Online Appendix*
Productivity Effects of Internationalisation through the Domestic Supply Chain

Bruno Merlevede[†] Angelos Theodorakopoulos[‡]

May 2021

*This document provides relevant supporting material for online publication.

[†]Department of Economics, Ghent University, Ghent, Belgium; e-mail: bruno.merlevede@ugent.be

[‡]Corresponding author: University of Oxford, Oxford Martin School, 34 Broad Street, Oxford OX1 3BD, United Kingdom; e-mail: angelos.theodorakopoulos@oxfordmartin.ox.ac.uk

A GNR Estimation Procedure

This section serves as an overview of the basic steps and assumptions needed to apply the non-parametric identification procedure of GNR when estimating equation (6) under a parametric specification of the production technology $\left(f_j(\cdot)\right)$ and firm fixed effects (ϕ_i). For a detailed and complete description refer to GNR. For simplicity and without loss of generality, we disregard the industry dimension j. The estimation discussed below is directly extended by allowing the functional form of the production technology $f(\cdot)$ to also vary by industry j. Therefore, under the presence of ϕ_i, the production function in (5) can be expressed as:[1]

$$y_{it} = f\left(k_{it}, l_{it}, m_{it}; \alpha\right) + \omega_{it} + \phi_i + \varepsilon_{it} \tag{A.1}$$

This case considers the classic environment of perfect competition in both input and output markets. Capital and labour are assumed to be predetermined inputs and therefore chosen one year prior to the realisation of productivity, ω_{it}, i.e. at $t-1$. The only flexible input in the specification is material, which is assumed to freely adjust in each period (variable) and have no dynamic implications (static).

Conditional on the state variables and other firm characteristics, a firm's static profit maximisation problem yields the first order condition with respect to the flexible input, material:

$$P_t^M = P_t \frac{\partial}{\partial M_t} F\left(K_{it}, L_{it}, M_{it}; \alpha\right) \mathrm{e}_{it}^{\omega} \mathscr{E} \tag{A.2}$$

where P_t^M and P_t is the price of material and output, respectively. Under perfect competition in input and output markets, they are constant across firms within the same country-industry but can vary across time. By the time firms make their annual decisions, ex-post shocks ε_{it} are not in their information set, and thus firms create expectations over them such that: $\mathscr{E} = E\left(\mathrm{e}^{\varepsilon_{it}}\right)$.

Combining the log of (A.2) with (A.1) and re-arranging terms, we retrieve a share equation:

$$s_{it} = ln\left(\widetilde{f}\left(k_{it}, l_{it}, m_{it}; \widetilde{\alpha}\right)\right) + ln\mathscr{E} - \varepsilon_{it} \tag{A.3}$$

where s_{it} is the log of the nominal share of material and $\widetilde{f}\left(k_{it}, l_{it}, m_{it}; \widetilde{\alpha}\right) = \frac{\partial}{\partial m_{it}} f\left(k_{it}, l_{it}, m_{it}; \alpha\right)$ is the output elasticity of the flexible input, material. Note that the share equation is net of the log additive TFP term ω_{it}, inducing the transmission bias, and the firm fixed effects ϕ_i.

A.1 Step One

A Non Linear Least Squares estimation of the share equation (A.3) is applied using the Gauss-Newton algorithm to minimise the sum of squared errors. Under a Cobb-Douglas production

[1] Given the data structure, we consider a large number of firms (N) and a small number of time series observations per firm (T). Thus, we rely on typical panel data asymptotic properties as $N \to \infty$ for fixed T.

technology $\widetilde{f}(k_{it}, l_{it}, m_{it}; \widetilde{\alpha}) \mathscr{E} = \alpha_m \mathscr{E} \equiv \widetilde{\alpha}_m$, where α_m is now a constant representing the output elasticity of the flexible input material. This step identifies ε_{it} (hence $\mathscr{E} \equiv \sum_{it} \frac{\varepsilon_{it}}{NT}$) and $\widetilde{\alpha}_m$, which in turn allows us to compute $\alpha_m \equiv \widetilde{\alpha}_m / \mathscr{E}$.

A.2 Step Two

By integrating up the output elasticity of the flexible input:

$$\int \widetilde{f}(k_{it}, l_{it}, m_{it}; \widetilde{\alpha})\, dm_{it} = f(k_{it}, l_{it}, m_{it}; \alpha) + \mathscr{F}\left(k_{it}, l_{it}; \widetilde{\widetilde{\alpha}}\right) \tag{A.4}$$

we retrieve the production technology $f(\cdot)$ to the part that remains to be identified $\mathscr{F}\left(k_{it}, l_{it}; \widetilde{\widetilde{\alpha}}\right)$. By differencing it with the production function (A.1) we get the following expression for TFP:

$$\omega_{it} = \mathscr{Y}_{it} + \mathscr{F}\left(k_{it}, l_{it}; \widetilde{\widetilde{\alpha}}\right) - \phi_i \tag{A.5}$$

where $\mathscr{Y}_{it}$ is the log of the expected output net of the term (A.4) computed in Step One. Under a Cobb-Douglass production technology, $\mathscr{Y}_{it} = y_{it} - \widehat{\varepsilon}_{it} - \widehat{\alpha}_m m_{it}$ and $\mathscr{F}\left(k_{it}, l_{it}; \widetilde{\widetilde{\alpha}}\right) = \widetilde{\widetilde{\alpha}}_k k_{it} + \widetilde{\widetilde{\alpha}}_l l_{it}$, where $\widetilde{\widetilde{\alpha}}_k \equiv -\alpha_k$ and $\widetilde{\widetilde{\alpha}}_l \equiv -\alpha_l$.

To proceed, we combine the assumption over the law of motion of TFP used in baseline specification (6) with (A.5) to generate the following estimating equation:

$$\begin{aligned}
\mathscr{Y}_{it} &= -\mathscr{F}\left(k_{it}, l_{it}; \widetilde{\widetilde{\alpha}}\right) + \rho_\omega \omega_{it-1} + \rho_p proxies_{jct-1} + \rho_x X_{jct-1} + \rho_{fe} d_{fe,t} + (1-\rho_\omega)\phi_i + \xi_{it} \\
&= -\mathscr{F}\left(k_{it}, l_{it}; \widetilde{\widetilde{\alpha}}\right) + \rho_\omega \left(\mathscr{Y}_{it-1} + \mathscr{F}\left(k_{it-1}, l_{it-1}; \widetilde{\widetilde{\alpha}}\right)\right) \\
&\qquad + \rho_p proxies_{jct-1} + \rho_x X_{jct-1} + \rho_{fe} d_{fe,t} + (1-\rho_\omega)\phi_i + \xi_{it}
\end{aligned} \tag{A.6}$$

where $d_{fe,t}$ is a full set of dummies with their corresponding parameters ρ_{fe} representing the relevant time-varying fixed effects $\phi_{fe,t}$ in (6). In the absence of ϕ_i, one can readily estimate (A.6) using a Generalised Method of Moments (GMM) estimator (see GNR for more details). However, in the presence of firm fixed effects, further transformations and assumptions are necessary. We turn to this next.

