Appendix to
Gravity Models, PPML Estimation and the Bias of the Robust Standard Errors

Michael Pfaffermayr

August 10, 2018

1 The Dummy PPML Estimator

In order to derive the limit distribution of the PPML estimator for α, we define $W = [Z, D]$, $G^* = W'VM^*VW$ with $M^* = diag(e^{w'_{ij}\vartheta^*})$, where $\vartheta^* = (\alpha^{*\prime}, \phi^{*\prime})'$ with $\phi^* = (\beta^{*\prime}, \gamma^{*\prime})'$ lies elementwise between $\widehat{\vartheta}$ and ϑ_0. Applying the mean-value theorem to the PPML-score yields

$$0 = W'V\varepsilon - G^* \left[\begin{array}{c} \widehat{\alpha} - \alpha_0 \\ \widehat{\phi} - \phi_0 \end{array} \right].$$

For missing values one may define the selection matrix V that is derived from the identity matrix by setting all ones in the main diagonal to zero if the corresponding observation is missing. Defining $\widetilde{Z}^* = M^{*\frac{1}{2}}Z$, $\widetilde{D}^* = M^{*\frac{1}{2}}D$ and $Q_{V\widetilde{D}^*} = I - V\widetilde{D}^* \left(\widetilde{D}^{*\prime}V\widetilde{D}^*\right)^{-1} \widetilde{D}^{*\prime}V$ and using the blocks of the partitioned inverse

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

one can write

$$C\left(\widehat{\alpha} - \alpha_0\right) = C^2G^{*11}\tfrac{1}{C}(Z'V - G^*_{12}G^{*-1}_{22}D'V)\varepsilon := B^{*-1}A^*\varepsilon,$$

where $B^* = \frac{1}{C^2}\left(\widetilde{Z}'^*VQ_{V\widetilde{D}^*}V\widetilde{Z}\right)$ and $A^* = \frac{1}{C}\left(\widetilde{Z}'^*VQ_{V\widetilde{D}^*}M^{*-\frac{1}{2}}\right)$. Following Fernandez-Val and Weidner (2017) and Wooldridge (1997) under a set of standard regularity conditions, the limit distribution of $\widehat{\alpha}$ can be derived as

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

where $V_\alpha = B_0^{-1}A_0\Omega_\varepsilon A_0'B_0^{-1}$ with $B_0 = p\lim_{C\to\infty} B^*$ is assumed to be invertible, $A_0\Omega_\varepsilon A_0' = p\lim_{C\to\infty} A^*\varepsilon\varepsilon' A^{*\prime}$. $\Omega_\varepsilon = E[\varepsilon\varepsilon']$ is the diagonal variance matrix of ε with typical element $\sigma^2_{\varepsilon,ij}$. Plugging in the estimated residuals $\widehat{\varepsilon}$, one can use $\frac{1}{C^2}\widehat{V_\alpha} = \frac{C^2-1}{C^4}B(\widehat{\alpha})^{-1}A(\widehat{\alpha})diag(\widehat{\varepsilon}\widehat{\varepsilon}')A(\widehat{\alpha})'B(\widehat{\alpha})^{-1}$ for inference in finite samples.

2 The bias of the standard errors of the Dummy PPML estimator

The bias of the standard errors of the dummy 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 and B, but use the residuals $\widehat{\varepsilon}$. Moreover, we take $\widetilde{W} = M_0^{\frac{1}{2}} W$ and $\widetilde{Z} = M_0^{\frac{1}{2}} Z$ as a non-stochastic matrices, whose elements are uniformly bounded. Then $\widehat{V}_\alpha$ can be written as

$$\widehat{V}_\alpha = B_0^{-1} A_0 diag(\widehat{\varepsilon}\widehat{\varepsilon}') A_0' B_0^{-1}.$$

Defining $H_{\widetilde{W}} = I_{C^2} - \widetilde{W}\left(\widetilde{W}'\widetilde{W}\right)^{-1}\widetilde{W}'$, the residuals under dummy PPML are given as (see Davidson and MacKinnon, 1993, 123-167)

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

since $\left\|\widehat{\vartheta} - \vartheta_0\right\| = \left\|C^2 G(\vartheta^*)^{-1} \frac{1}{C^2} W'\varepsilon\right\| \leq \left\|C^2 G(\vartheta^*)^{-1}\right\| \left\|\frac{1}{C^2} W'\varepsilon\right\| = O_p(C^{\frac{1}{2}}) O_p(C^{-1/2})$. This uses $G(\vartheta^*) = W'M^*W$ and

$$\begin{aligned} (K + 2C - 1)^{\frac{1}{2}} \lambda_{\min} &\leq \left\|\tfrac{1}{C^2} G(\vartheta^*)^{-1}\right\| \leq (K + 2C - 1)^{\frac{1}{2}} \lambda_{\max}, \\ \left\|C^2 G(\vartheta^*)^{-1}\right\| &\leq \tfrac{(K+2C-1)^{\frac{1}{2}}}{\lambda \min}, \end{aligned}$$

where $\lambda_{\min}$ and $\lambda_{\max}$ are the minium and maximum eigenvalues of $C^{-2} G(\vartheta)$ and $\lambda_{\min} > 0$ and $\lambda_{\max} < \infty$ in the compact parameter space is assumed.

