Technical Appendix
Cross-Section Gravity Models, PPML Estimation and the Bias Correction of the Two-Way Cluster-Robust Standard Errors

Michael Pfaffermayr

May, 2023

1 The PPML Estimator

The set of regularity assumptions are similar to those summarized in Pfaffermayr (2020) and include

1. The parameter space of α, $\Theta \subset \mathbb{R}^K$, is compact. α_0 is an interior point of Θ.

2. The system of multilateral resistances holds under the true model: $D'm(\alpha_0) - [\kappa', \theta']' = 0$, where D includes the exporter and importer dummies and $[\kappa', \theta']'$ is given, non-stochastic and of order $O(C)$.

3. $c_m < m_{ij}(\alpha) < (1 - c_m)$ for some positive constant $c_m < 0.5$ w.p. 1.

4. $Y_W = O_p(C^2)$ and independent of ε_{ij}, μ_i and v_j for all i and j.

5. η_{ij}, $i, j = 1, ..., C$ is independently distributed with $E[\eta_{ij}|z_{ij}]$ and $0 < \underline{\sigma}_\eta^2 < \sigma_{\eta,ij}^2 < \overline{\sigma}_\eta < \infty$

6. μ_i and ν_j, $i, j = 1, ..., C$ is independently distributed with $E[\mu_i|z_{ij}] = 0$ and $0 < \underline{\sigma}_\mu^2 < \sigma_{\mu,i}^2 < \overline{\sigma}_\mu < \infty$ and similarly for ν. In addition, $E[\eta_{ij}\mu_i|z_{ij}] = E[\eta_{ij}\nu_j|z_{ij}] = 0$ and $E[\mu_i\mu_{i'}] = 0$ and $E[\nu_j\nu_{j'}] = 0$ for $i \neq i'$ and $j \neq j'$.

7. Bounded support of $\varepsilon_{ij} = \eta_{ij} + \mu_i + \nu_j$ so that $m_{ij}(\alpha) + m_{ij}(\alpha)\varepsilon_{ij} > 0$ for $\alpha \in \Theta$, w.p. 1.

8. Explanatory variables: $Z \in \mathcal{Z} \subset \mathbb{R}^{C^2 \times K}$ possesses full column rank K, its elements are uniformly bounded by some constant c_z, i.e., $|z_{ij,k}| \leq c_z$ w.p. 1. All elements of Z vary at the bilateral level.

In order to derive the limit distribution of the PPML estimator for α, we define $W = [Z, D]$, $G^* = W'VM^*VW$ with $M^* = diag(e^{w'_{ij}\vartheta^*})$ and $\vartheta^* = (\alpha^{*\prime}, \phi^{*\prime})'$ with $\phi^* = (\beta^{*\prime}, \gamma^{*\prime})'$, which lies elementwise between $\widehat{\vartheta}$ and ϑ_0. For missing values one may define the selection matrix V that is derived from the identity matrix by setting the ones in the main diagonal to zero if the corresponding observation is missing. Note the typical element of the upper left block of G is bounded. $\frac{1}{C^2}|g_{kl}| \leq \frac{1}{C^2}\sum_{i=1}^{C}\sum_{j=1}^{C}|z_{ij,k}|\, v_{ij} m_{ij}(\alpha)\,|z_{ij,l}|$
$\leq \frac{1}{C^2}C^2 c_z^2(1 - c_m) = O(1)$.

Applying the mean-value theorem to the PPML-score yields

$$0 = W'V\varepsilon - G^* \begin{bmatrix} \widehat{\alpha} - \alpha_0 \\ \widehat{\phi} - \phi_0 \end{bmatrix}.$$

Defining $\widetilde{Z}^* = M^{*\frac{1}{2}}Z$, $\widetilde{D}^* = M^{*\frac{1}{2}}D$, $Q_{V\widetilde{D}^*} = I - V\widetilde{D}^*\left(\widetilde{D}^{*\prime}V\widetilde{D}^*\right)^{-1}\widetilde{D}^{*\prime}V$ and using the blocks of the partitioned inverse yields

$$
\begin{aligned}
G^{*11} &= \left(\widetilde{Z}'^{*}VQ_{V\widetilde{D}^*}V\widetilde{Z}^*\right)^{-1} \\
G^{*12} &= -G^{*11}G^{*}_{12}G^{*-1}_{22} = \left(\widetilde{Z}'^{*}VQ_{V\widetilde{D}^*}V\widetilde{Z}^*\right)^{-1}\widetilde{Z}'^{*}V\widetilde{D}^*\left(\widetilde{D}^{*\prime}V\widetilde{D}^*\right)^{-1},
\end{aligned}
$$

Thus, one can write

$$
\begin{aligned}
\tau_C\left(\widehat{\alpha}-\alpha_0\right) &= G^{*11}\tfrac{\tau_C}{C^2}C^2(Z'V - G^{*}_{12}G^{*-1}_{22}D'V)\varepsilon \\
&= G^{*11}\tfrac{\tau_C}{C^2}C^2(\widetilde{Z}'^{*}M^{*-1/2}V - \widetilde{Z}'^{*}V\widetilde{D}'^{*}\left(\widetilde{D}'^{*}V\widetilde{D}^*\right)^{-1}\widetilde{D}'^{*}M^{*-1/2}V)\varepsilon \\
&= C^2\left(\widetilde{Z}'^{*}VQ_{V\widetilde{D}^*}V\widetilde{Z}^*\right)^{-1}\tfrac{\tau_C}{C^2}\widetilde{Z}'^{*}VQ_{V\widetilde{D}^*}M^{*-1/2}\varepsilon := B(\alpha^*)^{-1}\tfrac{\tau_C}{C^2}A(\alpha^*)'\varepsilon,
\end{aligned}
$$

where $B(\alpha^*) = \frac{1}{C^2}\widetilde{Z}'^{*}VQ_{V\widetilde{D}^*}V\widetilde{Z}$ and $A(\alpha^*)' = \widetilde{Z}'^{*}VQ_{V\widetilde{D}^*}M^{*-\frac{1}{2}}$. Further, $B_0 = p\lim_{C\to\infty}B(\alpha^*)$ is assumed to exist and to be invertible, $\frac{\tau_C^2}{C^4}A_0'\Omega_\varepsilon A_0 = p\lim_{C\to\infty}\frac{\tau_C^2}{C^4} * A(\alpha^*)'\varepsilon\varepsilon'A(\alpha^*)$. τ_C is a normalization factor to be determined below.

We allow for two-way clustering of the disturbances in the exporter country and importer country dimension following Cameron, Gelbach and Miller (2011).

$$\Omega_\varepsilon = E\left[\varepsilon\varepsilon' \odot (S_x + S_m - S_d)\right]$$

$S_x = D_xD_x'$, $S_m = D_mD_m'$ and $S_d = I_{C^2}$ are selector matrices for the corresponding exporter country, importer country and the country pairs. $\odot$ denotes Hadarmards elementwise product. Plugging in the estimated residuals $\widehat{\varepsilon}$, one can use

$$\widehat{V}_\alpha = B(\widehat{\alpha})^{-1}A(\widehat{\alpha})'diag\left(\widehat{\varepsilon}\widehat{\varepsilon}' \odot \left(\tfrac{C-1}{C}S_x + \tfrac{C-1}{C}S_m - \tfrac{C^2-1}{C^2}S_d\right)\right)A(\widehat{\alpha})B(\widehat{\alpha})^{-1}$$

for inference in finite samples. The correction factors are the commonly used as small sample correction.

Proposition 3.2 of Tabord-Meehan (2019) can be applied to establish the limit distribution of $\tau_C\left(\widehat{\alpha}-\alpha_0\right) = \frac{\tau_C}{C^2}B_0^{-1}A_0\varepsilon$ when properly normalized by an appropriate factor τ_C. Note this limit distribution will be the same as that $B(\alpha^*)^{-1}\frac{\tau_C}{C^2}A(\alpha^*)'\varepsilon$. In the notation of Tabord-Meehan (2019) this proposition comprises the following assumptions. For simplicity, we first consider a single regressor and a single cluster ($\Omega_\varepsilon = E\left[\varepsilon\varepsilon' \odot S_x\right]$). In the notation of Tabord-Meehan (2019) the following conditions are imposed:

1. AF2: $\mathcal{M}^L = c\mathcal{M}^H$, $c = 1$ and $\mathcal{M}^H = C$ in the present notation, i.e., cluster size increases in sample size.

2. Condition 2.1: For $l \geq 3$, $\frac{(C^2/C)^{\frac{1}{l}}C}{\sigma_C} \to 0$ as $C \to 0$. Considering a single regressor we have $\sigma_C = Var\left[\sum_{i=1}^{C}\sum_{j=1}^{C} a_{ij}\varepsilon_{ij}\right]$ so that $A_0' = [a_{11}, ...a_{CC}]_{1\times C^2}$. This condition allows applying the limit theorem of Janson (1988) for random graphs.