A.2.1 First Difference GMM (DIF)

Following the dynamic panel literature, GNR augment their baseline estimator to account for firm fixed effects by first-differencing (A.6) such that:

$$\Delta \mathscr{Y}_{it} = -\Delta \mathscr{F}\left(k_{it}, l_{it}; \widetilde{\widetilde{\alpha}}\right) + \rho_\omega \Delta \omega_{it-1} + \rho_p \Delta proxies_{jct-1} + \rho_x \Delta X_{jct-1} + \rho_{fe} \Delta d_{fe,t} + \Delta \xi_{it} \tag{A.7}$$

where Δ is the first difference operator.[2] However, the above equation suffers from endogeneity induced by the correlation between $\Delta\omega_{it-1}$ and $\Delta\xi_{it}$. To solve for this, one could instrument with deeper lags in levels *à la* Arellano and Bond (1991). However, as discussed in section 3.3, such estimators are known to perform poorly due to weak instruments. Therefore, in the next section, we further augment the GNR estimator in the presence of firm fixed effects.

A.2.2 System GMM (SYS)

Following Blundell and Bond (1998), the SYS approach augments the DIF from the previous section by simultaneously estimating the equation in differences and levels:[3]

$$\begin{aligned}\begin{pmatrix}\Delta\mathscr{Y}_{it}\\ \mathscr{Y}_{it}\end{pmatrix} = &-\begin{pmatrix}\Delta\mathscr{F}\left(k_{it},l_{it};\tilde{\tilde{\alpha}}\right)\\ \mathscr{F}\left(k_{it},l_{it};\tilde{\tilde{\alpha}}\right)\end{pmatrix} + \rho_{\omega}\begin{pmatrix}\Delta\omega_{it-1}\\ \omega_{it-1}\end{pmatrix} + \rho_{p}\begin{pmatrix}\Delta proxies_{cjt-1}\\ proxies_{cjt-1}\end{pmatrix}\\ &+\rho_{x}\begin{pmatrix}\Delta X_{it-1}\\ X_{it-1}\end{pmatrix} + \rho_{fe}\begin{pmatrix}\Delta d_{fe,t}\\ d_{fe,t}\end{pmatrix} + \begin{pmatrix}\Delta\xi_{it}\\ \xi_{it}\end{pmatrix}\end{aligned} \tag{A.8}$$

where the same linear relationship with the same coefficients applies. This results in a stacked dataset with twice the number of firms and the same set of parameters used in levels.[4] By distinctly instrumenting each of the stacked equations, we form the $L \equiv \left(L^D + L^L\right)$ x 1 vector of stacked moment conditions:

$$E\left[m_i(\theta_o)\right] = E\left[\mathscr{Z}_i'\tilde{\xi}_i\right] = E\left[\begin{pmatrix}\mathscr{Z}_i^D & 0\\ 0 & \mathscr{Z}_i^L\end{pmatrix}'\begin{pmatrix}\Delta\xi_i\\ \xi_i\end{pmatrix}\right] = 0 \tag{A.9}$$

as a function of the K x 1 vector of unknown parameters $\theta_o = \left(\tilde{\tilde{\alpha}},\rho_{\omega},\rho_{p},\rho_{x},\rho_{fe}\right)$ with $L > K$, where $\Delta\xi_i = (\Delta\xi_{i2},\ldots,\Delta\xi_{i,T})'$, $\xi_i = (\xi_{i1},\ldots,\Delta\xi_{i,T})'$, $\mathscr{Z}_i^D$ is a $T-1$ x L^D instrument matrix used to distinctly instrument the equation in first-differences, and $\mathscr{Z}_i^L$ is a T x L^L instrument matrix used to distinctly instrument the equation in levels.

The choice of the instruments is based on the timing assumptions of the variables and thus how they correlate with the error term, i.e. predetermined, endogenous, or exogenous. On the one hand, capital and labour are assumed to be predetermined inputs chosen at time $t-1$ and are thus uncorrelated to any current or future innovations of productivity. On the other hand, the proxies and additional controls which are treated as endogenous—correlated with contemporary but not future productivity innovations—are instrumented with (deeper) lags. While we rely

[2]Standard to dynamic panel methods, linearity in the law of motion of TFP is a necessary condition for eliminating ϕ_i under first-differences.

[3]This approach requires additional stationarity restrictions on the initial conditions process (Arellano and Bover 1995).

[4]In the first-differenced equation, the industry ($\phi_j t$) and country ($\phi_c t$) specific linear time trends included in $\phi_{fe,t}$ from the levels equation now enter in $\Delta d_{fe,t}$ as a set of industry (ϕ_j) and country fixed effects (ϕ_c), respectively. In addition, the vector $\Delta d_{fe,t}$ is extended with zeros to annihilate any time-invariant terms such as the constant and, in the case of industry-j specific production technology $f_j(\cdot)$, industry-j dummies.

on the first lag only for the firm-level controls, we exploit all available lag information for the country-industry-year-level proxies to maintain maximal identifying variation. Therefore, similar to the persistence term, for the equation in first differences, $\mathscr{Z}_i^D$ contains (deeper lag) values in levels (*à la* Arellano and Bond 1991). For the equation in levels, $\mathscr{Z}_i^L$ contains (lag) values in first-differences (*à la* Blundell and Bond 1998; Arellano and Bover 1995). Note that we exclude redundant instruments by choosing the first available lag in $\mathscr{Z}_i^D$ and all available lags in $\mathscr{Z}_i^L$ (see Appendix C from Kiviet et al. 2017, for a complete description on redundant instruments).

We abstain from using additional lag lengths for the firm-level controls to avoid potential biases generated by instrument proliferation (Roodman 2009). In the same spirit, we further limit the instrument count by using a collapsed version of the instrument matrix, as suggested by Roodman (2009) among others. Kiviet et al. (2017) and Kiviet (2020) demonstrate how the combination of these two instrument reduction methods, i.e. removing long lags and collapsing, can improve estimation precision.