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

since $E\left[|\varepsilon_{ij}|^2\right] \leq \overline{\sigma}^2$ is assumed and Markov's inequality implies $\left|\varepsilon_{ij}^2\right| = O_p(1)$. Furthermore $\|w_{ij}\| = (K + 2C - 1)^{\frac{1}{2}} c_w$ for some constant c_w.

Inserting for $\widehat{\varepsilon} = M_0^{1/2} H_{\widetilde{W}} M_0^{-1/2} \varepsilon$ in $\widehat{V}_\alpha$ yields

$$\widehat{V}_\alpha = B_0^{-1} A_0 diag(M_0^{1/2} H_{\widetilde{W}} M_0^{-1/2} \varepsilon\varepsilon' M_0^{-1/2} H_{\widetilde{W}} M_0^{1/2}) A_0' B_0^{-1} + o_p(1).$$

Using the multiplicative error specification with $\Omega_\varepsilon = M_0^2 \Omega_\eta$, under regularity conditions the bias of $\widehat{V}_\alpha$ can be written as

$$\begin{aligned} & E\left[\widehat{V}_\alpha - V_\alpha\right] \\ = \ & B_0^{-1} A_0 M_0 diag\left(M_0^{-1/2} H_{\widetilde{W}} M_0^{1/2} \Omega_\eta M_0^{1/2} H_{\widetilde{W}} M_0^{-1/2} - \Omega_\eta\right) M_0 A_0' B_0^{-1} + o(1), \end{aligned}$$

see Chesher and Jewitt (1987) and Cribari–Neto, Ferrari and Cordeiro (2000). The matrix $M_0^{-1/2} H_{\widetilde{W}} M_0^{1/2}$ can written as

$$M_0^{-1/2} H_{\widetilde{W}} M_0^{1/2} = I_{C^2} - \underbrace{M_0^{-1/2} \widetilde{W} \left(\widetilde{W}' \widetilde{W} \right)^{-1} \widetilde{W}' M_0^{1/2}}_{P_{\widetilde{W}}} = I_{C^2} - P_{\widetilde{W}}$$

where $P_{\widetilde{W}}$ is idempotent. It follows that

$$\begin{aligned} & M_0^{-1/2} H_{\widetilde{W}} M_0^{1/2} \Omega_\eta M_0^{1/2} H_{\widetilde{W}} M_0^{-1/2} - \Omega_\eta \\ = & \left(I_{C^2} - P_{\widetilde{W}} \right) \Omega_\eta \left(I_{C^2} - P_{\widetilde{W}} \right)' - \Omega_\eta \\ = & -P_{\widetilde{W}} \Omega_\eta - \Omega_\eta P'_{\widetilde{W}} + P_{\widetilde{W}} \Omega_\eta P'_{\widetilde{W}}, \end{aligned}$$

leading to

$$E\left[\widehat{V}_\alpha - V_\alpha \right] = B_0^{-1} A_0 M_0 diag \left(P_{\widetilde{W}} \Omega_\eta P_{\widetilde{W}} - P_{\widetilde{W}} \Omega_\eta - \Omega_\eta P'_{\widetilde{W}} \right) M_0 A'_0 B_0^{-1} + o(1).$$

The results of Chesher and Jewitt (1987) can be directly applied. The ij, ij-diagonal element of $diag \left(P_{\widetilde{W}} \Omega_\eta P'_{\widetilde{W}} - P_{\widetilde{W}} \Omega_\eta - \Omega_\eta P'_{\widetilde{W}} \right)$ is given as

$$p'_{\widetilde{W},ij} \Omega_\eta p_{\widetilde{W},ij} - 2\omega_{ij} p_{\widetilde{W},ij} = p'_{\widetilde{W},ij} \left(\Omega_\eta - 2\omega_{ij} I_{C^2} \right) p_{\widetilde{W},ij},$$

where $p_{\widetilde{W},ij}$ be the ij-th column of $P_{\widetilde{W}}$ and observing that $p'_{\widetilde{W},ij} p_{\widetilde{W},ij} = p_{\widetilde{W},ij}$, since $P_{\widetilde{W}}$ is idempodent. As in Chesher and Jewitt (1987) the proportionate bias of $\widehat{V}_\alpha$ (bp) is defined as $pb(\widehat{V}_\alpha) = E\left[\frac{v'\left(\widehat{V}_\alpha - V_\alpha\right)v}{v' V_\alpha v} \right]$ for some vector $v \neq 0$. Ignoring the remainder being $o(1)$, it follows that

