Economics Letters 213 (2022) 110408

Contents lists available at ScienceDirect

Economics Letters

journal homepage: www.elsevier.com/locate/ecolet

A correlated random effects approach to the estimation of models with multiple fixed effects[☆]

Yimin Yang [*]

China Financial Policy Research Center, School of Finance, Renmin University of China, China

ARTICLE INFO

Article history:
Received 4 September 2021
Received in revised form 22 February 2022
Accepted 23 February 2022
Available online 4 March 2022

JEL classification:
C13
C23

Keywords:
Correlated random effects
Hausman test
Multiple fixed effects

ABSTRACT

This paper generalizes two important equivalence results in one-way panel models to models with multiple fixed effects. (1) We show that the correlated random effects estimator obtained via the use of the Mundlak device is identical to the fixed effects estimator. (2) A weighted variable addition test is shown to be equivalent to the fully robust Hausman specification test.

1. Introduction

Models with multiple fixed effects are now applied to a variety of areas, which include, transport, environment and migration. See Kabir et al. (2017) for a recent review. A prominent example is the two-way fixed effects models used in the international trade. Anderson's (1979) seminal work provides a theoretical foundation for the use of two-way model in analysing bilateral trade flows.

A random effects (RE) treatment to models with multiple fixed effects can be found in Matyas (2017, Chapter 2). The fixed effects (FE) estimation of such models is first considered by Matyas (1997) and is further explored for different settings by Baltagi et al. (2003), Egger and Pfaffermayr (2003), Balestra and Krishnakumar (2008) among others. See Balazsi et al. (2018) and Matyas (2017) for some excellent reviews.

As a middle ground between the RE and the FE estimations, this paper considers the estimation of linear models with multiple fixed effects within the correlated random effects (CRE) framework. We extend and generalize several equivalence results in one-way panel models to models with multiple fixed effects. In particular, we show that the CRE estimator obtained via the use of Mundlak device (Mundlak, 1978) is identical to the FE estimator. This equivalence is well known in one-way panel models and has been extended to various other contexts (e.g. Arkhangelsky and Imbens, 2019; Baltagi, 2006; Yang and Schmidt, 2021). As a byproduct of the CRE estimation, a weighted variable addition test is shown to be identical to the robust Hausman specification test (Hausman, 1978). While these results are well-known in one-way panel models, we give formal proofs of these results for models with multiple fixed effects.

This paper mainly focuses on the two-way linear model as extensions to models with more dimensions are straightforward. Section 2 presents two equivalence results related to the CRE estimator and the Hausman specification test. Section 3 concludes. All proofs can be found in the appendix of the supplementary material.

Some notations in order: let ι_d be the $d \times 1$ column vector with ones. Define $P_d = \iota_d(\iota_d'\iota_d)^{-1}\iota_d'$ and $M_d = I_d - P_d$. Let x_{ij} denote the covariates indexed by the pair (i, j), where $i = 1, 2, \ldots, N$ and $j = 1, 2, \ldots, L$. We assume that (i, j) is observed for all possible combinations. Let $\bar{x}_{i\bullet} = \frac{1}{L}\sum_{j=1}^{L} x_{ij}$ and $\bar{x}_{\bullet j} = \frac{1}{N}\sum_{i=1}^{N} x_{ij}$. $x_{i\bullet}$ represents the collection of $(x_{i1}, x_{i2}, \ldots, x_{iL})$ and $x_{\bullet j}$ is defined in a similar way.

2. Estimation and testing

This section considers the estimation and test of models with multiple fixed effects by the correlated random effects approach. A simple two-way linear model is specified as the following:

$$y_{ij} = x_{ij}'\beta + v_{ij}, \tag{1}$$

☆ I thank the editor Baltagi, an associate editor and two anonymous referees for their constructive comments, which led to a much improved presentation. I am also grateful to Peter Schmidt for his helpful comments on an earlier draft.

* Correspondence to: Mingde Main Building, Renmin University of China, Haidian District, Beijing, China.