3. Condition 2.6:

$$\begin{aligned}\Omega &= \lim_{C\to\infty} \frac{1}{C^{2+r}} Var\left[\sum_{i=1}^{C}\sum_{j=1}^{C} a_{ij}\varepsilon_{ij}\right] = \lim_{G\to\infty} \frac{1}{C^{2+r}} \sum_{i=1}^{C}\sum_{j=1}^{C}\sum_{i'=1}^{C}\sum_{'j=1}^{C} a_{ij}a_{i'j'}E[\varepsilon_{ij}\varepsilon_{i'j'}\ ] \\ &\leq \lim_{C\to\infty} \tfrac{C^2(2C-1)c_a}{C^{2+r}} = (2C^{1-r} - C^{-r})O(1),\end{aligned}$$

since $|a_{i'j'}a_{ij}E[\varepsilon_{ij}\varepsilon_{i'j'}\ ]| \leq c_a = O(1)$ for some positive constant c_a. Thereby $r \in [0, 1]$ and Ω is assumed be positive definite. At $r = 1$ we have

$$\lim_{C\to\infty} \frac{1}{C^3} Var\left[\sum_{i=1}^{C}\sum_{j=1}^{C} a_{ij}\varepsilon_{ij}\right] \leq (2 - C^{-1})O(1) = O(1)$$

Assuming $E\left[\varepsilon_{ij}\varepsilon_{i'j'}\ \right] = 0$ if both $i \neq i'$ and $j \neq j'$, there are $C^4 - C^2(2C-1)$ uncorrelated country pairs or $C^2(2C-1)$ are correlated ones. Hence, this condition holds at $r = 1$ (see Assumption 2.5 in Tabord-Meehan, 2019 and the corresponding remarks).

4. Assumption 3.1: The distribution of $\{(a_{ij}, m_{ij}(\alpha)\varepsilon_{ij})\}_{i,j=1}^{C}$ has bounded support.

With these assumptions Condition 2.1 (here we have $\mathcal{M}_H = C$) can be rewritten as

$$L_C = \frac{\left(\frac{C^2}{C}\right)^{1/l} C}{\Omega^{1/2}} = \underbrace{\frac{\left(\frac{C^2}{C}\right)^{1/l} C}{C^{\frac{2+r}{2}}}}_{R_1} * \underbrace{\left(\frac{1}{C^{2+r}} Var\left[\sum_{i=1}^{C}\sum_{j=1}^{C} a_{ij}\varepsilon_{ij}\right]\right)^{-1/2}}_{R_2}.$$

$R_2 = O(1)$ by Condition 2.6 and $R_1 = \frac{C^{1/l}}{C^{r/2}}$. Note $\frac{1}{l} - \frac{r}{2} = \frac{2-lr}{2l} < 0 \Leftrightarrow rl > 2$ or $r > \frac{2}{l}$. Since $l \geq 3$ is assumed, at $r = 1$ it follows that $R_1 \to 0$.

Under these assumptions Proposition 3.2 implies for scalar α that at $r = 1$ the normalization is given by $\tau_C = \left(\frac{C^2}{C^1}\right)^{\frac{1}{2}} = C^{\frac{1}{2}}$ and $C^{\frac{1}{2}}\left(\widehat{\alpha} - \alpha_0\right) \xrightarrow{d} N\left(0, V_\alpha\right)$, $V_a = \frac{1}{C^3}B_0^{-1}A_0\Omega A_0' B_0^{-1}$, where $B_0 = p\lim_{C\to\infty} \frac{1}{C^2}\widetilde{Z}'^{*}VQ_{V\widetilde{D}^*}V\widetilde{Z}$ and $\frac{1}{C^3}A_0'\Omega_\varepsilon A_0 =$

$p\lim_{C\to\infty}\frac{1}{C^3}A(\alpha^*)'\varepsilon\varepsilon' A(\alpha^*)$. For the general case define $X' = \widetilde{Z}_0' V Q_{V\widetilde{D}_0} M_0^{-\frac{1}{2}}$ so that $\frac{\tau_C}{C^2}A(\alpha_0)\varepsilon = \frac{\tau_C}{C^2}\sum_{i=1}^{C}\sum_{j=1}^{C} x_{ij}\varepsilon$, where x_{ij} is the $ij-$th column of X'. Now we apply Jansons's Theorem combined with the Cramer-Wold device.

Another DGP for the disturbance has been considered in MacKinnon, Nielsen and Webb (2021), who in turn use the results of Davezies, D'Haultfœuille and Guyonvarch (2021). They assume that the disturbances are separately exhangeable random variables. E.g., the random effects model fulfils this assumption. Define

$$\begin{aligned}
\Gamma_X &= \lim_{C\to\infty}\frac{1}{C^3}A_0'\left(E\left[\varepsilon\varepsilon'\right]\odot S_x\right)A_0 \\
\Gamma_M &= \lim_{C\to\infty}\frac{1}{C^3}A_0'\left(E\left[\varepsilon\varepsilon'\right]\odot S_m\right)A_0 \\
\Gamma_I &= \lim_{C\to\infty}\frac{1}{C^2}A_0'\left(E\left[\varepsilon\varepsilon'\right]\odot S_d\right)A_0
\end{aligned}$$

$$\begin{aligned}
V_x &= B_0^{-1}\Gamma_X B_0^{-1} \\
V_m &= B_0^{-1}\Gamma_M B_0^{-1}
\end{aligned}$$

If their Assumptions 1-6 as well as condition 16 hold true, $V_x + V_y > 0$ and setting their R to C

$$C\left(\widehat{\alpha}-\alpha_0\right)\xrightarrow{d} N\left(0, V_x + V_m\right)$$

Note the last matrix $V_I = B_0^{-1}\Gamma_I B_0^{-1}$ in Cameron, Gelbach and Miller (2011) disappears and $V_x + V_m$ is always positive definite.

2 The the PPML-residuals

The bias of the standard errors of the PPML estimator for $\widehat{\alpha}$ is best illustrated for the case of fully observed trade flows, setting $V = I_{C^2}$. To simplify the illustration of the bias of $\widehat{V}_\alpha$, we insert the true parameters into the matrices $A(\alpha)$ and $B(\alpha)$, but use the residuals $\widehat{\varepsilon}$. Moreover, W is treated as as a non-stochastic matrix, whose elements are assumed to be uniformly bounded. To simplify notation we skip the index 0 indicating the true parameters in M_0, A_0 and B_0. The residuals under of the PPML estimator using dummies are based on :

$$\widehat{\alpha}-\alpha_0 = B^{-1}\widetilde{Z}'Q_{\widetilde{D}}M^{-1/2}\varepsilon := B^{-1}A'\varepsilon, \text{ ignoring the remainder}$$

$$
\begin{aligned}
B &= \tilde{Z}'\left(I - M^{1/2}D\left(D'MD\right)^{-1}D'M^{1/2}\right)\tilde{Z} = \tilde{Z}'Q_{\widetilde{D}}\tilde{Z} \\
A &= \tilde{Z}'\left(I - M^{1/2}D\left(D'MD\right)^{-1}D'M^{1/2}\right)M^{-1/2} = \tilde{Z}'Q_{\widetilde{D}}M^{-1/2}
\end{aligned}
$$

Taylor series expansion of the residuals yields (see Davidson and MacKinnon, 1993, 162-167):

$$
\widehat{\varepsilon}_{ij} = \varepsilon_{ij} - m_{ij,0}w'_{ij}G(\vartheta_0)^{-1}W'\varepsilon + o_p\left(\underbrace{\left\|(\widehat{\vartheta} - \vartheta_0)\right\|}_{O_p(1)}\right) = \left(M_0^{1/2}Q_{\widetilde{W}}M_0^{-1/2}\varepsilon\right)_{ij} + o_p(1),
$$

where $G(\vartheta_0) = \widetilde{W}'\widetilde{W}$. Note $\left\|\widehat{\vartheta} - \vartheta_0\right\| = \left\|C^2G(\vartheta_0)^{-1}\frac{1}{C^2}W'\varepsilon\right\| \leq \left\|C^2G(\vartheta_0)^{-1}\right\|\left\|\frac{1}{C^2}W'\varepsilon\right\| = O(1)O_p(1)$. To see this observe that

$$
\left\|C^{-2}G(\vartheta)\right\|^2 = C^{-4}trace\left((\widetilde{Z}'\widetilde{Z})^2 + \widetilde{Z}'\widetilde{D}\widetilde{D}'\widetilde{Z} + \widetilde{D}'\widetilde{Z}\widetilde{Z}'\widetilde{D} + (\widetilde{D}'\widetilde{D})^2\right) = O(1).
$$

Note all four terms are positive definite and (Abadir and Magnus, 2005, p. 329)

$$
\begin{aligned}
C^{-4}trace\left((\widetilde{Z}'\widetilde{Z})^2\right) &\leq C^{-4}trace\left(\widetilde{Z}'\widetilde{Z}\right)^2 \leq C^{-4}K^2(C^2c_z^2)^2 = O(1) \\
C^{-4}trace\left(\widetilde{Z}'\widetilde{D}\widetilde{D}'\widetilde{Z}\right) &= C^{-4}trace\left(\widetilde{D}'\widetilde{Z}\widetilde{Z}'\widetilde{D}\right) = C^{-4}trace\left(\widetilde{Z}\widetilde{Z}'\widetilde{D}\widetilde{D}'\right) \\
&\leq C^{-4}trace\left(\widetilde{Z}\widetilde{Z}'\right)trace\left(\widetilde{D}\widetilde{D}'\right) \\
&= C^{-4}\left(KC^2c_z^2\right)(2C-1)C = O(1) \\
C^{-4}trace\left((\widetilde{D}'\widetilde{D})^2\right) &\leq C^{-4}trace\left(\widetilde{D}'\widetilde{D}\right)^2 = C^{-4}\left((2C-1)C)\right)^2 = O(1)
\end{aligned}
$$

