{dt}$ is low), the marginal cost $\mathcal{MC}_{fidt}$ needs to be sufficiently low for a firm to find it optimal to export its product to a market. Therefore, selection makes us more likely to observe low (high) marginal cost firms when the exchange rate is low (high), which leads to a *positive* correlation between the unobserved marginal cost and the exchange rate in the *observed* transactions and thus results in an *upward* selection bias.

We have focused on the selection bias in the above discussions. In general, the total bias caused by the unobserved marginal cost will also depend on the correlation between the marginal cost and the exchange rate in the absence of any selection effects. For example, if the marginal cost is positively correlated with exchange rates (e.g., due to a higher cost of imported inputs), then there will be an upward omitted variable bias even if we could observe the optimal price for all firms (including those that do not find it optimal to export). In this case, the omitted variable bias and the selection bias will reinforce each other and result in a significantly larger bias.

Finally, we note that, since preference shocks $\widehat{\alpha}_{fidt}$ do not affect the optimal price of the firm (see equation (OA2-1)), omitting them in the estimation of the markup elasticity to exchange rates will not result in any selection or omitted variable bias. By the same token, since the entry cost ζ_i does not affect the optimal price, changes in the entry cost will not cause any bias.

Simulation setup. We follow the same exchange rate data-generating process as in the paper:

$$\ln(\mathcal{E}_{dt}) = \sigma_{\mathcal{E}}(v_d * \mathcal{F}_t + u_{dt}) \tag{OA2-5}$$

where changes in $\mathcal{E}_{dt}$ are driven by (i) economic fundamentals of the origin country captured by $\mathcal{F}_t$, which can have differential effects in each destination market v_d, and (ii) a noise term u_{dt} that captures exchange rate changes due to financial market fluctuations, for example. $\sigma_{\mathcal{E}}$ controls for the relative size of exchange rate shocks.

The marginal cost $\mathcal{MC}_{fidt} = M_{fidt}/A_{fi}$ is comprised of two terms, where M_{fidt} denotes shocks to the firm's marginal costs due to firm-specific or macro reasons, and A_{fi} is the productivity of the firm-product drawn from a Pareto distribution. In contrast to the simulation setting in our paper, we now allow for firm-product-destination specific cost components and shocks:

$$\ln(M_{fidt}) = \begin{cases} \sigma_M(v_{fi} * \mathcal{F}_t + u_{fit}) & \text{in panel (a)} \\ \sigma_M(v_{fi} * \mathcal{F}_t + u_{fit}) + \sigma_D \varsigma_{fid} & \text{in panel (b)} \\ \sigma_M(v_{fi} * \mathcal{F}_t + u_{fit}) + \sigma_D \varsigma_{fid}(\mathcal{F}_t + u_{fidt}) & \text{in panel (c)} \end{cases} \tag{OA2-6}$$

As we discussed in the paper, the $\sigma_M(v_{fi} * \mathcal{F}_t + u_{fit})$ term in $\ln(M_{fidt})$ captures time-varying firm-product marginal costs that are positively correlated with exchange rates. The setting in panel (b) allows for a firm-product-destination-specific cost component ς_{fid}, whereas the setting in panel (c) permits the firm-product-destination-specific cost component to be time-varying and correlated with the shocks to the economic fundamentals $\mathcal{F}_t$.

Factors $\mathcal{F}_t, u_{dt}, u_{fit}$ and u_{fidt} are independently drawn from a standard normal distribution. Firm, product and destination specific effects v_{fi}, v_d and ς_{fid} are drawn from a standard uniform distribution. We set $\sigma_{\mathcal{E}} = 0.02$, $\sigma_M = 0.05$ and $\sigma_D = 0.075$ and give more weight to firm-product specific shocks so that most of the changes in the firms' trade patterns are driven by these unobserved shocks rather than by the observed bilateral exchange rate changes. We set the local distribution cost $\chi_i = 0.5$ so that the median distribution margin is around 40-50%, roughly in line with the recent empirical estimates (see, e.g., Berger et al. (2012)). We set the fixed cost of entry ζ_i so that about 20% of firms selling each product export.