E-mail address: yimin.yang@ruc.edu.cn.

https://doi.org/10.1016/j.econlet.2022.110408
0165-1765/

$$\nu_{ij} = \alpha_i + \gamma_j + e_{ij}, \tag{2}$$

where y_{ij} denotes the dependent variable indexed by the pair (i, j) and x_{ij} denotes the corresponding explanatory variables. We assume that x_{ij} varies across both i and j because the fixed effects estimation eliminates any variables constant at i or j. The regression error ν_{ij} is decomposed into three parts with α_i and γ_j representing the fixed effects for import country and export country in the bilateral trade analysis. e_{ij} is the usual idiosyncratic disturbance.[1]

Throughout this paper we maintain the following exogeneity assumption.

$$E(e_{ij}|x_{11}, x_{12}, \ldots, x_{NL}, \alpha_i, \gamma_j) = 0 \quad \text{for all } i \text{ and } j, \tag{3}$$

where $(x_{11}, x_{12}, \ldots, x_{NL})$ is the set of covariates of all (i, j) sample pairs as defined in the introduction section. The above assumption is an extension of the strict exogeneity assumption in panel data models. It is useful to rewrite the model in matrix form as

$$Y = X\beta + V = X\beta + (I_N \otimes \iota_L)\alpha + (\iota_N \otimes I_L)\gamma + e, \tag{4}$$

where $\alpha = (\alpha_1, \alpha_2, \ldots, \alpha_N)'$ and $\gamma = (\gamma_1, \gamma_2, \ldots, \gamma_L)'$, Y is a $NL \times 1$ vector with its ith block $(y_{i1}, \ldots, y_{iL})'$, and X, V and e are defined similarly.

The RE estimator assumes that α, γ and e are mutually uncorrelated and $E(\alpha_i\alpha_s) = \sigma_\alpha^2 \mathbb{1}[i = s]$, $E(\gamma_j\gamma_m) = \sigma_\gamma^2 \mathbb{1}[j = m]$, $E(e_{ij}e_{sm}) = \sigma_e^2 \mathbb{1}[i = s, j = m]$, where $\mathbb{1}(\cdot)$ is the usual indicator function. The RE estimator for the two-way model can be written as

$$\hat{\beta}_{RE} = (X'\Omega_1^{-1}X)^{-1}(X'\Omega_1^{-1}Y), \tag{5}$$

where $\Omega_1 = E(\mathrm{VV}') = L\sigma_\alpha^2(\mathrm{I}_N \otimes \mathrm{P}_L) + N\sigma_\gamma^2(\mathrm{P}_N \otimes \mathrm{I}_L) + \sigma_e^2(\mathrm{I}_N \otimes \mathrm{I}_L)$. Derivation of the Ω_1^{-1} is provided in Matyas (2017, Chapter 2). The consistency of the RE estimator requires that the covariates are exogenous with respect to all the heterogeneities.

The two sources of fixed effects in Eq. (2) are non-nested in the sense that knowing the export country i does not inform us the identity of the import country j. See Yang and Schmidt (2021) for more discussions of nested and non-nested models. Because there are no collinearity among different fixed effects, a fixed effects approach is available

$$\hat{\beta}_{FE} = (X'M_1X)^{-1}(X'M_1Y), \tag{6}$$

where $M_1 = I_{NL} - (I_N \otimes P_L) - (P_N \otimes I_L) + (P_N \otimes P_L)$. The FE estimator is consistent whenever Eq. (3) and proper rank conditions hold. Balazsi et al. (2018) provide an excellent review of FE estimators for models with multiple fixed effects.

The correlated random effects approach allows for correlations between the regressors and the fixed effects, although in some restricted way. A popular modelling technique is the Mundlak device. See Wooldridge (2010, Chapter 10) for examples. The linear projection of the fixed effects onto the covariates is given by

$$E^*(\alpha_i|\mathrm{x}_{11}, \mathrm{x}_{12}, \ldots, \mathrm{x}_{NL}) = \bar{\mathrm{x}}_{i\bullet}'\delta_1, \tag{7}$$

$$E^*(\gamma_j|\mathrm{x}_{11}, \mathrm{x}_{12}, \ldots, \mathrm{x}_{NL}) = \bar{\mathrm{x}}_{\bullet j}'\delta_2, \tag{8}$$