Further,

$$
\left\|C^{-2}W'\varepsilon\right\|^2 = \frac{1}{C^4}\sum_{h=1}^{K+2C-1}\sum_{i=1}^{C}\sum_{j=1}^{C}\sum_{k=1}^{C}\sum_{l=1}^{C}w_{ij,h}w_{kl,h}\varepsilon_{ij}\varepsilon_{kl} = O_p(1)
$$

Note $|E[\varepsilon_{ij}\varepsilon_{kl}]| \leq c_\sigma$ is assumed and $E[\varepsilon_{ij}\varepsilon_{kl}] = 0$ if $i \neq k$ and $j \neq l$. Let w_{ij} be the $K + 2C - 1 \times 1$ vector including the ij-th column of W'. Markov's inequality implies

$$
\begin{aligned}
P\left(\left\|C^{-2}W'\varepsilon\right\|^{2}\geq\kappa\right) &\leq \frac{E[\|C^{-2}W'\varepsilon\|^{2}]}{\kappa} \\
E[\left\|C^{-2}W'\varepsilon\right\|^{2}] &= \tfrac{1}{\kappa C^{4}}\sum_{h=1}^{K+2C-1}\sum_{i=1}^{C}\sum_{j=1}^{C}\sum_{k=1}^{C}\sum_{l=1}^{C}w_{ij,h}w_{kl,h}E\left[\varepsilon_{ij}\varepsilon_{kl}\right] \\
&= \tfrac{1}{\kappa C^{4}}\sum_{h=1}^{K+2C-1}\sum_{i=1}^{C}\sum_{j=1}^{C}\sum_{l=1}^{C}w_{ij,h}w_{il,h}E\left[\varepsilon_{ij}\varepsilon_{il}\right]\ \text{...[exporter cluster]} \\
&\quad +\tfrac{1}{\kappa C^{4}}\sum_{h=1}^{K+2C-1}\sum_{i=1}^{C}\sum_{j=1}^{C}\sum_{k=1}^{C}w_{ij,h}w_{kj,h}E\left[\varepsilon_{ij}\varepsilon_{kj}\right]\text{... [importer cluster]} \\
&\quad -\tfrac{1}{\kappa C^{4}}\sum_{h=1}^{K+2C-1}\sum_{i=1}^{C}\sum_{j=1}^{C}w_{ij,h}w_{il,h}E\left[\varepsilon_{ij}\varepsilon_{ij}\right]\text{...[double counting]}
\end{aligned}
$$

$$
\begin{aligned}
E[\left\|C^{-2}W'\varepsilon\right\|^{2}] &\leq \tfrac{2}{\kappa C^{4}}\left(K+2C-1\right)C^{3}c_{\sigma}c_{w}^{2}-\tfrac{1}{\kappa C^{4}}\left(K+2C-1\right)C^{2}c_{\sigma}c_{w}^{4} \\
&= \tfrac{1}{\kappa C^{4}}\left(K+2C-1\right)\left(2C^{3}-C^{2}\right)c_{\sigma}c_{w}^{2} \\
&= \tfrac{(K+2C-1)(2C-1)}{\kappa C^{2}}c_{\sigma}c_{w}^{2}=O(1)
\end{aligned}
$$

This uses $\|w_{ij}\|=(K+2C-1)^{\frac{1}{2}}c_{w}$ for some constant c_{w}. It follows that

$$\widehat{\varepsilon}=M^{1/2}Q_{\widetilde{W}}M^{-1/2}\varepsilon++o_{p}(1)b$$

for some vector b whose elements are $O_{p}(1)$ (see Davidson and MacKinnon, 1993, p. 166).

3 The bias of the variance estimator

Consider the case of a single cluster, e.g., exporter country cluster, at true M skipping the index 0. Let $\ddot{Z}'=\widetilde{Z}'Q_{\widetilde{D}}=[\ddot{Z}_{1}',...,\ddot{Z}_{C}']$, $\ddot{Z}_{l}=\left(\widetilde{Z}'Q_{\widetilde{D}}\right)_{l,K\times C}$, $l=$

$1, ..., C$. Remember

$$
\begin{aligned}
\widehat{\alpha} - \alpha_0 &= \left(\ddot{Z}'\ddot{Z}\right)^{-1} \ddot{Z}'M^{-1/2}\varepsilon \\
\widehat{\varepsilon} &= M^{1/2}Q_{\widetilde{W}}M^{-1/2}\varepsilon + +o_p(1)b \\
V_{\widehat{\alpha}} &= \left(\ddot{Z}'\ddot{Z}\right)^{-1} \ddot{Z}'M^{-1/2}\Omega M^{-1/2}\ddot{Z}\left(\ddot{Z}'\ddot{Z}\right)^{-1} \\
&= \left(\ddot{Z}'\ddot{Z}\right)^{-1} \left(\sum_{l=1}^{C} \ddot{Z}_l'M_l^{-1/2}\Omega_l M_l^{-1/2}\ddot{Z}_l\right)\left(\ddot{Z}'\ddot{Z}\right)^{-1},
\end{aligned}
$$

where $E[\varepsilon_l\varepsilon_l'] = \Omega_{l,C\times C}$, $\Omega = diag[\Omega_l]$ is a block diagonal matrix. M_l is a $C \times C$ diagonal matrix extracted from $M = diag(M_l)$.

Now consider the vector of residuals referring to cluster l given as $\widehat{\varepsilon}_l = M_l^{1/2}Q_{W,l}M^{-1/2}\varepsilon$. $Q_{\widetilde{W},l}$ is comprised of the C rows of $Q_{\widetilde{W}}$ referring to exporter cluster l. Ignoring the remainder, we have

$$
E[\widehat{\varepsilon}\widehat{\varepsilon}'] = M^{1/2}Q_{\widetilde{W}}M^{-1/2}\Omega M^{-1/2}Q_{\widetilde{W}}M^{1/2}
$$

and

$$
\begin{aligned}
E\left[\widehat{V}_{\alpha,x}\right] &= \left(\ddot{Z}'\ddot{Z}\right)^{-1} \ddot{Z}'M^{-1/2}E[\widehat{\varepsilon}'\widehat{\varepsilon} \odot S_x]M^{-1/2}\ddot{Z}'\left(\ddot{Z}'\ddot{Z}\right)^{-1} \\
&= \left(\ddot{Z}'\ddot{Z}\right)^{-1} \ddot{Z}'M^{-1/2}\left[\left(M^{1/2}Q_{\widetilde{W}}M^{-1/2}\Omega M^{-1/2}Q_{\widetilde{W}}M^{1/2}\right) \odot S_x\right]M^{-1/2}\ddot{Z}'\left(\ddot{Z}'\ddot{Z}\right)^{-1} \\
&= \left(\ddot{Z}'\ddot{Z}\right)^{-1} \left(\sum_{l=1}^{C} \ddot{Z}_l'Q_{\widetilde{W},l}M^{-1/2}\Omega M^{-1/2}Q_{\widetilde{W},l}'\ddot{Z}_l\right)\left(\ddot{Z}'\ddot{Z}\right)^{-1},
\end{aligned}
$$

where S_x selects the appropriate blocs of Ω referring to the exporter clusters. The bias of $\widehat{V}_{\alpha,x}$ thus emerges because the elements of the sum over the exporter blocs deviate form their counterparts in $V_{\widehat{\alpha}}$

$$
\ddot{Z}_l'Q_{\widetilde{W},l}M^{-1/2}\Omega M^{-1/2}Q_{\widetilde{W},l}'\ddot{Z}_l \neq \ddot{Z}_l'M_l^{-1/2}\Omega_l M^{-1/2}\ddot{Z}_l.
$$

4 The Jackknife variance estimator $\widehat{V}_{\alpha}^{JK}$

The illustration of the Jackknife estimator follows Bell and McCaffrey (2002) considering clustering by exporter countries only. We assume fully observed data ($V = I$) and plug in the true parameters into M. Again we skip the index 0 indicating true parameters. The Jackknife variance matrix is defined as

$$\widehat{V}_{\alpha}^{JK} = \tfrac{C-1}{C} \sum_{l=1}^{C} \left(\widehat{\alpha}_{[l]} - \widehat{\alpha}\right) \left(\widehat{\alpha}_{[l]} - \widehat{\alpha}\right)',$$

where $\widehat{\alpha}_{[i]}$ is the JK-estimate that leaves out cluster l. Some authors replace $\widehat{\alpha}$ by $\frac{1}{C}\sum_{l=1}^{C} \widehat{\alpha}_{[l]}$. However, as Bell and McCaffrey (2002) argue, in simulations both methods provide similar results (see also Hansen, 2019, p. 326 for the linear model). We can write

$$\begin{aligned} \widehat{\alpha} - \alpha_0 &= \left(\ddot{Z}'\ddot{Z}\right)^{-1} \ddot{Z}' M^{-1/2} \varepsilon \\ &= \left(\ddot{Z}'\ddot{Z}\right)^{-1} \ddot{Z}' Q_{\widetilde{D}} M^{-1/2} \varepsilon \\ &= \left(\ddot{Z}'\ddot{Z}\right)^{-1} \ddot{Z}' \widetilde{\varepsilon}. \end{aligned}$$

where we define $\ddot{Z} = Q_{\widetilde{D}}\widetilde{Z}$, $\ddot{Z} = \left[\ddot{Z}_1, .., \ddot{Z}_C\right]$ and $\widetilde{\varepsilon} = Q_{\widetilde{D}} M^{-1/2}\varepsilon$. $\widehat{\alpha}_{[l]} - \alpha_0$ is given by

$$\widehat{\alpha}_{[l]} - \alpha_0 = \left(\ddot{Z}'\ddot{Z} - \ddot{Z}_l'\ddot{Z}_l\right)^{-1} \left(\ddot{Z}'\widetilde{\varepsilon} - \ddot{Z}_l'\widetilde{\varepsilon}_l\right),$$