Finally, the time-varying fixed effects $d_{fe,t}$ are assumed to be exogenous and thus orthogonal to the productivity shocks. To extract redundant instruments, we consider $d_{fe,t}$ only in $\mathscr{Z}_i^L$. It is important to mention here that $d_{fe,t}$ includes a constant which is by default exogenous and identifies the global mean.[5] Specifically, under a Cobb-Douglas production technology:

$$\mathscr{Z}_i^D = \begin{pmatrix} k_{i1} & l_{i1} & 0 & 0 & 0 \\ k_{i2} & l_{i2} & \mathscr{Y}_{i1} & proxies_{cj1} & X_{i1} \\ k_{i3} & l_{i3} & \mathscr{Y}_{i2} & proxies_{cj2} & X_{i2} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ k_{i,T-1} & l_{i,T-1} & \mathscr{Y}_{i,T-2} & proxies_{cj,T-2} & X_{i,T-2} \end{pmatrix}$$

and

$$\mathscr{Z}_i^L = \begin{pmatrix} \Delta k_{i1} & \Delta l_{i1} & 0 & 0 & \cdots & 0 & 0 & d_{fe,1} \\ \Delta k_{i2} & \Delta l_{i2} & \Delta \mathscr{Y}_{i1} & \Delta proxies_{cj1} & & 0 & \Delta X_{i1} & d_{fe,2} \\ \Delta k_{i3} & \Delta l_{i3} & \Delta \mathscr{Y}_{i2} & \Delta proxies_{cj2} & & 0 & \Delta X_{i2} & d_{fe,3} \\ \vdots & \vdots & \vdots & \vdots & \ddots & O & \vdots & \vdots \\ \Delta k_{iT} & \Delta l_{iT} & \Delta \mathscr{Y}_{i,T-1} & \Delta proxies_{cj,T-1} & \cdots & \Delta proxies_{cj1} & \Delta X_{iT-1} & d_{fe,T} \end{pmatrix}$$

This step proceeds with a GMM estimation which uses the sample analog of the population moment conditions (A.9) to construct an estimator for θ (Hansen 1982). The GMM estimator $\hat{\theta}$

[5]In our case of industry specific production technology $f_j(\cdot)$, we replace the constant with a full set of industry-j dummies which are assumed to be exogenous.

minimises the quadratic form:

$$J(\theta) = \left(\frac{1}{N} \sum_{i=1}^{N} m_i(\theta) \right)' W \left(\frac{1}{N} \sum_{i=1}^{N} m_i(\theta) \right) \tag{A.10}$$

with respect to θ where W is an L x L positive semi-definite weighting matrix. Given that the GMM objective function is of quadratic form, we solve for the minimum using the Gauss-Newton non-linear algorithm which involves iteration to convergence for a given W (one-step). Note that for inference we rely on bootstrapping and that for the SYS moment conditions there is no simple one-step efficient W. Therefore, as a choice of a suboptimal weighting matrix to yield a consistent one-step GMM estimator, we follow Blundell and Bond (2000) in setting $W = \left(\frac{1}{N} \sum_{i=1}^{N} \mathscr{Z}_i' H_i \mathscr{Z}_i \right)^{-1}$. This contains a block diagonal matrix $H_i = diag\left(D_i D_i', I_{T-1}\right)$, where D_i is a $T-1$ x T matrix with -1s in the diagonal, 1s in the first upper sub-diagonal, and zeros elsewhere, and I_{T-1} is an identity matrix of size $T-1$.[6]

By minimising the sample analogue of the GMM criterion function, we retrieve estimates for the remaining parameters of the production technology $\left(\widehat{\widehat{\widehat{\alpha}}}\right)$ as well as the Markov process parameters and the time varying fixed effects $\left(\widehat{\rho}_\omega, \widehat{\rho}_p, \widehat{\rho}_x, \widehat{\rho}_{fe}\right)$. For a Cobb-Douglas production technology the estimated production function is:

$$y_{it} = \widehat{\alpha}_k k_{it} + \widehat{\alpha}_l l_{it} + \widehat{\alpha}_m m_{it} + \omega_{it} + \phi_i + \widetilde{\varepsilon}_{it} \tag{A.11}$$

Using the estimated parameters from this two-step procedure, i.e. $\widehat{\widehat{\alpha}}$ from step one and $\widehat{\widehat{\widehat{\alpha}}}$ from step two, we can now compute productivity $\widehat{\omega_{it} + \phi_i}$ and other relevant functionals, e.g. returns to scale $RTS = \widehat{\alpha}_k + \widehat{\alpha}_l + \widehat{\alpha}_m$.

On a technical matter, it is important to mention that, within the GMM algorithm we 'net out' the exogenous fixed effects $d_{fe,t}$ using a partitioned regression (Frisch and Waugh 1933; Lovell 1963; Giles 1984). This approach reduces the parameter space that the GMM algorithm needs to search over, which has two empirical advantages. First, it exponentially reduces estimation time given both the iterative nature of the GMM algorithm and the bootstrapping procedure used to obtain standard errors. Second, we find empirically that it helps to avoid possible non-convergence issues of the estimator related to the presence of local-minima and flat regions in the criterion function.

Implementation of the partitioning is straightforward. Within each iteration of the GMM algorithm, we use the moment conditions related to the fixed effects to retrieve Ordinary Least

[6] Alternatively, for the consistent estimation of this one-step GMM estimator, one could use the suggestion from Windmeijer (2000), where the lower-left and upper-right zero quadrants of matrix H_i are replaced by matrix D_i' and D_i, respectively. Or, more simply, one could follow Arellano and Bover (1995) and Blundell and Bond (1998) in setting $W = \left(\frac{1}{N} \sum_{i=1}^{N} \mathscr{Z}_i' \mathscr{Z}_i \right)^{-1}$. For a detailed overview on this topic, see Kiviet et al. (2017) and the Online Appendix of Kripfganz and Schwarz (2019).

Squares (OLS) estimates for ρ_{fe}. Specifically, in each iteration, for a given a set of starting values for the parameters θ other than ρ_{fe},[7] we regress the left hand side variable minus the right-hand side part of equation (A.8)—excluding the part related to the fixed effects—on the set of fixed effects. In turn, the ρ_{fe} 'OLS estimates' are used in (A.8) to complete the GMM iteration under the full set of instruments $\mathscr{Z}_i$. We repeat this approach within each iteration of the GMM algorithm until convergence is achieved. As such, we can now calculate the time-consuming cross-products and inversions of large matrices needed for the OLS outside of the iterative procedure. Also, the GMM parameter space is reduced drastically. For example, in our baseline model, from a total of 140 parameters we now need to estimate through the computationally intensive GMM only 49 parameters since the additional 91 parameters related to the fixed effects are partitioned and obtained from a computationally fast OLS regression.

[7] We use OLS estimates of equation (A.8).

B Over- and Underidentification Test Statistics

In this section we describe the construction of the relevant statistics to test the model for overidentification and underidentification. The construction of the bootstrap p-values for the test statistics is discussed next in the Online Appendix C.3.

Overidentification.—We test the validity of overidentifying moment conditions using the Hansen-J test (Hansen 1982). For the construction of the Hansen-J statistic we use $J = NJ(\widehat{\theta})$, which is computed using the consistent one-step GMM estimates $\widehat{\theta}$ and the weight matrix specified as before.