where E^* is the linear projection operator, $\bar{\mathrm{x}}_{i\bullet}$ and $\bar{\mathrm{x}}_{\bullet j}$ are sample averages defined in the introduction. A less restrictive alternative is the Chamberlain (1980) device which specifies the linear projections as a linear combination of $\mathrm{x}_{i\bullet}$ and $\mathrm{x}_{\bullet j}$ with unrestricted coefficients. While the Chamberlain device is more flexible, the Mundlak device has an advantage of conserving on degrees of freedom, which turns out to be crucial in our setup. When N and L are large, the number of unrestricted coefficients associated with the Chamberlain device grows linearly in N and L. On the other hand, the number of parameters in the Mundlak specification is finite. Substituting Eqs. (7) and (8) into the two-way linear model gives

$$y_{ij} = x_{ij}'\beta + \bar{x}_{i\bullet}'\delta_1 + \bar{x}_{\bullet j}'\delta_2 + \varepsilon_{ij}, \tag{9}$$

where $\varepsilon_{ij} = (\alpha_i - \bar{x}_{i\bullet}'\delta_1) + (\gamma_j - \bar{x}_{\bullet j}'\delta_2) + e_{ij}$ is uncorrelated with the covariates. The pooled OLS or the RE estimator of Eq. (9) is what we call CRE estimator for the two-way linear models. Define $\Delta_1 = \left[X'(M_1 + d_1(I_N \otimes P_L - P_N \otimes P_L) + d_2(P_N \otimes I_L - P_N \otimes P_L))X\right]^{-1} d_1X'(I_N \otimes P_L - P_N \otimes P_L)X$ and $\Delta_2 = \left[X'(M_1 + d_1(I_N \otimes P_L - P_N \otimes P_L) + d_2(P_N \otimes I_L - P_N \otimes P_L))X\right]^{-1} d_2X'(P_N \otimes I_L - P_N \otimes P_L)X$, where d_1 and d_2 are some constants associated with the variances of the error components. We summarize the properties of the CRE estimator in the following theorem.

Theorem 1. *Both the random effects estimator and the pooled OLS estimator of β in Eq. (9) are identical to the fixed effects estimator defined in Eq. (6). The Wald statistic for testing $\Delta_1\delta_1 + \Delta_2\delta_2 = 0$ is equivalent to the Hausman specification test based on $\hat{\beta}_{RE} - \hat{\beta}_{FE}$ for testing $E(\alpha|x_{11}, x_{12}, \ldots, x_{NL}) = 0$ and $E(\gamma|x_{11}, x_{12}, \ldots, x_{NL}) = 0$.*

It is well known that the CRE estimator is identical to the FE estimator in one-way panel models. For example, Baltagi (2006) uses a system estimation approach to establish the equivalence between these two estimators. Similar results have been extended to many other contexts. For instance, Arkhangelsky and Imbens (2019) show that the FE estimator of the treatment effects is identical to the OLS estimator when the unobserved group heterogeneity is approximated by the group averages of the covariates and the treatment values. Yang and Schmidt (2021) obtain a similar equivalence result for multi-level panel models. Our setting differs from theirs as we have two non-nested fixed effects.

The second part of Theorem 1 is slightly different from the classical result. Here we have two different fixed effects so there are two sets of additional parameters in the augmented regression. We show in the appendix that the RE estimator or the pooled OLS estimator of δ_1 is identical to $\hat{\beta}_{o\bullet} - \hat{\beta}_{FE}$, where $\hat{\beta}_{o\bullet} = \left(\sum_{i=1}^N(\bar{x}_{i\bullet} - \bar{x}_{\bullet\bullet})(\bar{x}_{i\bullet} - \bar{x}_{\bullet\bullet})'\right)^{-1}\left(\sum_{i=1}^N(\bar{x}_{i\bullet} - \bar{x}_{\bullet\bullet})(\bar{y}_{i\bullet} - \bar{y}_{\bullet\bullet})\right)$. It follows that a test of $\delta_1 = 0$ is equivalent to the Hausman test based on $\hat{\beta}_{o\bullet} - \hat{\beta}_{FE}$ as in the one-way panel models. The same reasoning applies to δ_2.