$\ddot{Z}_l$ comprises the rows of $\ddot{Z}$ referring to exporter country l. Following Cook and Weisberg (1982 , p. 136) and using the updating formula in their Appendix A.2 one has

$$\begin{aligned} \left(\ddot{Z}'\ddot{Z} - \ddot{Z}_l'\ddot{Z}_l\right)^{-1} &= \left(\ddot{Z}'\ddot{Z}\right)^{-1} - \left(\ddot{Z}'\ddot{Z}\right)^{-1} \ddot{Z}_l' \left(I - \ddot{Z}_l \left(\ddot{Z}'\ddot{Z}\right)^{-1} \ddot{Z}_l'\right)^{-1} \ddot{Z}_l \left(\ddot{Z}'\ddot{Z}\right)^{-1} \\ &= \left(\ddot{Z}'\ddot{Z}\right)^{-1} - \left(\ddot{Z}'\ddot{Z}\right)^{-1} \ddot{Z}_l' \left(I - P_{\ddot{Z},ll}\right)^{-1} \ddot{Z}_l \left(\ddot{Z}'\ddot{Z}\right)^{-1} \end{aligned}$$

Thereby, we define $P_{\ddot{Z},ll} = \ddot{Z}_l \left(\ddot{Z}'\ddot{Z}\right)^{-1} \ddot{Z}_l'$ and assume that $\left(I - P_{\ddot{Z},ll}\right)^{-1}$ exists.

$$\begin{aligned} &\widehat{\alpha}_{[l]} - \alpha_0 \\ = &\left(\left(\ddot{Z}'\ddot{Z}\right)^{-1} + \left(\ddot{Z}'\ddot{Z}\right)^{-1} \ddot{Z}_l' \left(I - P_{\ddot{Z},ll}\right)^{-1} \ddot{Z}_l \left(\ddot{Z}'\ddot{Z}\right)^{-1}\right) \left(\ddot{Z}'\widetilde{\varepsilon} - \ddot{Z}_l\widetilde{\varepsilon}_l\right) \\ = &\underbrace{\left(\ddot{Z}'\ddot{Z}\right)^{-1} \ddot{Z}'\widetilde{\varepsilon}}_{\widehat{\alpha}-\alpha_0} \end{aligned}$$

$$- \left(\ddot{Z}'\ddot{Z}\right)^{-1} \ddot{Z}_l' \left(- \left(I - P_{\ddot{Z},ll}\right)^{-1} \ddot{Z}_l \left(\ddot{Z}'\ddot{Z}\right)^{-1} \ddot{Z}'\widetilde{\varepsilon} + \left[\underbrace{I + \left(I - P_{\ddot{Z},ll}\right)^{-1} \ddot{Z}_l \left(\ddot{Z}'\ddot{Z}\right)^{-1} \ddot{Z}_l'}_{\left(I - P_{\ddot{Z},ll}\right)^{-1}} \right] \widetilde{\varepsilon}_l \right)$$

$$= \widehat{\alpha} - \alpha_0 - \left(\ddot{Z}'\ddot{Z}\right)^{-1} \ddot{Z}_l' \left(I - P_{\ddot{Z},ll}\right)^{-1} \left(\widetilde{\varepsilon}_l - \ddot{Z}_l \left(\ddot{Z}'\ddot{Z}\right)^{-1} \ddot{Z}'\widetilde{\varepsilon} \right).$$

Note $I + \left(I - P_{\ddot{Z},ll}\right)^{-1} \ddot{Z}_{ll} \left(\ddot{Z}'\ddot{Z}\right)^{-1} \ddot{Z}_{ll} = \left(I - P_{\ddot{Z},ll}\right)^{-1}$, since

$$I + \left(I - P_{\ddot{Z},ll}\right)^{-1} P_{\ddot{Z},ll} - \left(I - P_{\ddot{Z},ll}\right)^{-1} = I + \left(I - P_{\ddot{Z},ll}\right)^{-1} \left(P_{\ddot{Z},ll} - I\right) = 0.$$

Davis (2002) shows that one can decompose the symmetric projection matrix $P_{\widetilde{W}}$ as

$$\begin{aligned}
P_{\widetilde{W}} &= P_{\widetilde{D}} + P_{Q_{\widetilde{D}}\widetilde{Z}} = \widetilde{D}(\widetilde{D}'\widetilde{D})^{-1}\widetilde{D}' + \ddot{Z}\left(\ddot{Z}'\ddot{Z}\right)^{-1}\ddot{Z}' \\
Q_{\widetilde{W}} &= (I - P_{Q_{\widetilde{D}}\widetilde{Z}})(I - P_{\widetilde{D}}) = I - P_{Q_{\widetilde{D}}\widetilde{Z}} - P_{\widetilde{D}} + P_{Q_{\widetilde{D}}\widetilde{Z}} P_{\widetilde{D}} \\
&= I - P_{Q_{\widetilde{D}}\widetilde{Z}} - P_{\widetilde{D}}
\end{aligned}$$

since $P_{Q_{\widetilde{D}}\widetilde{Z}} P_{\widetilde{D}} = Q_{\widetilde{D}}\widetilde{Z}\left(\widetilde{Z}'Q_{\widetilde{D}}\widetilde{Z}\right)^{-1}\widetilde{Z}'Q_{\widetilde{D}}P_{\widetilde{D}} = 0$. It follows that

$$\begin{aligned}
\left(I - \ddot{Z}\left(\ddot{Z}'\ddot{Z}\right)^{-1}\ddot{Z}'\right)\widetilde{\varepsilon} &= \left(I - \ddot{Z}\left(\ddot{Z}'\ddot{Z}\right)^{-1}\ddot{Z}'\right) Q_{\widetilde{D}} M^{-1/2}\varepsilon = Q_{\widetilde{W}} M^{-1/2}\varepsilon \\
\widetilde{\varepsilon}_l - \ddot{Z}_l\left(\ddot{Z}'\ddot{Z}\right)^{-1}\ddot{Z}'\widetilde{\varepsilon} &= Q_{\widetilde{W},l} M^{-1/2}\varepsilon = \left(M^{-1/2}\widehat{\varepsilon}\right)_l
\end{aligned}$$

where $Q_{\widetilde{W},l}$ denotes the C rows of $Q_{\widetilde{W}}$ that correspond to cluster l. Hence,

$$\widehat{\alpha}_{[l]} - \widehat{\alpha} = -\left(\ddot{Z}'\ddot{Z}\right)^{-1}\ddot{Z}'\left(I - P_{\ddot{Z},ll}\right)^{-1}\left(M^{-1/2}\widehat{\varepsilon}\right)_l$$

Inserting yields

$$\widehat{V}_{\alpha,x}^{JK} = \tfrac{C-1}{C}\left(\ddot{Z}'\ddot{Z}\right)^{-1}\sum_{l=1}^{C}\left(\ddot{Z}_l'\left(I - P_{\ddot{Z},ll}\right)^{-1}\left(M^{-1/2}\widehat{\varepsilon}\right)_l\left(M^{-1/2}\widehat{\varepsilon}\right)_l'\left(I - P_{\ddot{Z},ll}\right)^{-1}\ddot{Z}_l\right)\left(\ddot{Z}'\ddot{Z}\right)^{-1}$$

Comparing $V_{\alpha,x}$ and $\widehat{V}_{\alpha,x}$ and using $M^{1/2}\widehat{\varepsilon} = Q_{\widetilde{W}}M^{-1/2}\varepsilon$ illustrates that the bias originates from the difference

$$E\left[\ddot{Z}'M^{-\frac{1}{2}}\varepsilon\varepsilon'M^{-\frac{1}{2}}\ddot{Z}\right] - E\left[\ddot{Z}'\underbrace{Q_{\widetilde{W}}M^{-1/2}\varepsilon\varepsilon'M^{-\frac{1}{2}}Q_{\widetilde{W}}}_{M^{1/2}\widehat{\varepsilon}}\ddot{Z}\right]$$

Following Niccodemi and Wansbeek (2022, p. 3) and Bell and McCaffrey (2002) assume that $\ddot{Z}'\ddot{Z} = C\ddot{Z}_l'\ddot{Z}_l$ and $Var[\varepsilon] = \sigma^2 M$, setting $\sigma^2 = 1$ for without loss generality. It follows that

(i)

$$\tfrac{C-1}{C}\sum_{l=1}^{C}\left(\widehat{\alpha}_{[l]} - \widehat{\alpha}\right)\left(\widehat{\alpha}_{[l]} - \widehat{\alpha}\right)'$$

$$= \tfrac{C-1}{C}\left(\ddot{Z}'\ddot{Z}\right)^{-1}\left[\sum_{l=1}^{C}\ddot{Z}_l'\left(I - P_{\ddot{Z},ll}\right)^{-1}\left(\widetilde{\varepsilon}_l - \ddot{Z}_l\left(\ddot{Z}'\ddot{Z}\right)^{-1}\ddot{Z}'\widetilde{\varepsilon}\right)_l * \right.$$
$$\left.\left(\widetilde{\varepsilon}_l - \ddot{Z}_l\left(\ddot{Z}'\ddot{Z}\right)^{-1}\ddot{Z}'\widetilde{\varepsilon}\right)_l'\left(I - P_{\ddot{Z},ll}\right)^{-1}\ddot{Z}_l\right]\left(\ddot{Z}'\ddot{Z}\right)^{-1}.$$