Underidentification.—We test for weak identification using the underidentification test established by Windmeijer (2021) for models with complex data structures, i.e. clustered and potentially heteroskedastic dynamic panel data models estimated by GMM. This test builds upon the simple test for weak instruments by Sanderson and Windmeijer (2016). Windmeijer (2021) highlights that the underidentification test is equivalent to an overidentification test when regressing any endogenous variable on the remaining regressors of the original model using the same set of instruments. For the case of the k^{th} proxy ($proxies^{k}_{cjt-1}$), the auxiliary model is:

$$\begin{aligned}\begin{pmatrix}\Delta proxies^{k}_{cjt-1}\\ proxies^{k}_{cjt-1}\end{pmatrix} = -\begin{pmatrix}\Delta\mathscr{F}\left(k_{it}, l_{it}; \widetilde{\widetilde{\alpha}}\right)\\ \mathscr{F}\left(k_{it}, l_{it}; \widetilde{\widetilde{\alpha}}\right)\end{pmatrix} + \rho_{\omega}\begin{pmatrix}\Delta\omega_{it-1}\\ \omega_{it-1}\end{pmatrix} + \rho^{s}_{p}\begin{pmatrix}\Delta proxies^{-k}_{cjt-1}\\ proxies^{-k}_{cjt-1}\end{pmatrix}\\ + \rho_{x}\begin{pmatrix}\Delta X_{it-1}\\ X_{it-1}\end{pmatrix} + \rho_{fe}\begin{pmatrix}\Delta d_{fe,t}\\ d_{fe,t}\end{pmatrix} + \begin{pmatrix}\Delta\xi_{it}\\ \xi_{it}\end{pmatrix}\end{aligned} \tag{B.1}$$

where the dependent variable from the original model is replaced with $proxies^{k}_{cjt-1}$. In turn, this is excluded from the set of regressors ($proxies^{-k}_{cjt-1}$). The same estimation as the original model is thus performed while keeping the instrument matrix $\mathscr{Z}_i$ unchanged. Overall, for each endogenous explanatory variable, the overidentification Hansen-J test for the relevant auxiliary model serves as an underidentification test under the null hypothesis that the model is underidentified. The test relies on the choice of the left-hand side variable, and thus can only inform whether the particular endogenous variable is poorly predicted by the instruments. The Hansen-J test statistic is computed in the same way as above.

C Bootstrap

In this section we discuss how to operationalise the cluster bootstrap procedure for our dynamic panel data model estimated from an overidentified GMM. In turn, we show how to compute cluster bootstrap standard errors and p-values for two model specification tests, i.e. overidentification and underidentification tests.

C.1 Implementation and Calculation of Standard Errors

We implement a cluster bootstrap where we first define clusters G at the industry-country level, i.e. firm-year observations can be arbitrarily correlated within but independent across clusters.[8] We form clusters at the industry-country and not at the industry-country-year level of our regressors of interest in order to ensure that—given the dynamic representation of our model—the full time-series of each firm is retained when creating the bootstrap samples below (Horowitz 2001). If anything, this choice allows for a more flexible error structure, whereby errors within the cluster can also be arbitrarily correlated over time. We then randomly draw with replacement G times over entire clusters, i.e. blocks of firms, from the original sample to generate the b^{th} bootstrap sample, where $b = 1 \dots B$.[9] We repeat this exercise for $B = 99$ times and for each parameter estimate from the original sample $\widehat{\theta}$, $\widehat{\theta}_b$ is the estimate from the b^{th} bootstrap replication and $\bar{\theta}$ is the mean of all the $\widehat{\theta}_b$s. As such, the bootstrap standard error can be written as:

$$se(\widehat{\theta}) = \left(\frac{1}{B-1} \sum_{b=1}^{B} \left(\widehat{\theta}_b - \bar{\theta} \right)^2 \right)^{1/2} \tag{C.1}$$

Calculated as such, the computed standard errors can be used for statistical inference similar to any other asymptotically valid standard errors.

C.2 Recentering

For reliable bootstrap inference and testing of the over-identified GMM estimation described in Online Appendix A.2.2, we follow Hall and Horowitz (1996) to recenter the bootstrap moment conditions.[10] Specifically, for each b^{th} bootstrap sample, the GMM estimator $\widehat{\theta}_b$ minimises the following criterion function:

$$\widetilde{J}(\theta) = \left(\frac{1}{N} \sum_{i^b=1}^{N} \left(m_{i^b}(\theta) - \frac{1}{N} \sum_{i=1}^{N} m_i(\widehat{\theta}) \right) \right)' W_b \left(\frac{1}{N} \sum_{i^b=1}^{N} \left(m_{i^b}(\theta) - \frac{1}{N} \sum_{i=1}^{N} m_i(\widehat{\theta}) \right) \right) \tag{C.2}$$

[8] We form a total of 360 clusters for this application. This comes from the fact that we have 19 industries in each of the 19 countries, and exclude CPA classification 10 in Norway (due to a lack of data).

[9] Note that we use a pairs bootstrap, i.e. draw pairs of Y (left-hand side variable), X (right-hand side variables).

[10] See Bond and Windmeijer (2005) for such an application when comparing the finite sample performance of various test procedures for a range of dynamic panel data models using GMM. Alternatively, one could follow Brown and Newey (2002) by drawing bootstrap samples under a specific weighting of the original data ensuring that the moment conditions hold.

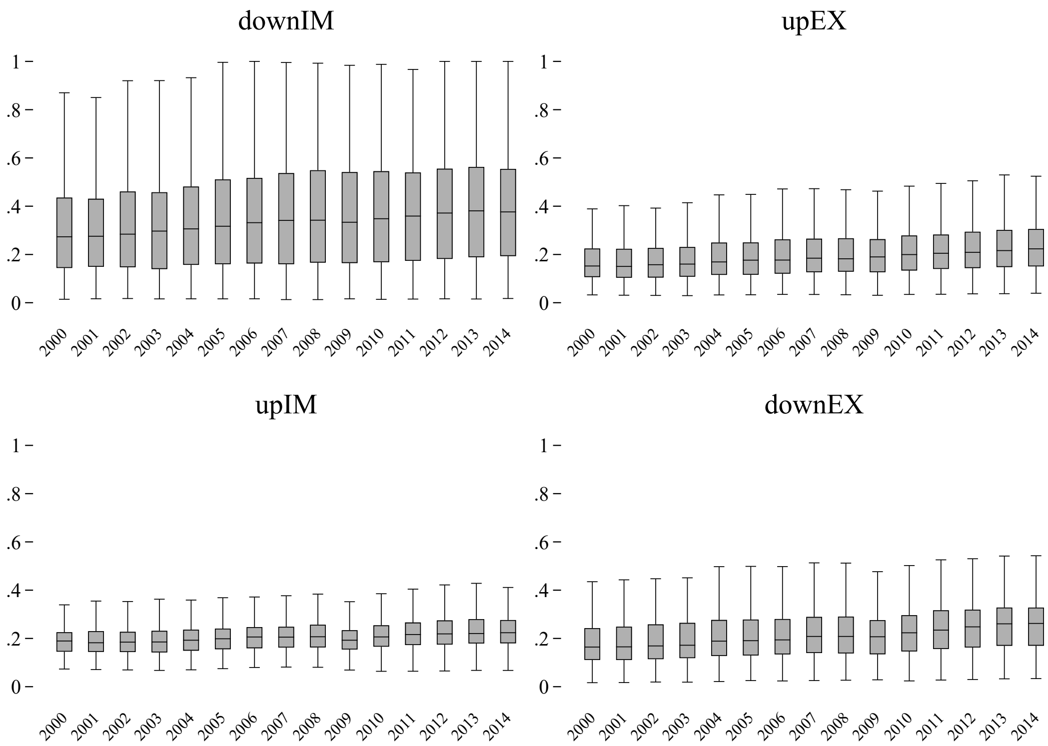
where the bootstrap moment conditions are recentered relative to the original sample moment conditions under the consistent one-step GMM estimates from the original sample $\widehat{\theta}$. W_b is constructed similarly to the weighting matrix used in (A.10) under the bootstrap sample. As such, the bootstrap version of the Hansen-J statistic is now based on $J_b = N\widetilde{J}(\widehat{\theta}_b)$.