It is shown in the appendix that the RE estimator can be obtained by the pooled OLS of y_{ij} on $\ddot{x}_{ij} + s_1\bar{x}_{i\bullet} + s_2\bar{x}_{\bullet j}$ for some constants s_1 and s_2, where $\ddot{x}_{ij}$ is the transpose of the row vector of M_1X. Thus the deviations between the RE and FE estimators are due to correlation between $s_1\bar{x}_{i\bullet} + s_2\bar{x}_{\bullet j}$ and $\alpha_i + \gamma_j$. In other words, the null of the Hausman test based on $\hat{\beta}_{RE} - \hat{\beta}_{FE}$ is the same as a test of $E[(s_1\bar{x}_{i\bullet} + s_2\bar{x}_{\bullet j})(\alpha_i + \gamma_j)] = 0$. A simple joint test of $\delta_1 = \delta_2 = 0$ is still valid for testing the correlations between the covariates and the fixed effects but the associated Wald statistic is no longer equivalent to the Hausman statistic based on the difference $\hat{\beta}_{RE} - \hat{\beta}_{FE}$. Now the Hausman specification test based on $\hat{\beta}_{RE} - \hat{\beta}_{FE}$ is identical to a modified variable addition test of $\Delta_1\delta_1 + \Delta_2\delta_2 = 0$.[2] The two weighting matrices Δ_1 and Δ_2 can be roughly interpreted as the proportion of two different between variations of X to the total variation of X.

[1] Note that if we take j as the time index t, then the two-way linear model is exactly the same as linear panel data models with both individual fixed effects and time fixed effects, which are already well-studied in the literature.

[2] The traditional Hausman specification based on $\hat{\beta}_{RE} - \hat{\beta}_{FE}$ requires that $\hat{\beta}_{RE}$ is more efficient under the null. The traditional Hausman test is not robust to the failure of the second moment assumptions while we can easily make the modified variable addition test robust to heteroskedasticity and serial correlation by using the robust standard errors.

Theorem 1 is very similar to the result obtained in Wooldridge (2021) for the two-way Mundlak regression. Our paper is written independently of his and has the extension to higher-order models. Next we consider a three-dimensional model with three types of heterogeneity.

$$y_{ijt} = x_{ijt}'\beta + \nu_{ijt}, \tag{10}$$

$$\nu_{ijt} = \alpha_i + \gamma_j + \lambda_t + e_{ijt}. \tag{11}$$

The regression error now has an additional component λ_t, which is interpreted as the unobserved time fixed effects. The RE estimator of the three-dimensional models can be found in Maytas (2017, Chapter 2) and the FE estimator is proposed by Matyas (1997).

$$\hat{\beta}_{RE} = (X'\Omega_2^{-1}X)^{-1}(X'\Omega_2^{-1}Y), \tag{12}$$

$$\hat{\beta}_{FE} = (X'M_2X)^{-1}(X'M_2Y), \tag{13}$$

where $\Omega_2 = LT\sigma_\alpha^2(I_N \otimes P_L \otimes P_T) + NT\sigma_\gamma^2(P_N \otimes I_L \otimes P_L) + NL\sigma_\lambda^2(P_N \otimes P_L \otimes I_T) + \sigma_e^2 I_{NLT}$ and $M_2 = I_{NLT} - (I_N \otimes I_L \otimes P_T) - (I_N \otimes P_L \otimes I_T) - (P_N \otimes I_{LT}) + (I_N \otimes P_L \otimes P_T) + (P_N \otimes I_L \otimes P_T) + (P_N \otimes P_L \otimes I_T) + (P_N \otimes P_L \otimes P_T)$. The correlated random effects approach models the heterogeneity via the Mundlak device as the following

$$E^*(\alpha_i|x_{111}, x_{112}, \dots, x_{NLT}) = \bar{x}_{i\bullet\bullet}'\delta_1, \tag{14}$$

$$E^*(\gamma_j|x_{111}, x_{112}, \dots, x_{NLT}) = \bar{x}_{\bullet j\bullet}'\delta_2, \tag{15}$$