Simulation results. Tables OA2-1 and OA2-2 show the estimates under three different marginal cost processes described in (OA2-6) for the Corsetti and Dedola (2005) model discussed above and the Kimball (1995) model in section 6 of the paper, respectively.[11]

We compare the performance of our TPSFE estimator (column 7) along with six alternative approaches (columns 1-6) and the benchmark estimates from an infeasible estimator (column 8). Specifically, column (1) shows the OLS estimates from regressing $\ln(P_{fidt})$ on $\ln(\mathcal{E}_{dt})$. Column (2) shows the estimates that would have been obtained from productivity and marginal cost estimation approaches, where we add the mean marginal cost of a firm's product in a period (i.e., $\overline{\mathcal{MC}}_{fit} \equiv \frac{1}{n^D_{fit}} \sum_{d \in D_{fit}} \mathcal{MC}_{fidt}$) as an additional control variable to the OLS specification in column (1). Column (3) shows the estimates from the original Knetter (1989) approach. Column (4) shows results from the "S-difference" specification of Gopinath et al. (2010). Columns (5) and (6) report estimates using firm-product-destination + time and firm-product-time + destination fixed effects, respectively. Column (7) reports the estimates from our TPSFE estimator. Finally, in the last column (8), we report the benchmark estimates from an infeasible estimator by running an OLS regression which includes *all* the unobserved variables (e.g., the true marginal cost $\mathcal{MC}_{fidt}$) in the specification. This regression gives the best linear relationship that an econometrician could get without specifying the underlying theoretical model.

The key takeaways in panel (a) of the two tables are the same as those we discussed in section 6 of the paper: the marginal cost estimation approach (2) and the fixed effect approaches (6) and (7) give estimates that are very close to the benchmark best linear estimates. Panel (b) of both tables show that, similar to the case of adding firm-product-destination-specific demand conditions

[11] Since demand shocks do not result in any bias in the estimation of markup elasticities in Corsetti and Dedola (2005), we also shut down the markup-relevant demand shocks in the simulations of the Kimball model (by setting $\ln(D_{fidt}) = 0$) to make the simulation results of the two models more comparable. We allow for firm-product-destination-specific markup-irrelevant demand shifters α_{fid} in both models.

Table OA2-1: Comparison of Estimators – Corsetti and Dedola (2005)

Sample	(1) OLS	(2) OLS with $\overline{\mathcal{MC}}_{fit}$	(3) $d+t$ FE	(4) S-diff	(5) $fid+t$ FE	(6) $fit+d$ FE	(7) TPSFE	(8) Best Linear
Panel (a): firm-product-time cost shocks								
All	1.30	0.15	1.48	0.31	0.31	0.12	0.12	0.15
	(0.02)	(0.00)	(0.03)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
HD ($\rho = 4$)	1.45	0.20	1.45	0.38	0.38	0.17	0.20	0.20
	(0.03)	(0.00)	(0.03)	(0.01)	(0.01)	(0.00)	(0.00)	(0.00)
LD ($\rho = 12$)	1.14	0.08	1.14	0.24	0.24	0.07	0.07	0.08
	(0.02)	(0.00)	(0.03)	(0.01)	(0.01)	(0.00)	(0.00)	(0.00)
Panel (b): firm-product-time cost shocks + firm-product-destination specific cost component								
All	1.29	0.16	1.47	0.31	0.30	0.15	0.12	0.14
	(0.02)	(0.00)	(0.03)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
HD ($\rho = 4$)	1.44	0.21	1.44	0.38	0.37	0.19	0.19	0.20
	(0.03)	(0.00)	(0.03)	(0.01)	(0.01)	(0.00)	(0.00)	(0.00)
LD ($\rho = 12$)	1.14	0.11	1.14	0.24	0.24	0.10	0.07	0.08
	(0.02)	(0.00)	(0.03)	(0.01)	(0.01)	(0.00)	(0.00)	(0.00)
Panel (c): firm-product-*destination*-time cost shocks								
All	1.29	0.23	1.46	0.83	0.38	0.23	0.15	0.15
	(0.02)	(0.00)	(0.03)	(0.01)	(0.01)	(0.00)	(0.01)	(0.00)
HD ($\rho = 4$)	1.44	0.27	1.42	0.89	0.46	0.26	0.27	0.21
	(0.03)	(0.01)	(0.03)	(0.01)	(0.01)	(0.01)	(0.01)	(0.00)
LD ($\rho = 12$)	1.14	0.19	1.14	0.77	0.31	0.21	0.10	0.08
	(0.02)	(0.01)	(0.03)	(0.01)	(0.01)	(0.01)	(0.01)	(0.00)