(ii)

$$\left(\widetilde{\varepsilon}_l - \ddot{Z}_l\left(\ddot{Z}'\ddot{Z}\right)^{-1}\ddot{Z}'\widetilde{\varepsilon}\right)_l =$$
$$\left(\underbrace{[0, ..., I_C, ..., 0]}_{E_l} - \ddot{Z}_l\left(\ddot{Z}'\ddot{Z}\right)^{-1}\ddot{Z}'\right)Q_{\widetilde{D}}M^{-1/2}\varepsilon = \left(Q_{\widetilde{D},l} - \ddot{Z}_l\left(\ddot{Z}'\ddot{Z}\right)^{-1}\ddot{Z}'\right)M^{-1/2}\varepsilon,$$

since $\ddot{Z}'Q_{\widetilde{D}} = \ddot{Z}'$. It follows that

$$
\begin{aligned}
& E\left[\left(E_l - \ddot{Z}_l\left(\ddot{Z}'\ddot{Z}\right)^{-1}\ddot{Z}'\right)Q_{\widetilde{D}}M^{-1/2}\varepsilon\varepsilon' M^{-1/2}Q_{\widetilde{D}}\left(E_l - \ddot{Z}_l\left(\ddot{Z}'\ddot{Z}\right)^{-1}\ddot{Z}'\right)'\right] \\
& \left(E_l - \ddot{Z}_l\left(\ddot{Z}'\ddot{Z}\right)^{-1}\ddot{Z}'\right)Q_{\widetilde{D}}\left(E_l - \ddot{Z}_l\left(\ddot{Z}'\ddot{Z}\right)^{-1}\ddot{Z}'\right)' \\
= \; & E_l Q_{\widetilde{D}} E_l' - \ddot{Z}_l\left(\ddot{Z}'\ddot{Z}\right)^{-1}\ddot{Z}'Q_{\widetilde{D}}E_l - E_l Q_{\widetilde{D}}\ddot{Z}\left(\ddot{Z}'\ddot{Z}\right)^{-1}\ddot{Z}_l' \\
& + \ddot{Z}_l\left(\ddot{Z}'\ddot{Z}\right)^{-1}\ddot{Z}'Q_{\widetilde{D}}\ddot{Z}\left(\ddot{Z}'\ddot{Z}\right)^{-1}\ddot{Z}_l' \\
= \; & I_C - \ddot{Z}_l\left(\ddot{Z}'\ddot{Z}\right)^{-1}\ddot{Z}_l' - \widetilde{D}_l(\widetilde{D}\widetilde{D})^{-1}\widetilde{D}_l'.
\end{aligned}
$$

(iii) Consider

$$
\left(I_C - P_{\ddot{Z},ll}\right)^{-1}\left(I_C - \ddot{Z}_l\left(\ddot{Z}'\ddot{Z}\right)^{-1}\ddot{Z}_l' - \widetilde{D}_l(\widetilde{D}\widetilde{D})^{-1}\widetilde{D}_l'\right)\left(I - P_{\ddot{Z},ll}\right)^{-1}.
$$

Note

$$
\begin{aligned}
\left(I_C - P_{\ddot{Z},ll}\right)^{-1} &= I_C + \ddot{Z}_l\left(\ddot{Z}'\ddot{Z} - \ddot{Z}_l'\ddot{Z}_l'\right)^{-1}\ddot{Z}_l' \\
&= I_C + \tfrac{1}{C-1}\ddot{Z}_l\left(\ddot{Z}_l'\ddot{Z}_l'\right)^{-1}\ddot{Z}_l'.
\end{aligned}
$$

$$
\begin{aligned}
\ddot{Z}_l &= Q_{\widetilde{D},l}Z \\
\ddot{Z}_l'\widetilde{D}_l &= Z'Q_{\widetilde{D},l}'\widetilde{D}_l \\
&= Z'\left[E_l' - \widetilde{D}(\widetilde{D}'\widetilde{D})^{-1}\widetilde{D}_l'\right]_{C^2\times C}\widetilde{D}_{l,C\times C} \\
&= Z'\left[E_l'\widetilde{D}_l - \widetilde{D}(\widetilde{D}'\widetilde{D})^{-1}\widetilde{D}_l'\widetilde{D}_l\right]
\end{aligned}
$$

$$\widetilde{D}_{l,C\times C} = \left[\begin{array}{ccccc} 0_{C\times 1} & ... & b_{l,C\times 1} & ... & 0 \end{array}\right], \ b_l' b_l = \theta_l = \sum_{j=1}^{C} m_{lj}$$

$$E_l' \widetilde{D}_l - \widetilde{D}(\widetilde{D}'\widetilde{D})^{-1}\widetilde{D}_l'\widetilde{D}_l$$

$$= \left[\begin{array}{c} 0_{C\times C} \\ ... \\ \widetilde{D}_l \\ ... \\ 0_{C\times C} \end{array}\right] - \left[\begin{array}{c} \widetilde{D}_1 \\ ... \\ \widetilde{D}_l \\ ... \\ \widetilde{D}_l \end{array}\right]_{C^2\times C} \left[\begin{array}{ccccc} \theta_1^{-1} & 0 & ... & & \\ .. & .. & & & \\ & & \theta_l^{-1} & & \\ & & & ... & \\ & & & & \theta_C^{-1} \end{array}\right] \left[\begin{array}{ccccc} 0 & 0 & ... & & \\ .. & .. & & & \\ & & \theta_l & & \\ & & & ... & \\ & & & & 0 \end{array}\right]$$

$$= \left[\begin{array}{c} 0_{C\times C} \\ ... \\ \widetilde{D}_l \\ ... \\ 0_{C\times C} \end{array}\right] - \left[\begin{array}{c} \widetilde{D}_1 \\ ... \\ \widetilde{D}_l \\ ... \\ \widetilde{D}_l \end{array}\right]_{C^2\times C} \left[\begin{array}{ccccc} 0 & 0 & ... & & \\ .. & .. & & & \\ & & 1 & & \\ & & & ... & \\ & & & & 0 \end{array}\right]_{C\times C}$$

$$= \left[\begin{array}{c} 0_{C\times C} \\ ... \\ \widetilde{D}_l \\ ... \\ 0_{C\times C} \end{array}\right] - \left[\begin{array}{c} \widetilde{D}_{1,C\times C} \\ ... \\ \widetilde{D}_l \\ ... \\ \widetilde{D}_l \end{array}\right]_{C^2\times C} \left[\begin{array}{ccccc} 0 & 0 & ... & & \\ .. & .. & & & \\ & & 1_{l,l} & & \\ & & & ... & \\ & & & & 0 \end{array}\right]_{C\times C}$$

$k \neq l$

$$\widetilde{D}_{k,C\times C} \left[\begin{array}{ccccc} 0 & 0 & ... & & \\ .. & .. & & & \\ & & 1_{l,l} & & \\ & & & ... & \\ & & & & 0 \end{array}\right]_{C\times C}$$

$$= \left[\begin{array}{ccccc} 0_{C\times 1} & ... & b_{k,C\times 1} & ... & 0 \end{array}\right] \left[\begin{array}{ccccc} 0 & 0 & ... & & \\ .. & .. & & & \\ & & 1_{l,l} & & \\ & & & ... & \\ & & & & 0 \end{array}\right] = 0_{C\times C}$$

$k = l$

$$
\begin{aligned}
&= \begin{bmatrix} 0_{C\times 1} & ... & b_{l,C\times 1} & ... & 0 \end{bmatrix} \begin{bmatrix} 0 & 0 & ... & & \\ .. & .. & & & \\ & & 1_{l,l} & & \\ & & & ... & \\ & & & & 0 \end{bmatrix} \\
&= \begin{bmatrix} 0_{C\times 1} & ... & b_{l,C\times 1} & ... & 0 \end{bmatrix} = \widetilde{D}_{l,C\times C}
\end{aligned}
$$

It follows that

$$E_l'\widetilde{D}_l - \widetilde{D}(\widetilde{D}'\widetilde{D})^{-1}\widetilde{D}_l'\widetilde{D}_l = 0$$

and

$$\ddot{Z}_l'\widetilde{D}_l = Z'\left[E_l'\widetilde{D}_l - \widetilde{D}(\widetilde{D}'\widetilde{D})^{-1}\widetilde{D}_l'\widetilde{D}_l\right] = 0_{k\times C}.$$