C.3 p-values for Over- and Underidentification Tests

For each bootstrap sample b we calculate the relevant test statistic J_b and create its bootstrap empirical distribution. Recall that the underidentification test is equivalent to the overidentification test where J_b is the bootstrap Hansen-J statistic from the auxiliary model outlined in Online Appendix B. The percentile in the bootstrap distribution of J_b is then given by $p_J = \frac{1}{B}\sum_{b=1}^{B} \mathbb{1}(J_b > J)$, where the indicator function is equal to one each time the bootstrap sample statistic J_b is strictly larger than the original sample statistic J, and 0 otherwise. If $p_J < \alpha$, the test rejects the null hypothesis—at size α.

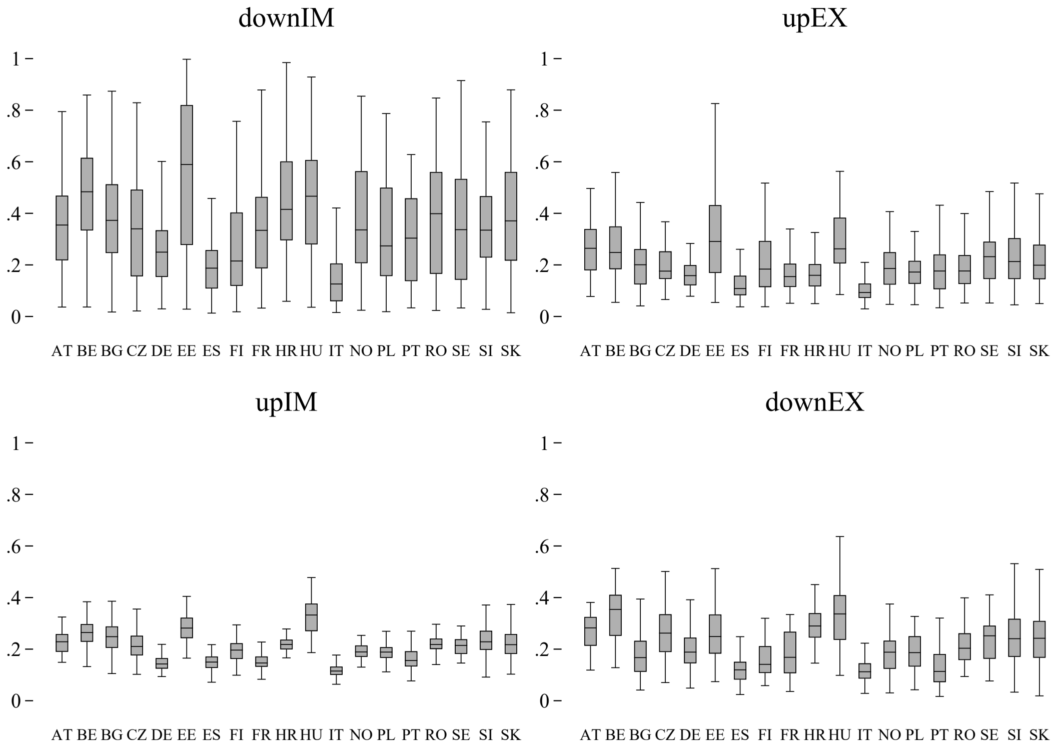
D Additional Figures and Tables

Figure D.1: Inter-industry importing and exporting by year



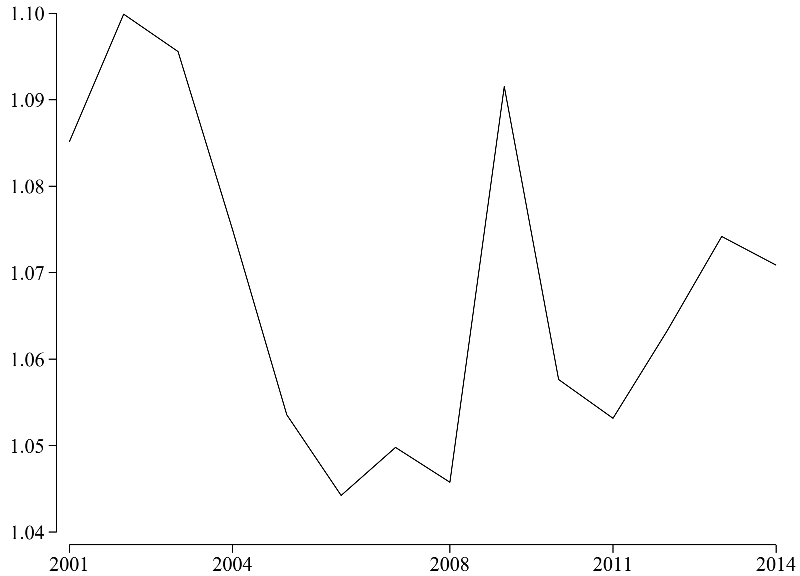
Source: Authors' calculations based on WIOT.
Notes: Let x represent the variable of interest. The upper, middle and lower hinge of the box represents the 25th ($x_{[25]}$), 50th ($x_{[50]}$) and 75th ($x_{[75]}$) percentile, respectively. Define $x_{(i)}$ as the ith ordered value of x. The upper adjacent line has a value $x_{(i)}$ such that $x_{(i)} \leq U$ and $x_{(i+1)} > U$, where $U = x_{[75]} + 1.5(x_{[75]} - x_{[25]})$. The lower adjacent line has a value $x_{(i)}$ such that $x_{(i)} \geq L$ and $x_{(i+1)} < L$, where $L = x_{[25]} - 1.5(x_{[75]} - x_{[25]})$.