Note: Estimates and standard errors are calculated based on the average of 10 simulations of each setting.

Table OA2-2: Comparison of Estimators – Kimball (1995)

Sample	(1) OLS	(2) OLS with $\overline{\mathcal{MC}}_{fit}$	(3) $d+t$ FE	(4) S-diff	(5) $fid+t$ FE	(6) $fit+d$ FE	(7) TPSFE	(8) Best Linear
Panel (a): firm-product-time cost shocks								
All	1.36	0.17	1.50	0.36	0.35	0.17	0.15	0.17
	(0.02)	(0.00)	(0.02)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
HD ($\rho = 4$)	1.51	0.27	1.51	0.46	0.45	0.26	0.27	0.27
	(0.02)	(0.00)	(0.02)	(0.01)	(0.01)	(0.00)	(0.00)	(0.00)
LD ($\rho = 12$)	1.21	0.09	1.21	0.26	0.26	0.09	0.09	0.09
	(0.02)	(0.00)	(0.03)	(0.01)	(0.01)	(0.00)	(0.00)	(0.00)
Panel (b): firm-product-time cost shocks + firm-product-destination specific cost component								
All	1.34	0.20	1.48	0.35	0.35	0.21	0.16	0.17
	(0.02)	(0.00)	(0.02)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
HD ($\rho = 4$)	1.49	0.29	1.49	0.45	0.45	0.29	0.27	0.27
	(0.02)	(0.00)	(0.02)	(0.01)	(0.01)	(0.00)	(0.00)	(0.00)
LD ($\rho = 12$)	1.20	0.12	1.21	0.26	0.26	0.13	0.09	0.09
	(0.02)	(0.00)	(0.03)	(0.01)	(0.01)	(0.00)	(0.00)	(0.00)
Panel (c): firm-product-*destination*-time cost shocks								
All	1.35	0.27	1.49	0.86	0.43	0.30	0.17	0.17
	(0.02)	(0.00)	(0.02)	(0.01)	(0.01)	(0.00)	(0.01)	(0.00)
HD ($\rho = 4$)	1.50	0.35	1.50	0.92	0.53	0.36	0.29	0.27
	(0.02)	(0.00)	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)	(0.00)
LD ($\rho = 12$)	1.21	0.21	1.21	0.79	0.33	0.24	0.10	0.09
	(0.02)	(0.01)	(0.03)	(0.01)	(0.01)	(0.01)	(0.01)	(0.00)

Note: Estimates and standard errors are calculated based on the average of 10 simulations of each setting.

discussed in the paper, allowing for firm-product-destination-specific cost components results in biased estimates in specifications (2) and (6). However, a key difference is that the presence of unobserved marginal cost components will result in an upward selection bias (as opposed to a downward bias in the case of markup-relevant demand shocks). As we can see from panel (b) of both tables, the estimates of specifications (2) and (6) tend to be larger than the benchmark estimates in column (8) and the difference in the estimates is larger for low differentiation goods, reflecting that the goods with a high elasticity of substitution are more sensitive to cost changes. Finally, in the very challenging case of exchange rates correlated with firm-product-destination-time cost shocks in panel (c), we see our TPSFE estimator outperforms alternative approaches and gives estimates closer to the benchmark estimates in column (8). This is particularly true for the low differentiation goods that are more sensitive to cost changes.