(iv) Therefore, we have

$$\left(I_C + \tfrac{1}{C-1}\ddot{Z}_l\left(\ddot{Z}_l'\ddot{Z}_l'\right)^{-1}\ddot{Z}_l'\right)\widetilde{D}_l(\widetilde{D}\widetilde{D})^{-1}\widetilde{D}_l'\left(I_C + \tfrac{1}{C-1}\ddot{Z}_l\left(\ddot{Z}_l'\ddot{Z}_l'\right)^{-1}\ddot{Z}_l'\right)' = \widetilde{D}_l(\widetilde{D}\widetilde{D})^{-1}\widetilde{D}_l',$$

observing that $\ddot{Z}_l'\widetilde{D}_l = 0_{k\times C}$, and

$$
\begin{aligned}
&\left(I_C - P_{\ddot{Z},ll}\right)^{-1}\left(I_C - \ddot{Z}_l\left(\ddot{Z}'\ddot{Z}\right)^{-1}\ddot{Z}_l' - \widetilde{D}_l(\widetilde{D}\widetilde{D})^{-1}\widetilde{D}_l'\right)\left(I - P_{\ddot{Z},ll}\right)^{-1} \\
= &\left(I_C + \tfrac{1}{C-1}\ddot{Z}_l\left(\ddot{Z}_l'\ddot{Z}_l'\right)^{-1}\ddot{Z}_l'\right) - \widetilde{D}_l(\widetilde{D}\widetilde{D})^{-1}\widetilde{D}_l'
\end{aligned}
$$

using

$$
\begin{aligned}
\left(I_C - P_{\ddot{Z},ll}\right)^{-1}\widetilde{D}_l(\widetilde{D}\widetilde{D})^{-1}\widetilde{D}_l' &= \widetilde{D}_l(\widetilde{D}\widetilde{D})^{-1}\widetilde{D}_l' + \tfrac{1}{C-1}\ddot{Z}_l\left(\ddot{Z}_l'\ddot{Z}_l'\right)^{-1}\ddot{Z}_l'\widetilde{D}_l(\widetilde{D}\widetilde{D})^{-1}\widetilde{D}_l' \\
&= \widetilde{D}_l(\widetilde{D}\widetilde{D})^{-1}\widetilde{D}_l'.
\end{aligned}
$$

(v) Since

$$
\begin{aligned}
&\ddot{Z}_l'\left(I_C + \tfrac{1}{C-1}\ddot{Z}_l\left(\ddot{Z}_l'\ddot{Z}_l'\right)^{-1}\ddot{Z}_l'\right)\ddot{Z}_l \\
= &\ddot{Z}_l'\ddot{Z}_l\left(\tfrac{C-1+1}{C-1}\right) = \ddot{Z}_l'\ddot{Z}_l\left(\tfrac{C}{C-1}\right)
\end{aligned}
$$

and

$$\sum_{l=1}^{C}\ddot{Z}_l'\widetilde{D}_l(\widetilde{D}\widetilde{D})^{-1}\widetilde{D}_l'\ddot{Z}_l = 0$$

it follows that

$$\frac{C-1}{C}\left(\ddot{Z}'\ddot{Z}\right)^{-1}E\left[\sum_{l=1}^{C}\ddot{Z}_l'\left(\left(I_C+\frac{1}{C-1}\ddot{Z}_l\left(\ddot{Z}_l'\ddot{Z}_l'\right)^{-1}\ddot{Z}_l'\right)-\widetilde{D}_l(\widetilde{D}\widetilde{D})^{-1}\widetilde{D}_l'\right)\ddot{Z}_l\right]\left(\ddot{Z}'\ddot{Z}\right)^{-1}$$

$$=\frac{C-1}{C}\left(\ddot{Z}'\ddot{Z}\right)^{-1}\sum_{l=1}^{C}\ddot{Z}_l'\ddot{Z}_l\left(\frac{C-1+1}{C-1}\right)\left(\ddot{Z}'\ddot{Z}\right)^{-1}=\left(\ddot{Z}'\ddot{Z}\right)^{-1}.$$

Hence, under the restrictive assumptions that (i) $Var[\varepsilon]=\sigma^2 M$ and $\ddot{Z}'\ddot{Z}=C\ddot{Z}_l'\ddot{Z}_l$ the jackknife-estimator is free of bias, resembling Theorem 2 of Bell and McCaffrey (2002) for the linear regression case. The results is also related to Theorem 2 of Pustejovsky and Tipton (2018).

4.1 The bias correction of Bell and McAffrey (2002)

Bell and McAffrey (2002) correct the PPML-residuals with a matrix $F=diag(F_l)$ so that.

$$\begin{aligned}\widehat{V}_\alpha^{BM} &= \left(\ddot{Z}'\ddot{Z}\right)^{-1}\ddot{Z}FM^{-1/2}\left(\widehat{\varepsilon}\widehat{\varepsilon}'\odot S_x\right)M^{-1/2}F'\ddot{Z}'\left(\ddot{Z}'\ddot{Z}\right)^{-1}\\ E[\widehat{V}_\alpha^{BM}] &= V_\alpha\end{aligned}$$

Remember $M^{-1/2}\widehat{\varepsilon}=Q_{\widetilde{W}}M^{-1/2}\varepsilon$. The correction matrix $F=diag(F_l)$ has to fulfil the following equation to guarantee unbiasedness under the working variance assumption:

$$E[AM^{-1/2}\left(\widehat{\varepsilon}\widehat{\varepsilon}'\odot S_x\right)M^{-1/2}F']=F\left(Q_{\widetilde{W}}M^{-1/2}\Omega M^{-1/2}Q_{\widetilde{W}}\right)F'=M^{-1/2}\Omega M^{-1/2}$$

This expression can be rewritten as

$$F_lQ_{\widetilde{W},l}M^{-1/2}\Omega M^{-1/2}Q_{\widetilde{W},l}'F_l'=M_l^{-1/2}\Omega_l M_l^{-1/2}$$

Under $\Omega=diag(\kappa_l M_l)$ as considered by Bell and McAffrey (2002) this condition reduces to

$$F_lQ_{\widetilde{W},ll}F_l'=I_C$$

where $Q_{\widetilde{W},ll}=Q_{\widetilde{W},l}Q_{\widetilde{W},l}'$. However, in this case clustering is not necessary and the correction reduces to that commonly used for heteroskedastic standard errors.

A solution exists, if $Q_{\widetilde{W},ll}$ is invertible.

$$F_lQ_{\widetilde{W},ll}F_l'=I_C \text{ or } F_lQ_{\widetilde{W},ll}F_l'F_l=F_l\rightarrow Q_{\widetilde{W},ll}^{-1/2}.$$

Consider the eigenvalue decomposition $Q_{\widetilde{W},ll} = P_l \Lambda_l P_l'$ so that $F_l = P_l \Lambda_l^{-1/2} P_l'$.

$$F_l Q_{\widetilde{W},ll} F_l' = P_l \Lambda_l^{-1/2} P_l' P_l \Lambda P_l' P_l \Lambda^{-1/2} P_l' = P_l P_l' = I_C$$

Hence, if $Q_{\widetilde{W},ll}$ is invertible we have

$$\widehat{V}_\alpha^{BM} = \left(\ddot{Z}'\ddot{Z}\right)^{-1} \left(\sum_{l=1}^{C} \ddot{Z}_l' Q_{\widetilde{W},ll}^{-1/2} M_l^{-1/2} \widehat{\varepsilon}_l \widehat{\varepsilon}_l' M_l^{-1/2} Q_{\widetilde{W},ll}^{-1/2} \ddot{Z}_l\right)$$

(see Imbens and Kolesar, 2016, p. 709 and Pustejovsky and Tipton 2018, p. 675). However, there is evidence of cases where $Q_{\widetilde{W},ll}^{-1/2}$ is singular. In the gravity context, this will be the case if variables with unilateral variation such as the exporter and importer country dummies are included.

5 Bias correction and the BRL criterion

Following Pustejovsky and Tipton (2018), we want to find matrices F_l such that

$$E\left[\widehat{V}_{\alpha,x}^{PT}\right] = \left(\ddot{Z}'\ddot{Z}\right)^{-1} \left(\sum_{l=1}^{C} \ddot{Z}_l' F_l Q_{\widetilde{W},l} M_l^{-1/2} \Omega M_l^{-1/2} Q_{\widetilde{W},l}' F_l' \ddot{Z}_l\right) \left(\ddot{Z}'\ddot{Z}\right)^{-1} = V_{\alpha,x}.$$

So for each exporter l, F_l solves

$$\begin{aligned} & E\left[\ddot{Z}_l' F_l \left(M_l^{-1/2} \widehat{\varepsilon}\widehat{\varepsilon}' M_l^{-1/2} \odot S_x\right) F_l' \ddot{Z}_l\right] \qquad (1) \\ = \; & \ddot{Z}_l' F_l Q_{\widetilde{W},l} M^{-1/2} \left(\Omega \odot S_x\right) M^{-1/2} Q_{\widetilde{W},l}' F_l' \ddot{Z}_l \\ = \; & \ddot{Z}_l' M_l^{-1/2} \Omega_l M_l^{-1/2} \ddot{Z}_l \end{aligned}$$

Theorem 1 Pustejovsky and Tipton (2018) shows that

$$F_l = T_l \left(G_l^{+}\right)^{1/2} T_l' \text{ with } G_l = T_l' Q_{W,l} \Omega Q_{W,l}' T_l$$

solves (1). $\Omega_l = T_l T_l'$ is the Cholesky factorization with T_l a lower triangular matrix. The proof uses the eigenvalue decomposition $G_l = V_l \Lambda_l V_l'$ and $G_l^{+1/2} = V_l \Lambda_l^{1/2} V'$. V_l is the matrix of eigenvectors with those eigenvectors skipped that refer to zero eigenvalues. Λ_l is the diagonal matrix with the corresponding eigenvalues.