$$E^*(\lambda_t|x_{111}, x_{112}, \dots, x_{NLT}) = \bar{x}_{\bullet\bullet t}'\delta_3, \tag{16}$$

where $\bar{x}_{\bullet\bullet t} = \frac{1}{NL}\sum_{i=1}^{N}\sum_{j=1}^{L} x_{ijt}$. The idea of using $\bar{x}_{\bullet\bullet t}$ as a proxy for the unobserved fixed time effects is similar to the technique used in the common correlated effects estimator developed by Pesaran (2006) and has been applied to heterogeneous hierarchical panel models studied by Kapetanios et al. (2021). Substituting Eqs. (14), (15) and (16) into the three-dimensional models gives

$$y_{ijt} = x_{ijt}'\beta + \bar{x}_{i\bullet\bullet}'\delta_1 + \bar{x}_{\bullet j\bullet}'\delta_2 + \bar{x}_{\bullet\bullet t}'\delta_3 + \varepsilon_{ijt}. \tag{17}$$

If the sample averages properly account for the correlations between the regressors and the heterogeneity, then the pooled OLS or RE estimators are consistent. The following theorem summarizes the properties of the CRE estimators in Eq. (17).

Theorem 2. *Both the random effects estimator and the pooled estimator of β in Eq. (17) are identical to the fixed effects estimator defined in Eq. (13). The Wald statistic for testing $\tilde{\Delta}_1\delta_1 + \tilde{\Delta}_2\delta_2 + \tilde{\Delta}_3\delta_3 = 0$, where the definitions of the $\tilde{\Delta}$s can be found in the appendix, is equivalent to the Hausman specification test based on $\hat{\beta}_{RE} - \hat{\beta}_{FE}$ for testing $E(\alpha|x_{111}, x_{112}, \dots, x_{NLT}) = 0$, $E(\gamma|x_{111}, x_{112}, \dots, x_{NLT}) = 0$ and $E(\lambda|x_{111}, x_{112}, \dots, x_{NLT}) = 0$.*

The first part of Theorem 2 is the same as before. In the three-dimensional linear model, there are three types of fixed effects, which induces three additional sets of parameters. A little algebra shows that the random effects estimator $\hat{\beta}_{RE}$ in Eq. (12) can be rewritten as

$$\hat{\beta}_{RE} = \tilde{\Delta}_1\hat{\beta}_{o\bullet\bullet} + \tilde{\Delta}_2\hat{\beta}_{\bullet o\bullet} + \tilde{\Delta}_3\hat{\beta}_{\bullet\bullet o} + (\mathrm{I} - \tilde{\Delta}_1 - \tilde{\Delta}_2 - \tilde{\Delta}_3)\hat{\beta}_{FE}, \tag{18}$$

where the definitions of $\tilde{\Delta}$s can be found in the appendix and $\hat{\beta}_{\bullet\bullet o}$ is the between-group estimator using data $(\bar{x}_{\bullet\bullet t}, \bar{y}_{\bullet\bullet t})_{t=1}^{T}$. The Wald statistic testing $\delta_1 = 0$ is identical to the Hausman statistic based on $\hat{\beta}_{o\bullet\bullet} - \hat{\beta}_{FE}$. The same interpretation is true for δ_2 and δ_3. While the Hausman test for each type of heterogeneity is equivalent to the corresponding variable addition test in Eq. (17), the Hausman statistic based on $\hat{\beta}_{RE} - \hat{\beta}_{FE}$ is not identical to the joint test of $\delta_1 = \delta_2 = \delta_3 = 0$. Here the Hausman test based on $\hat{\beta}_{RE} - \hat{\beta}_{FE}$ is identical to a weighted variable addition test of $\tilde{\Delta}_1\delta_1 + \tilde{\Delta}_2\delta_2 + \tilde{\Delta}_3\delta_3 = 0$ in the three-dimensional linear model.

As discussed in Baltagi et al. (2003), the three-dimensional model designed by Eqs. (10) and (11) can be embedded in a more general setting

$$\nu_{ijt} = \mu_{ij} + \alpha_{it} + \gamma_{jt} + e_{ijt}. \tag{19}$$

The fixed effects of export country and import country are varying across time. Matyas (2017) provides a comprehensive review of the RE and FE estimators of this generalized model.

