UNSW Business School Research Paper No. 2016 ECON 15

Multidimensional Parameter Heterogeneity in Panel Data Models

Timothy Neal

Multidimensional Parameter Heterogeneity in Panel Data Models

By Timothy Neal *

Draft: October 15, 2016

This article introduces an approach to estimation for static or dynamic panel data models that feature intercept and slope heterogeneity across individuals and over time. It is able to estimate each individual observation coefficient as well as the average coefficient over the sample, and allows for correlation between the heterogeneity and the regressors. Asymptotic theory establishes the consistency and asymptotic normality of the estimates as N and T jointly go to infinity. Finally, Monte Carlo simulations demonstrate that the estimator performs well in environments where fixed effects and mean group estimators are inconsistent and severely biased.

JEL: C13, C22, C23, C33

Keywords: Panel Data; Parameter heterogeneity; Dynamic Panels; Estimation.

Parameter heterogeneity has long been of interest in econometrics, reflecting the inherent instability of economic relationships that can arise from consumer tastes, structural change, aggregation problems, or misspecification. Consider a dynamic panel data model with minimal restrictions on the parameter heterogeneity:

$$y_{it} = \alpha_{it} + \gamma_{it} y_{it-1} + \beta_{it} x_{it} + u_{it} \tag{1}$$

* University of New South Wales, timothy.neal@unsw.edu.au. Thank you to Luke Hartigan, Stéphane Bonhomme, Gerdie Everaert, Seojeong Lee, and Maurice Bun for constructive comments that improved the article.

where $\beta_{it} = \beta + \lambda_i + \lambda_t$, and $\gamma_{it} = \gamma + \delta_i + \delta_t$.

Most of the focus in the panel data literature has been on heterogeneity in the intercept term, here denoted by α_{it}. The fixed effects, random effects, and first difference estimators have been used extensively in applied research to account for individual heterogeneity in the intercept (constraining α_{it} to not vary over time). It is also possible to adapt the fixed effects estimator to consider fixed time effects as well (i.e. the two-way fixed effects estimator). Mesters and Koopman (2014) and Boneva, Linton and Vogt (2015) are two recent examples of panel data estimators that account for both cross-section and time variation in the intercept term.

Heterogeneity in the slope coefficients γ_{it} and β_{it} has also been a topic of great interest. Swamy (1970), which introduces the random coefficients model, considers the case where $\gamma_{it} = 0 \ \forall \ i \ \& \ t$ and β_{it} is restricted to $\beta_i = \lambda + \lambda_i$ where λ_i is a random variable uncorrelated with the regressors and $E(\lambda_i) = 0$. A highly related approach was introduced in Pesaran and Smith (1995) and is called mean group OLS. Regressions are undertaken on each individual to obtain consistent estimates for γ_i and β_i, and these are then averaged to derive a consistent estimate of γ and β (usually as a simple average e.g. $\hat{\beta}_{MG} = N^{-1} \sum_{i=1}^{N} \hat{\beta}_i$). Explicit estimates of β_i can also be of inherent economic interest, for instance when i represents a country or industry. Both models assume that the slope coefficient is constant over time for each individual, where $\lambda_t = 0 \ \forall \ t$.

Meanwhile, time varying parameters have been studied extensively in time series econometrics, usually in the form $\beta_t = \rho\beta_{t-1} + \epsilon_t$ where ϵ_t is a stochastic process and β_t is estimated using a Kalman filter. Pagan (1980) restricts $\rho = 0$ and $E(\epsilon_t) = \bar{\beta}$. Cooley and Prescott (1976), comparatively, sets $\rho = 1$ and $E(\epsilon_t) = 0$. For a useful survey of these panel and time series models see Hsiao and Pesaran (2004). There are some studies that have looked at time varying parameters in panel data as well, such as Degui, Chen and Gao (2011), Lee (2015), and Liu and Hanssens (1981).

Given that parameter heterogeneity can exist across both individuals and time, and can be of significant economic interest in either case, it is reasonable to ask if it is possible to estimate a slope coefficient that varies across both dimensions. Hsiao (1975) was the first to consider a static version of (1) where the individual effects λ_i and the time effects λ_t were both random processes where $E(\lambda_i) = E(\lambda_t) = 0$. Importantly, the model required that the heterogeneity be uncorrelated with the regressors, which is necessarily untrue in dynamic panel models. Baltagi, Feng and Kao (2016) allows for both forms of heterogeneity, again with a static version of (1), but restricts the time heterogeneity to be in the form of a structural break. Hsiao (1974) concludes, with Pesaran (2015) and Balestra (1996) in agreement, that:

> *"If the coefficients of the explanatory variables are fixed and different over time as well as across cross-sectional units, then the parameters to be estimated will increase with the number of sample observations. Not only is there no point at which to pool the data, but there may not exist any consistent estimator at all."*

The purpose of this article is to demonstrate that it is in fact possible to consistently estimate γ_{it} and β_{it}, where the heterogeneity is assumed to be additive (as in (1)) and the panel is moderate to large in both N and T. The approach exploits the ability of large panel data models to pool data across different dimensions in order to triangulate a consistent estimate of a slope coefficient associated with a single observation in the sample. To the author's knowledge this has not yet been considered in the literature.