It follows that

$$
\begin{aligned}
& \ddot{Z}_l' F_l Q_{\widetilde{W},l} M^{-1/2} \Omega M^{-1/2} Q_{\widetilde{W},l}' F_l' \ddot{Z}_l \\
= & \ddot{Z}_l' T_l \left(G_l^{+}\right)^{1/2} \left(\underbrace{T_l' Q_{\widetilde{W},l} M^{-1/2} \Omega M^{-1/2} Q_{\widetilde{W},l}' T_l}_{G_l} \right) \left(G_l^{+}\right)^{1/2} T_l' \ddot{Z}_l \\
= & \ddot{Z}_l' T_l V_l \Lambda_l^{-1/2} V_l' V_l \Lambda_l V_l' V_l \Lambda_l^{-1/2} V_l' T_l' \ddot{Z}_l \\
= & \ddot{Z}_l' T_l V_l V_l' T_l' \ddot{Z}_l
\end{aligned}
$$

Their Theorem 1 shows that

$$
\ddot{Z}' T_l V_l V_l' T_l' \ddot{Z} = \ddot{Z}_l' M_l^{-1/2} \Omega_l M_l^{-1/2} \ddot{Z},
$$

since $\ddot{Z}' T_l$ is in the column space of $Q_{\widetilde{W},l}$ that is spanned by the eigenvectors V_l. Thus we have

$$
\begin{aligned}
E\left[\widehat{V}_{\alpha,x}^{PT}\right] &= \left(\ddot{Z}'\ddot{Z}\right)^{-1} \left(\sum_{l=1}^{C} \ddot{Z}_l' F_l \left(M_l^{-1/2} E\left[\widehat{\varepsilon}\widehat{\varepsilon}'\right] M_l^{-1/2} \odot S_x \right) F_l' \ddot{Z}_l \right) \left(\ddot{Z}'\ddot{Z}\right)^{-1} \\
&= \left(\ddot{Z}'\ddot{Z}\right)^{-1} \left(\sum_{l=1}^{C} \ddot{Z}_l' M_l^{-1/2} \Omega_l M_l^{-1/2} \ddot{Z}_l \right) \left(\ddot{Z}'\ddot{Z}\right)^{-1}
\end{aligned}
$$

The procedure needs a working variance matrix Ω_l that is unknown making this correction infeasible . However, as second best solution one can choose a feasible one that reduces the bias as discussed in the main text. If G_l is invertible the bias correction reduces to that of Bell and McCaffrey (2002) setting $F_l = Q_{\ddot{Z},ll}^{-1/2}$, since $Q_{\widetilde{W}} = Q_{\widetilde{D}} Q_{\ddot{Z}}$.

6 Stata code for projected robust bias corrected standard error

```
clear all
matrix drop _all
macro drop _all
program drop _all
sca drop _all
timer clear
clear mata
set matastrict on

capture log close
capture set more off
version 16

cd "C:\seadrive_root\Michael\My Libraries\PPML_cluster\Second revision OBES\stata"
**cd "C:\seadrive_root\Michael_1\My Libraries\PPML_cluster\Second revision OBES\stata"

/**********************************************
*** Log file
**********************************************/
log using stata_example, replace

/**********************************************
*** Data
**********************************************/
use stata_example, clear

/* Note data must be balanced and missings are set to
zero. The dummy variable V takes the value of 1
if data of a country pair are observed and zero else */


/**********************************************
*** Globals
**********************************************/
global re=  " border distw contig comlang_off colony  rta " /* explanatory variables*/
global b=63 /*numer of countries*/
global ka=0 /* working variance assumption */


/**********************************************
*** Outliers
**********************************************/
replace V=0 if (ex==61 & im==61)
replace V=0 if (ex==62 & im==62)
replace V=0 if (ex==63 & im==63)
**replace V=0 if s <= 0

/*
    exname |        61         62         63 |     Total
-----------+---------------------------------+----------
       DEU |        63          0          0 |        63
       JPN |         0         63          0 |        63
       USA |         0          0         63 |        63
-----------+---------------------------------+----------
     Total |        63         63         63 |       189
*/
**********************************************/

/**********************************************
*** Mata procdure for Pustejovsky and
*** Tipton corrected  standard errors
**********************************************/
qui mata
function pt_avcr(real matrix QsZs, real matrix Hs, real matrix WOMs, real matrix WOMCs, ///
            real matrix WOM,  real matrix S, real matrix Mhi, real matrix OM,  ///
            real scalar cc )
{
real scalar K
```

```
real matrix B, Br, Sr, Sri, U, F, AVCR, Vt, s

B=WOMCs'*((Hs*WOMs*Hs'):*S)*WOMCs
svd(B, U, s, Vt)
Sr=diag(s)
K=rows(Hs)-cc
U=U[.,1..K]
Sr=Sr[1..K,1..K]
Sri=sqrt(invsym(Sr))

F=WOMCs*U*Sri*U'*WOMCs'
Br=(Mhi*OM*Mhi):*S
AVCR=QsZs'*F*Br*F'*QsZs
return(AVCR)

printf("function pt  \n")
}

function pt (real vector b)
{
real scalar cl
real vector y, yp, f, ec
real matrix Dx, Dfx, Dm, D,Ds, S, Z, Zs, ZSS, QsZs
real matrix W, Ws, Q, Qs, H, Hs, GIZZ, PhiZZ, Vpoipt, Spoipt
real matrix s, U, V, Vt,VV, M, Mh, Mhi, OM , WOM, WOMC, WOMCs, WOMs
real matrix AVCRx, AVCRm, AVCRxm

printf("Data \n")
st_view(y,., "s")
st_view(Dm,., "ibn.im")
st_view(Dx,., "ibn.ex")
st_view(Dfx,., "ibn.ex")
st_view(Z,., "$re")
st_view(V, ., "V")
Dx=Dx[., 1..$b-1]
D=(Dm,Dx)
W=(Z,D)
VV=diag(V)

printf("y yp, and M \n")
y=V:*y
yp=exp(W*b)
M=diag(yp)
Mh=diag(V:*(yp:^(0.5)) )
Mhi=diag(V:*(yp:^(-0.5)) )
yp=V:*yp

printf("OM \n")
U=.
s=.
Vt=.
OM= (y-yp)*(y-yp)'

printf("Tilde transformations \n")
Zs=Mh*Z
Ds=Mh*D
Ws=Mh*W

printf("GIZZ\n")
Q=I(rows(Z))-VV*D*luinv(D'*VV*M*D)*D'*VV*M
ZSS=VV*Q*Z
GIZZ=luinv(quadcross(ZSS,yp,ZSS))

printf("Qs Hs\n")
Qs=I(rows(Z))-VV*Ds*luinv(Ds'*VV*Ds)*Ds'*VV
Hs=I(rows(Ws))-VV*Ws*luinv(Ws'*VV*Ws)*Ws'*VV
QsZs=Qs*Zs