If we proceed to model μ_{ij} α_{it} and γ_{jt} by the Mundlak device, the equivalence between the CRE estimator and the FE estimator no longer holds. Algebraically, the residuals obtained from regressing x_{ijt} on $(\bar{x}_{ij\bullet}, \bar{x}_{i\bullet t}, \bar{x}_{\bullet jt})$ are different from those obtained from regressing x_{ijt} on a full set of dummy variables indexing the pairs (i, j), (i, t) and (j, t). Thus in the general design of Eq. (19), the CRE estimator is no longer equivalent to the FE estimator. The equivalence between the Hausman specification test and the modified variable addition test is unlikely to hold.

3. Conclusion

This paper considers the estimation of linear models with multiple fixed effects within the correlated random effects framework. We extend two important equivalence results in one-way panel models to models with multiple fixed effects. We show that the CRE estimator obtained via the use of the Mundlak device is identical to the fixed effects estimator. As a byproduct of the correlated random effects estimation, we demonstrate that a modified variable addition test is identical to the fully robust Hausman specification test. These results are well known in the one-way panel data models but are new in models with multiple fixed effects.

Acknowledgement

This paper is supported by National Natural Science Foundation of China (No. 72103191).

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.econlet.2022.110408.

References

Anderson, J.E., 1979. A theoretical foundation for the gravity equation. Am. Econ. Rev. 69 (1), 106–116.
Arkhangelsky, D., Imbens, G.W., 2019. The Role of the Propensity Score in Fixed Effects Models. CEMFI Working Paper.
Balazsi, L., Matyas, L., Wansbeek, T., 2018. The estimation of multidimensional fixed effects panel data models. Econometric Rev. 37 (3), 212–227.
Balestra, P., Krishnakumar, J., 2008. Fixed effects models and fixed coefficients models. In: Mátyás, L., Sevestre, P. (Eds.), The Econometrics of Panel Data: Fundamentals and Recent Developments in Theory and Practice. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 23–48.
Baltagi, B.H., 2006. An alternative derivation ofMundlak's fixed effects results using system estimation. Econom. Theory 22 (6), 1191–1194.
Baltagi, B.H., Egger, P., Pfaffermayr, M., 2003. A generalized design for bilateral trade flow models. Econom. Lett. 80 (3), 391–397.
Chamberlain, G., 1980. Analysis of covariance with qualitative data. Rev. Econom. Stud. 47 (1), 225–238.
Egger, P., Pfaffermayr, M., 2003. The proper panel econometric specification of the gravity equation: A three-way model with bilateral interaction effects. Empir. Econ. 28 (3), 571–580.
Hausman, J., 1978. Specification tests in econometrics. Econometrica 46 (6), 1251–1271.
Kabir, M., Salim, R., Al-Mawali, N., 2017. The gravity model and trade flows: Recent developments in econometric modeling and empirical evidence. Econ. Anal. Policy 56, 60–71.

Kapetanios, G., Serlenga, L., Shin, Y., 2021. Estimation and inference for multi-dimensional heterogeneous panel datasets with hierarchical multi-factor error structure. J. Econometrics 220 (2), 504–531.
Matyas, L., 1997. Proper econometric specification of the gravity model. World Econ. 20 (3), 363–368.
Matyas, L. (Ed.), 2017. The Econometrics of Multi-Dimensional Panels. Springer-Verlag.
Mundlak, Y., 1978. On the pooling of time series and cross section data. Econometrica 46 (1), 69–85.
Pesaran, M.H., 2006. Estimation and inference in large heterogeneous panels with a multifactor error structure. Econometrica 74 (4), 967–1012.
Wooldridge, J.M., 2010. Econometric Analysis of Cross Section and Panel Data, Vol. 1, second ed. The MIT Press.
Wooldridge, J.M., 2021. Two-Way Fixed Effects, the Two-Way Mundlak Regression, and Difference-in-Differences Estimators. Working Paper.
Yang, Y., Schmidt, P., 2021. An econometric approach to the estimation of multi-level models. J. Econometrics 220 (2), 532–543.