Figure D.2: Inter-industry importing and exporting by country



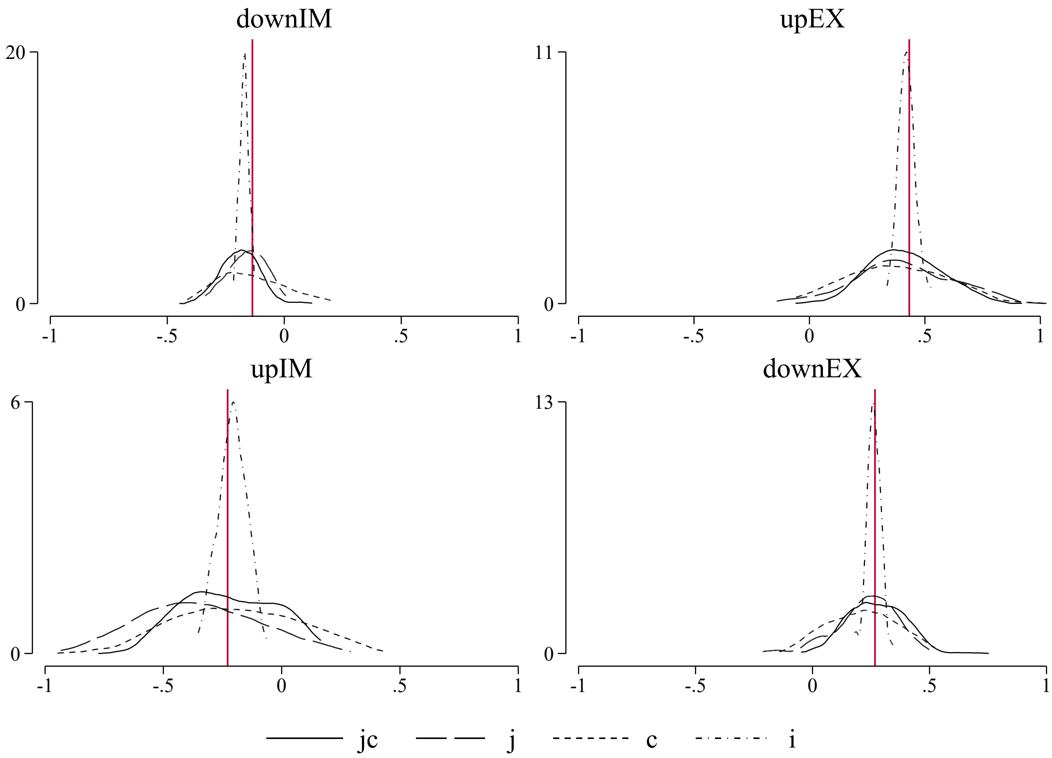
Source: Authors' calculations based on WIOT.
Notes: Let x represent the variable of interest. The upper, middle and lower hinge of the box represents the 25th ($x_{[25]}$), 50th ($x_{[50]}$) and 75th ($x_{[75]}$) percentile, respectively. Define $x_{(i)}$ as the ith ordered value of x. The upper adjacent line has a value $x_{(i)}$ such that $x_{(i)} \leq U$ and $x_{(i+1)} > U$, where $U = x_{[75]} + 1.5(x_{[75]} - x_{[25]})$. The lower adjacent line has a value $x_{(i)}$ such that $x_{(i)} \geq L$ and $x_{(i+1)} < L$, where $L = x_{[25]} - 1.5(x_{[75]} - x_{[25]})$.

Figure D.3: Markup by year



Source: Authors' calculations based BvDEP data and the estimated extension of the baseline model (6) assuming monopolistic competition in the output market and CES preferences.

Figure D.4: Distributions of bootstraped values for different clustering levels



Source: Authors' calculations based on BvDEP.
Notes: For each proxy variable of interest, the plotted distributions represent the kernel densities of the point estimates from the 99 replications of the cluster bootstrap for different types of clustering, i.e. industry-country (*jc*), industry (*j*), country (*c*), and firm (*i*). The vertical (red) line represents the point estimate of each variable from the original sample.

Table D.1: Summary Statistics for sample selection criteria

Sample Selection Criteria	(1) Observations	(2) # Firms	(3) Sales	(4) # Employees
1. Active Legal Status	99.74	99.98	99.94	99.86
2. Consolidated Accounts	95.82	96.48	88.23	82.47
3. No Missing Data	61.85	71.82	70.46	75.14
4. >20 Employees	30.66	23.70	92.43	89.29
5.i. BACON 30th percentile	93.84	94.85	78.44	88.71
5.ii. >2 consecutive observations	94.24	76.97	93.29	95.37
6.i. BACON 15th percentile	97.82	98.25	94.37	97.12
6.ii. >2 consecutive observations	94.42	77.48	89.33	95.14

Notes: This table reports the remaining percentage coverage of the firm-level sample across four different categories (columns 1-4), after applying each sample selection criterion (in each row). Each selection criterion is applied sequentially and thus the table reads from top to bottom rows. For the first two criteria, results are reported relative to the original sample. For criterion 5.i and 6.i, results are relative to the sample after the first four criteria are applied, and for criteria 5.ii and 6.ii results are relative to the sample after the application up to criterion 5.i and 6.i, respectively.

Table D.2: Output elasticities of inputs and returns to scale

CPA	Baseline				Imperfect Competition			
Industry	α_k	α_l	α_m	RTS	α_k	α_l	α_m	RTS
5	0.162	0.119	0.519	0.800	0.133	0.108	0.555	0.796
6	0.199	0.229	0.339	0.767	0.206	0.241	0.362	0.809
7	0.192	0.174	0.487	0.853	0.167	0.171	0.521	0.858
8	0.134	0.142	0.478	0.753	0.125	0.141	0.510	0.776
9	0.141	0.204	0.325	0.671	0.150	0.218	0.347	0.714
10	0.189	0.127	0.537	0.853	0.076	0.053	0.573	0.702
11	0.152	0.130	0.476	0.758	0.154	0.137	0.508	0.799
12	0.113	0.189	0.351	0.653	0.118	0.205	0.375	0.698
13	0.184	0.187	0.461	0.831	0.182	0.191	0.492	0.865
14	0.188	0.250	0.390	0.828	0.194	0.265	0.417	0.875
15	0.166	0.190	0.476	0.832	0.174	0.199	0.508	0.881
16	0.177	0.246	0.354	0.776	0.214	0.275	0.378	0.867
17	0.146	0.215	0.385	0.746	0.182	0.245	0.411	0.838
18	0.173	0.159	0.445	0.777	0.195	0.178	0.475	0.849
19	0.164	0.212	0.410	0.786	0.202	0.244	0.438	0.884
20	0.168	0.218	0.493	0.879	0.194	0.242	0.526	0.962
21	0.201	0.244	0.373	0.818	0.219	0.258	0.398	0.875
22	0.163	0.228	0.415	0.806	0.157	0.235	0.444	0.835
23	0.203	0.218	0.293	0.714	0.227	0.250	0.313	0.790
All-mean	0.172	0.200	0.416	0.788	0.182	0.213	0.445	0.839
All-median	0.173	0.212	0.410	0.786	0.194	0.241	0.438	0.849
All-st.dev.	0.018	0.044	0.066	0.041	0.030	0.056	0.070	0.048

Notes: α_k, α_l, α_m are point estimates of the output elasticities of capital, labour and material, respectively, under the baseline model (Baseline) and an extension accounting for monopolistic competition in output and CES preferences (Imperfect Competition). $RTS = \alpha_k + \alpha_l + \alpha_m$ is the returns to scale of production. The last three rows report the mean, median and standard deviation of the point estimates across all industries.