$$\begin{aligned} v' \left(\widehat{V}_\alpha - V_\alpha \right) v &= v' B_0^{-1} A_0 M_0 diag \left(P_{\widetilde{W}} \Omega_\eta P_{\widetilde{W}} - P_{\widetilde{W}} \Omega_\eta - \Omega_\eta P_{\widetilde{W}} \right) M_0 A'_0 B_0^{-1} v \\ &= z' diag(P_{\widetilde{W}} \Omega_\eta P_{\widetilde{W}} - P_{\widetilde{W}} \Omega_\eta - \Omega_\eta P_{\widetilde{W}})) z \\ &= z' diag(p'_{\widetilde{W},ij} \left(\Omega_\eta - 2\omega_{ij} I_{C^2} \right) p_{\widetilde{W},ij}) z \end{aligned}$$

with $z = M_0 A'_0 B_0^{-1} v$. The proportionate bias in general depends on the degree of heteroskedasticity and on the features of the data as represented by the leverage $p_{\widetilde{W},ij,ij}$ of $P_{\widetilde{W}}$, which is of order $O(C^{-1})$, since $trace(P_{\widetilde{W}}) = K + 2C - 1$. Further, a lower and upper bound of bp can be established:

$$\inf_z \left(pb(\widehat{V}_\alpha)\right) \geq \min_{ij} \sum_{l=1,l\neq i}^{C} \sum_{k=1,l\neq j}^{C} \frac{\sigma^2_{\eta,lk}}{\sigma^2_{\eta,ij}} p^2_{\widetilde{W},ij,lk} + p_{\widetilde{W},ij,ij}(p_{\widetilde{W},ij,ij} - 2).$$

$$\sup_z \left(pb(\widehat{V}_\alpha)\right) \leq \max_{ij} \sum_{l=1,l\neq i}^{C} \sum_{k=1,l\neq j}^{C} \frac{\sigma^2_{\eta,lk}}{\sigma^2_{\eta,ij}} p^2_{\widetilde{W},ij,lk} + p_{\widetilde{W},ij,ij}(p_{\widetilde{W},ij,ij} - 2).$$

Idempodency of $P_{\widetilde{W}}$ implies that $0 \leq \sum_{l=1,l\neq i}^{C} \sum_{k=1,l\neq j}^{C} p^2_{\widetilde{W},ij,lk} = p_{\widetilde{W},ij,ij}\left(1 - p_{\widetilde{W},ij,ij}\right)$ $\leq \frac{1}{2}$. Since $\frac{\sigma^2_{\eta,lk}}{\sigma^2_{\eta,ij}} \leq \frac{\overline{\sigma}^2_\eta}{\underline{\sigma}^2_\eta}$ and since the elements of the main diagonal of $P_{\widetilde{W}}$ are of order $O(C^{-1})$, we have

$$\sup_z \left(pb(\widehat{V}_\alpha)\right) \leq \max_{ij} \left[\left(\frac{\overline{\sigma}^2_\eta}{\underline{\sigma}^2_\eta} - 1\right) p_{\widetilde{W},ij,ij}(1 - p_{\widetilde{W},ij,ij}) - p_{\widetilde{W},ij,ij}\right] = O(C^{-1})$$

Since $\sum_{l=1,l\neq i}^{C} \sum_{k=1,l\neq j}^{C} \frac{\sigma^2_{\eta,lk}}{\sigma^2_{\eta,ij}} p^2_{\widetilde{W},ij,lk} > 0$ and $p_{\widetilde{W},ij,ij}(p_{\widetilde{W},ij,ij} - 2)$ is decreasing in $p_{\widetilde{W},ij,ij}$

$$\inf_{ij} \left(pb(\widehat{V}_\alpha)\right) \geq \max_{ij} \left(p_{\widetilde{W},ij,ij}\right) \left(\max_{ij}(p_{\widetilde{W},ij,ij} - 2\right) = O(C^{-1}).$$

References

Chesher A. and I. Jewitt (1987), The Bias of a Heteroskedasticity Consistent Covariance Matrix Estimator, *Econometrica* **55**(5), 1217–1222.

Cribari–Neto, F., S. L. Ferrari and G. M. Cordeiro, (2000). Improved Heteroscedasticity-Consistent Covariance Matrix Estimators. *Biometrika* **87**(4), 907-918.

Davidson, R. and MacKinnon, J. G. (1993). Estimation and Inference in Econometrics. Oxford University Press, Oxford UK.

Fernandez-Val, I. and M. Weidner (2016), Individual and Time Effects in Nonlinear Panel Models with large N, T, *Journal of Econometrics* **192(1),** 291–312.

Wooldridge, J. (1997), Quasi-Likelihood Methods for Count Data, in Pesaran H. and P. Schmidt (eds.), Handbook of Applied Econometrics Volume II: Micro-economics, Blackwell, Oxford UK.