```

```
printf("start x cluster \n")
WOM=M* (I(rows(M))+Dfx*Dfx' )*M
WOMs=Mhi*WOM*Mhi
WOMC=cholesky(WOM)
WOMCs=Mhi*WOMC
S=Dfx*Dfx'
cl=cols(Dfx) /*number of clusters for singularity*/

printf("start pt_avcr \n")
AVCRx=pt_avcr(QsZs,Hs,WOMs, WOMCs, WOM,S,Mhi,OM,cl)
printf("end pt_avcr \n")
printf("done x cluster \n")

printf("start m cluster  \n")
WOM=M* (I(rows(M))+ Dm*Dm') *M
WOMs=Mhi*WOM*Mhi
WOMC=cholesky(WOM)
WOMCs=Mhi*WOMC
S=Dm*Dm'
cl=cols(Dm) /*number of clusters for singularity*/

printf("start pt_avcr \n")
AVCRm=pt_avcr(QsZs,Hs,WOMs, WOMCs, WOM,S,Mhi,OM,cl)
printf("end pt_avcr \n")
printf("done m cluster \n")

printf("start xm cluster \n")
WOM=M*M
WOMs=Mhi*WOM*Mhi
WOMC=cholesky(WOM)
WOMCs=Mhi*WOMC
S=I(rows(M))
cl=0 /*number of clusters for singularity*/
printf("start pt_avcr \n")
AVCRxm=pt_avcr(QsZs,Hs,WOMs, WOMCs, WOM,S,Mhi,OM,cl)
printf("end pt_avcr \n")
printf("done xn cluster \n")

printf("VC matrix \n")
Vpoipt=GIZZ*(AVCRx+AVCRm-AVCRxm)*GIZZ'
Spoipt=diagonal(Vpoipt):^0.5
return(Spoipt)
}
end

/**********************************************
*** Mata procdure for projected standard errors
**********************************************/
mata
function pr (real vector b)
{
real vector y, yp, f, ec, V, c
real matrix Dx, Dfx, Dm, D,Ds, Z, Zs, W, Ws, Q,H, GIZZ, PhiZZ, Vpoihc, Spoihc
real matrix VV, M, Mh, Mhi, OMc

printf("Data \n")
st_view(y,., "s")
st_view(Dm,., "ibn.im")
st_view(Dx,., "ibn.ex")
st_view(Dfx,., "ibn.ex")
st_view(Z,., "$re")
st_view(V, ., "V")
Dx=Dx[., 1..$b-1]
D=(Dm,Dx)
W=(Z,D)

printf("y, yp and M \n")
y=V:*y
```

```
yp=exp(W*b)

M=diag(yp)
Mh=diag(V:*(yp:^(0.5)) )
Mhi=diag(V:*(yp:^(-0.5)) )
yp=V:*yp

printf("Tilde transformations \n")
Ws=Mh*W
Zs=Mh*Z
Ds=Mh*D

printf("Projection matrices \n")
Q=(diag(V)-Ds*cholinv(quadcross(Ds,Ds))*Ds')
H=(diag(V)-Ws*cholinv(quadcross(Ws,Ws))*Ws')

printf("Within transformation  of Z \n")
Zs=Q*Zs

printf("Correction factor of the PPMl-residuals \n")
f=(diagonal(H'*( diag(yp:^(1+$ka))    )*H))+0.00001*(J(rows(y),1,1)-V)
/*0.00001 avoids missings, multiplication by V eliminates this*/

c=sqrt((yp:^(1+$ka)):/f)
ec=(V:*(H*Mhi*(V:*(y-yp)) ):*c)

st_store(.,"cstata",c)
st_store(.,"fstata",f)
st_store(.,"ecstata",ec)

printf("VC matrix \n")
OMc=diag(ec*ec')
GIZZ=cholinv(quadcross(Zs,Zs))
PhiZZ=Zs'*OMc*Zs
PhiZZ=quadcross(Zs,diagonal(OMc),Zs)
Vpoihc=GIZZ*PhiZZ*GIZZ'
Spoihc=diagonal(Vpoihc):^0.5
return(Spoihc)
}
end
**********************************************/

/*********************************************
*** Mata procedure jacknife
**********************************************/
qui mata
function jk (real vector b)
{

real vector y, yp, f, ec
real matrix Dx,Dfx, Dm, D,Ds, Z, Zs, ZS, W, Ws, Q, QZs, H, GIZZ, PhiZZ, Vpoijk, Spoijk
real matrix V, M, Mh, Mhi, OMc, OMx, OMm, OMxm
real matrix  iZssZs, PZppx, iPZppx, PhiZZx, PZppm, iPZppm,  PhiZZm, PZppxm
real matrix  iPZppxm , PhiZZxm

printf("Data \n")
st_view(y,., "s")
st_view(Dm,., "ibn.im")
st_view(Dfx,., "ibn.ex")
st_view(Z,., "$re")
st_view(V, ., "V")
Dx=Dfx[., 1..$b-1]
D=(Dm,Dx)
W=(Z,D)

printf("y, yp and M \n")
y=V:*y
yp=exp(W*b)
```

```
M=diag(yp)
Mh=diag(V:*(yp:^(0.5)) )
Mhi=diag(V:*(yp:^(-0.5)) )
yp=V:*yp

printf("Tilde transformations \n")
Ws=Mh*W
Zs=Mh*Z
Ds=Mh*D

printf("start Jk\n")
OMx=( Mhi*(y-yp)*(y-yp)' *Mhi):*( Dfx*Dfx')
OMm=( Mhi*(y-yp)*(y-yp)' *Mhi):*( Dm*Dm')
OMxm=(Mhi*(y-yp)*(y-yp)' *Mhi):*( I(rows(D)) )

V=diag(V)
Q=I(rows(Z))-V*D*luinv(D'*V*M*D)*D'*V*M
ZS=V*Q*Z
GIZZ=luinv(quadcross(ZS,yp,ZS))

Zs=Mh*V*Q*Z
iZssZs=luinv(Zs'*Zs)
iZssZs
QZs=(I(rows(Z))-Zs*iZssZs*Zs')
PZppx=QZs:*(Dfx*Dfx')
iPZppx=luinv(PZppx)
PhiZZx=Zs'*iPZppx*V*OMx*V*iPZppx'*Zs

PZppm=QZs:*(Dm*Dm')
iPZppm=luinv(PZppm)
PhiZZm=Zs'*iPZppm*V*OMm*V*iPZppm'*Zs

PZppxm=QZs:*( I(rows(D)) )
iPZppxm=luinv(PZppxm)
PhiZZxm=Zs'*iPZppxm*V*OMxm*V*iPZppxm'*Zs

printf("VC matrix \n")
Vpoijk=GIZZ*(PhiZZx+PhiZZxm+PhiZZxm)*GIZZ'
Spoijk=diagonal(Vpoijk):^0.5
printf("done Jk\n")
return(Spoijk)
}
end
***********************************************/

/***********************************************
*** PPML under heteroskedasticity
***********************************************/
qui glm s $re  ibn.im ib$b.ex if V==1, nocons   ///
    irls family(poisson) robust

predict sp
gen r=s-sp
sum r,d
list ex im s sp r if r < -0.005, clean

sca Kall = e(k)
sca K=Kall-2*$b
di Kall " " K

/***********************************************
*** Residual check
***********************************************/
qui {
mat bpoi=e(b)'
mat bpoi=bpoi[1..Kall-1,1]
mata
st_addvar("double", "cstata")
```

```
st_addvar("double", "fstata")
st_addvar("double", "ecstata")
bpoi=st_matrix("bpoi")
Spoihc=pr(bpoi)
end
}

list ex im s sp r if r < -0.005, clean
list ex im s sp r *stata if cstata > 10, clean

/**********************************************
*** Save est. robust parameters and standard errors
**********************************************/
qui {
mat bpoi=e(b)'
mat Vpoi=e(V)
mat bpoi=bpoi[1..Kall-1,1]
mat Vpoi=Vpoi[1..K,1..K]
mat Spoi=vecdiag(Vpoi)'
mat Spoi=cholesky( diag(Spoi))
mat Spoi=vecdiag(Spoi)'
mat out=(bpoi[1..K,1],Spoi)
}

/**********************************************
*** PPML under two-way clustering
**********************************************/
qui vcemway glm s $re   ibn.im ib$b.ex if V==1, nocons ///
    irls family(poisson) cluster(ex im)

/**********************************************
*** Save est. cluster-robust
*** parameters and standard errors
**********************************************/
qui {
mat Vc=e(V)
mat Vc1=e(V_raw) /*used in Vppml procedure*/
mat Vc2=e(V_modelbased)
mat Vc=Vc[1..K,1..K]
mat Vc1=Vc1[1..K,1..K]
mat Vc2=Vc2[1..K,1..K]

mat Sc=vecdiag(Vc)'
mat Sc=cholesky( diag(Sc))
mat Sc=vecdiag(Sc)'
mat out=(out,Sc)
mata: Sc=st_matrix("Sc")
}

/**********************************************
*** Run mata for  bias corrected
*** standard errors
**********************************************/
qui mata
bpoi=st_matrix("bpoi")
Spoipt=pt(bpoi) /*Pustejovsky and Tipton */
Spoihc=pr(bpoi) /*projected*/
Spoijk=jk(bpoi) /*jackknife*/
end

/**********************************************
*** Results table
**********************************************/
qui {
mata st_matrix("outm", Spoipt)
mat out=(out,outm)
mata st_matrix("outm", Spoijk)
mat out=(out,outm)
```

```
mata st_matrix("outm", Spoihc)
mat PPML_results=(out,outm)
}

local names ="b  sc pt  jk  pr "
matrix colnames PPML_results =b  het cl pt jk pr
estout matrix(PPML_results, fmt( %4.2f) )
```

References

Abadir, Karim M., Magnus, Jan R., 2005. Matrix algebra. Vol. 1. Cambridge University Press.

Bell, Robert M., McCaffrey, Daniel F., 2002. Bias reduction in standard errors for linear regression with multi-stage samples. Survey Methodology 28(2), 169–182.

Cook, Dennis R, Weisberg, Sanford, 1982. Residuals and influence in regression. New York: Chapman and Hall.

Cameron, Colin A. , Gelbach, Jonah B., Miller, Douglas L., 2011. Robust inference with multi-way clustering. Journal of Business and Economics Statistics 29(2), 238–249.

Davidson, Russell, and MacKinnon, James G., 1993. Estimation and inference in econometrics. Oxford University Press, Oxford UK.

Davezies, Laurent, D'Haultfœuille Xavier, Guyonvarch, Yannick Guyonvarch, 2021. Empirical process results for exchangeable arrays. Annals of Statistics 49(2), 845–862.

Davis, Peter, 2002. Estimating multi-way error components models with unbalanced data structures. Journal of Econometrics 106(1), 67–95.

Hansen, Bruce, 2019. Econometrics, unpublished manuscript.

Henderson, Harold V., and Shayle R. Searle, 1981. On deriving the inverse of a sum of matrices. Siam Review 23(1), 53–60.

Imbens, Guido W. and Kolesar, Michal, 2016. Robust standard errors in small samples: Some practical advice. Review of Economics and Statistics 98, 701–712.

Janson, Svante, 1988. Normal convergence by higher Semiinvariants with applications to sums of dependent random variables and random graphs, The Annals of Probability 16, 305–12.

MacKinnon, James G., Nielsen, Morten. Ø., Webb, Matthew. D., 2021. Wild bootstrap and asymptotic inference with multi-way clustering. Journal of Business and Economics Statistics 39(2), 505-519.

Niccodemi, Gianmaria, and Tom Wansbeek, 2022. A New Estimator for Standard Errors with Few Unbalanced Clusters. Econometrics 10.1: 6.

Pfaffermayr, Michael, 2020. Constrained poisson pseudo maximum likelihood estimation of structural gravity models. International Economics 161, 188–198.

Pustejovsky, James E., Tipton, Elizabeth, 2018. Small-sample methods for cluster-robust variance estimation and hypothesis testing in fixed effects models, Journal of Business and Economics Statistics 36(4), 672–683.

Tabord-Meehan, Max, 2019. Inference with dyadic data: Asymptotic behaviour of the dyadic-robust t-statistic. Journal of Business and Economics Statistics 37(4), 671-680.