Table D.3: TFP effects from inter-industry importing and exporting under baseline specification with alternative combinations of proxies

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
$downIM_{jct-1}$	-0.136	-0.083	-0.001				0.052			
	(0.091)	(0.066)	(0.111)				(0.068)			
$upEX_{jct-1}$	0.432***	0.370**		0.389**				0.360**		
	(0.166)	(0.154)		(0.175)				(0.149)		
$upIM_{jct-1}$	-0.228	-0.086			0.205				0.348	
	(0.207)	(0.224)			(0.208)				(0.221)	
$downEX_{jct-1}$	0.267**	0.304**				0.304*				0.337**
	(0.134)	(0.126)				(0.174)				(0.142)
IM_{jct-1}	0.254**		0.408***	0.279*	0.346**	0.356**				
	(0.117)		(0.146)	(0.158)	(0.164)	(0.172)				
EX_{jct-1}	0.002		-0.069	-0.078	-0.057	-0.102				
	(0.142)		(0.177)	(0.154)	(0.145)	(0.162)				
Observations	1,018,643	1,018,643	1,018,643	1,018,643	1,018,643	1,018,643	1,018,643	1,018,643	1,018,643	1,018,643

Notes: $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$. All regressions include: the persistence term; dummies for domestic and foreign ownership links; year, industry and country fixed effects; and industry and country linear time trends. Standard errors are computed using a cluster (at the industry-country) bootstrap with 99 replications and are reported in parentheses below point estimates.

Table D.4: TFP effects from inter-industry importing and exporting under baseline specification and robustness to alternative assumptions

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
		Proxies			Size		BACON	Bootstrap		Clustering	
	Baseline	0 diagonal	Broad	Small	Large	Dummy	Trim p15	499 Reps	j-level	c-level	i-level
$downIM_{jct-1}$	-0.136	-0.123	-0.283	-0.122	-0.040	-0.112	-0.135	-0.136	-0.136*	-0.136	-0.136***
	(0.091)	(0.087)	(0.256)	(0.110)	(0.091)	(0.088)	(0.119)	(0.097)	(0.082)	(0.143)	(0.020)
$upEX_{jct-1}$	0.432***	0.415***	0.780***	0.710***	0.273*	0.492***	0.527**	0.432***	0.432*	0.432**	0.432***
	(0.166)	(0.144)	(0.262)	(0.244)	(0.148)	(0.166)	(0.226)	(0.163)	(0.224)	(0.220)	(0.039)
$upIM_{jct-1}$	-0.228	0.072	-0.403	-0.376	-0.364*	-0.251	-0.106	-0.228	-0.228	-0.228	-0.228***
	(0.207)	(0.295)	(0.288)	(0.311)	(0.207)	(0.198)	(0.285)	(0.231)	(0.268)	(0.291)	(0.061)
$downEX_{jct-1}$	0.267**	0.114	0.341*	0.153	0.168	0.255**	0.329*	0.267**	0.267*	0.267*	0.267***
	(0.134)	(0.102)	(0.186)	(0.143)	(0.110)	(0.127)	(0.175)	(0.128)	(0.138)	(0.149)	(0.030)
$D * downIM_{jct-1}$						0.152*					
						(0.092)					
$D * upEX_{jct-1}$						-0.289					
						(0.216)					
$D * upIM_{jct-1}$						-0.349					
						(0.255)					
$D * downEX_{jct-1}$						-0.121					
						(0.175)					
Observations	1,018,643	1,018,643	1,018,643	435,223	583,420	1,018,643	1,065,512	1,018,643	1,018,643	1,018,643	1,018,643

Notes: $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$. All regressions include: the persistence term; intra-industry importing and exporting; dummies for domestic and foreign ownership links; year, industry and country fixed effects; and industry and country linear time trends. D is a dummy variable equal to one when firms are classified as small-sized and zero when medium-large-sized. D is not reported since it is time invariant and thus not separately identified from the fixed effects. Column 4 regression also includes the interactions of intra-industry importing and exporting with D. Standard errors are computed using a cluster (at the industry-country—except for columns 7-9 at the industry, country, and firm, respectively) bootstrap with 99 replications (except for column 5 with 499 replications) and are reported in parentheses below point estimates.

Table D.5: Heterogeneity in TFP effects from inter-industry importing and exporting with sample splits

	(1) Baseline	(2) Domestic Links	(3) Foreign Links	(4) Relatively Downstream	(5) Relatively Upstream	(6) Relatively Low-tech	(7) Relatively High-tech	(8) Labour Intensive	(9) Capital Intensive	(10) Downstream & Low-tech	(11) Downstream & High-tech	(12) Downstream & Labour-int	(13) Downstream & Capital-int
$downIM_{jct-1}$	-0.136	-0.171	-0.013	-0.214**	-0.146	-0.140	-0.155	-0.333**	-0.229*	-0.120	-0.230*	-0.126	-0.248**
	(0.091)	(0.107)	(0.063)	(0.095)	(0.166)	(0.128)	(0.110)	(0.151)	(0.121)	(0.154)	(0.118)	(0.141)	(0.118)
$upEX_{jct-1}$	0.432***	0.556***	0.036	0.680***	-0.163	1.176***	0.044	1.031***	0.163	1.427***	0.143	1.421***	0.182
	(0.166)	(0.183)	(0.081)	(0.184)	(0.295)	(0.266)	(0.163)	(0.342)	(0.140)	(0.399)	(0.176)	(0.315)	(0.203)
$upIM_{jct-1}$	-0.228	-0.302	-0.112	-0.408	0.379	-1.144***	-0.062	-0.698*	-0.138	-1.309*	-0.390	-1.336**	-0.441
	(0.207)	(0.241)	(0.126)	(0.396)	(0.272)	(0.439)	(0.249)	(0.360)	(0.292)	(0.711)	(0.459)	(0.566)	(0.480)
$downEX_{jct-1}$	0.267**	0.236*	0.262***	0.188	0.346*	0.240	0.333**	0.281	0.177	0.268	0.149	0.217	0.082
	(0.134)	(0.123)	(0.073)	(0.155)	(0.194)	(0.173)	(0.143)	(0.185)	(0.167)	(0.205)	(0.183)	(0.224)	(0.176)
Observations	1,018,643	890,661	84,461	533,966	484,676	403,163	615,479	608,339	410,304	305,033	228,933	317,264	216,702

Notes: $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$. All regressions include: the persistence term; intra-industry importing and exporting; dummies for domestic and foreign ownership links; year, industry and country fixed effects; and industry and country linear time trends. Column 2 and 3 do not include SHH^{for}, SUB^{for} and SHH^{dom}, SUB^{dom}, respectively, as these variables are used as criteria to split the sample. Standard errors are computed using a cluster (at the industry-country) bootstrap with 99 replications and are reported in parentheses below point estimates.