References

Balazsi, Laszlo, Laszlo Matyas, and Tom Wansbeek, "The Estimation of Multidimensional Fixed Effects Panel Data Models," *Econometric Reviews*, 2018, *37* (3), 212–227.

Berger, David, Jon Faust, John H. Rogers, and Kai Steverson, "Border prices and retail prices," *Journal of International Economics*, 2012, *88* (1), 62–73.

Berman, Nicolas, Philippe Martin, and Thierry Mayer, "How Do Different Exporters React to Exchange Rate Changes?," *The Quarterly Journal of Economics*, 2012, *127* (1), 437–492.

Charbonneau, Karyne B, "Multiple Fixed Effects in Binary Response Panel Data Models," *The Econometrics Journal*, 2017, *20* (3), S1–S13.

Corsetti, Giancarlo and Luca Dedola, "A Macroeconomic Model of International Price Discrimination," *Journal of International Economics*, 2005, *67* (1), 129–155.

Fernández-Val, Iván and Martin Weidner, "Individual and Time Effects in Nonlinear Panel Models with Large N, T," *Journal of Econometrics*, 2016, *192* (1), 291–312.

Gopinath, Gita, Oleg Itskhoki, and Roberto Rigobon, "Currency Choice and Exchange Rate Pass-Through," *The American Economic Review*, 2010, *100* (1), 304–336.

Heckman, James J, "Sample Selection Bias As a Specification Error," *Econometrica: Journal of the Econometric Society*, 1979, pp. 153–161.

Honoré, Bo, Francis Vella, and Marno Verbeek, "Attrition, Selection Bias and Censored Regressions," in "The Econometrics of Panel Data," Springer, 2008, pp. 385–418.

Kimball, Miles S., "The Quantitative Analytics of the Basic Neomonetarist Model," *Journal of Money, Credit and Banking*, 1995, *27* (4), 1241–1277.

Knetter, Michael M., "Price Discrimination by US and German Exporters," *The American Economic Review*, 1989, *79* (1), 198–210.

Kyriazidou, Ekaterini, "Estimation of a Panel Data Sample Selection Model," *Econometrica: Journal of the Econometric Society*, 1997, pp. 1335–1364.

Levinsohn, James and Amil Petrin, "Estimating Production Functions Using Inputs to Control for Unobservables," *The Review of Economic Studies*, 2003, *70* (2), 317–341.

Loecker, Jan De, Pinelopi K. Goldberg, Amit K. Khandelwal, and Nina Pavcnik, "Prices, Markups, and Trade Reform," *Econometrica*, 2016, *84* (2), 445–510.

Matyas, Laszlo, *The Econometrics of Multi-dimensional Panels*, Springer, 2017.

Olley, Steven and Ariel Pakes, "The Dynamics of Productivity in the Telecomunications Equipment Industry," *Econometrica*, 1996, *64*, 1263–97.

Verbeek, Marno and Theo Nijman, "Incomplete Panels and Selection Bias," in "The Econometrics of Panel Data," Springer, 1996, pp. 449–490.

Wansbeek, Tom and Arie Kapteyn, "Estimation of the Error-components Model with Incomplete Panels," *Journal of Econometrics*, 1989, *41* (3), 341–361.

Wooldridge, Jeffrey M., "On Estimating Firm-level Production Functions Using Proxy Variables to Control for Unobservables," *Economics Letters*, 2009, *104* (3), 112–114.