Table D.6: TFP effects from inter-industry importing and exporting under baseline specification and robustness to alternative assumptions

	(1)	(2)	(3)
		D=1 if Low-tech	
	Baseline	CEEC	WEC
$downIM_{jct-1}$	-0.136	0.134	-0.083
	(0.091)	(0.161)	(0.099)
$upEX_{jct-1}$	0.432***	0.124	-0.265
	(0.166)	(0.200)	(0.172)
$upIM_{jct-1}$	-0.228	0.043	1.021***
	(0.207)	(0.275)	(0.332)
$downEX_{jct-1}$	0.267**	-0.012	0.393***
	(0.134)	(0.164)	(0.134)
$D * downIM_{jct-1}$		-0.346	-0.249
		(0.227)	(0.160)
$D * upEX_{jct-1}$		0.875**	0.610
		(0.386)	(0.437)
$D * upIM_{jct-1}$		-0.933	-0.024
		(0.587)	(0.486)
$D * downEX_{jct-1}$		0.511	-0.238
		(0.361)	(0.243)
Observations	1,018,643	277,003	741,640

Notes: $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$. All regressions include: the persistence term; intra-industry importing and exporting; dummies for domestic and foreign ownership links; year, industry and country fixed effects; and industry and country linear time trends. D is a dummy variable equal to one when firms are in Low-tech industries and zero otherwise. D is not reported since it is time invariant and thus not separately identified from the fixed effects. Column 2 regression also includes the interactions of intra-industry importing and exporting with D. CEEC refers to Cenrtal Eastern European Countries while WEC refers to Western European Countries. Standard errors are computed using a cluster (at the industry-country) bootstrap with 99 replications and are reported in parentheses below point estimates.

Table D.7: Upstreamness measure

Production Line Position	CPA	(1) *Mean EU*	(2) *Mean*
1	12	1.42	1.29
2	22	1.47	1.35
3	21	1.48	1.38
4	17	1.52	1.24
5	20	1.55	1.21
6	6	1.56	1.29
7	5	1.59	1.48
8	19	1.60	1.31
9	18	1.87	1.34
10	10	1.98	2.42
11	23	2.03	2.02
12	11	2.17	1.50
13	13	2.31	1.70
14	16	2.32	1.86
15	7	2.38	1.87
16	14	2.38	1.98
17	8	2.45	1.70
18	15	2.64	1.60
19	9	2.84	2.55

Notes: The upstreamness measures are computed as in Fally (2012) and Antràs et al. (2012) using WIOT. In column 1, we consider EU as one economy and, thus, for each industry-year we use the sum of WIOT tables across all 19 EU countries to construct the measures. In column 2, we use all available granular information to compute the measures, i.e. each industry-country-year WIOT tables separately. *Mean EU* is the per industry mean of the computed EU wide upstreamness measure across time. *Mean* is the per industry mean of the upstreamness measure across all EU countries and time. Larger values represent more upstream industries.

References

Antràs, P., D. Chor, T. Fally, and R. Hillberry (2012). Measuring the upstreamness of production and trade flows. *American Economic Review: Papers & Proceedings 102*(3), 412–416.

Arellano, M. and S. Bond (1991). Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations. *Review of Economic Studies 58*(2), 277–297.

Arellano, M. and O. Bover (1995). Another look at the instrumental variable estimation of error-components models. *Journal of Econometrics 68*(1), 29–51.

Blundell, R. and S. Bond (1998). Initial conditions and moment restrictions in dynamic panel data models. *Journal of Econometrics 87*(1), 115–143.

Blundell, R. and S. Bond (2000). GMM estimation with persistent panel data: an application to production functions. *Econometric Reviews 19*(3), 321–340.

Bond, S. and F. Windmeijer (2005). Reliable inference for GMM estimators? Finite sample properties of alternative test procedures in linear panel data models. *Econometric Reviews 24*(1), 1–37.

Brown, B. W. and W. K. Newey (2002). Generalized Method of Moments, efficient bootstrapping, and improved inference. *Journal of Business & Economic Statistics 20*(4), 507–517.

Fally, T. (2012). Production staging: Measurement and facts. *Boulder, Colorado, University of Colorado Boulder*, 155–168.

Frisch, R. and F. V. Waugh (1933). Partial time regressions as compared with individual trends. *Econometrica 1*(4), 387–401.

Giles, D. E. A. (1984). Instrumental variables regressions involving seasonal data. *Economics Letters 14*(4), 339–343.

Hall, P. and J. L. Horowitz (1996). Bootstrap critical values for tests based on generalized-method-of-moments estimators. *Econometrica 64*(4), 891–916.

Hansen, L. P. (1982). Large sample properties of Generalized Method of Moments estimators. *Econometrica 50*(4), 1029–1054.

Horowitz, J. L. (2001). Chapter 52 - The bootstrap. Volume 5 of *Handbook of Econometrics*, pp. 3159–3228. Elsevier.

Kiviet, J., M. Pleus, and R. Poldermans (2017). Accuracy and efficiency of various GMM inference techniques in dynamic micro panel data models. *Econometrics 5*(1).

Kiviet, J. F. (2020). Microeconometric dynamic panel data methods: Model specification and selection issues. *Econometrics and Statistics 13*, 16 – 45.

Kripfganz, S. and C. Schwarz (2019). Estimation of linear dynamic panel data models with time-invariant regressors. *Journal of Applied Econometrics*, 1–21.

Lovell, M. C. (1963). Seasonal adjustment of economic time series and multiple regression analysis. *Journal of the American Statistical Association 58*(304), 993–1010.

Roodman, D. (2009). A note on the theme of too many instruments. *Oxford Bulletin of Economics and Statistics 71*(1), 135–158.

Sanderson, E. and F. Windmeijer (2016). A weak instrument F-test in linear IV models with multiple endogenous variables. *Journal of Econometrics 190*(2), 212–221. Endogeneity Problems in Econometrics.

Windmeijer, F. (2000). *Efficiency Comparisons for a System GMM Estimator in Dynamic Panel Data Models*, pp. 175–184. Boston, MA: Springer US.

Windmeijer, F. (2021). Testing underidentification in linear models, with applications to dynamic panel and asset pricing models. *Journal of Econometrics*.