The result is potentially of significant interest to economists for several reasons. Since time varying parameters and individual varying parameters are often found to exist in time series and panel data applications respectively, a model that is robust to both forms of heterogeneity will likely have wide applicability. The intercept and slope heterogeneity is also allowed to be correlated with the regressors, which is particularly important for dynamic panel data estimation. Most importantly, even if an applied researcher is only interested in the average effect over the sample, fixed effects and mean group estimators will be inconsistent and (potentially severely) biased in this environment.

For the sake of simplicity the technique is developed in an environment that assumes cross-sectional independence (implying additive not interactive fixed effects) and exogenous regressors. However, there is no reason why future work can't weaken these assumptions just as Pesaran (2006) and Chudik and Pesaran (2015) did the same for mean group OLS. Furthermore, the assumption of additive slope heterogeneity may also be relaxed in future work.

The finite sample performance of this technique is tested using Monte Carlo simulations. The results show that in a dynamic panel data model where the slope coefficient varies across both dimensions, the estimator proposed in this paper is consistent where fixed effects and mean group estimators are not. Moreover, the simulations suggest that the estimator is also relatively efficient, preventing a bias/variance trade-off from emerging.

The rest of the article is organised in the following way. Section 1 outlines the estimation problem in a dynamic panel framework. Section 2 presents a consistent estimator for the individual coefficients in this environment. Section 3 presents a consistent estimate for the average coefficient. Section 4 conducts a Monte Carlo simulation study that tests this approach against a number of alternatives under varying assumptions. Section 5 discusses the potential for future extensions of the approach. Section 6 contains some concluding remarks. The Appendix provides proofs of the asymptotic results obtained in Sections 3 and 4.

1. Multidimensional Parameter Heterogeneity in a Dynamic Panel Model

Consider the following estimation problem:

$$y_{it} = \alpha_{it} + \gamma_{it} y_{it-1} + \boldsymbol{x}'_{it} \boldsymbol{\beta}_{it} + u_{it} \tag{2}$$

where $\boldsymbol{x}_{it} = (x_{1it}, x_{2it}, \cdots, x_{Kit})$ is a K x 1 vector of regressors, $\boldsymbol{\beta}_{it} = (\beta_{1it}, \beta_{2it},$

$\cdots$, β_{Kit}) is a K x 1 vector of coefficients that vary across individuals and over time, γ_{it} is the autoregressive coefficient, and u_{it} is the idiosyncratic error term. Further assume that the regressors are driven by an autoregressive process:

$$\boldsymbol{x}_{it} = \boldsymbol{\mu}_i + \boldsymbol{x}_{it-1}\rho_x + \boldsymbol{e}_{it} \tag{3}$$

where $\boldsymbol{\mu}_i = (\mu_{1i}, \mu_{2i}, \cdots, \mu_{Ki})$ and $\boldsymbol{e}_{it} = (e_{1it}, e_{2it}, \cdots, e_{Kit})$ are K x 1 vectors.

The coefficients have the structure:

$$\alpha_{it} = \alpha + c_i + c_t \tag{4}$$

$$\gamma_{it} = \gamma + \delta_i + \delta_t \tag{5}$$

$$\boldsymbol{\beta}_{it} = \boldsymbol{\beta} + \boldsymbol{\lambda}_i + \boldsymbol{\lambda}_t \tag{6}$$

where each possess a constant effect across all observations α, γ, and $\boldsymbol{\beta} = (\beta_1$, β_2, $\cdots$, $\beta_K)$, individual effects that vary across every unit in the panel c_i, δ_i, and $\boldsymbol{\lambda_i} = (\lambda_{1i}$, λ_{2i}, $\cdots$, $\lambda_{Ki})$, and finally time effects that vary between each time period c_t, δ_t, and $\boldsymbol{\lambda_t} = (\lambda_{1t}$, λ_{2t}, $\cdots$, $\lambda_{Kt})$. Accordingly, in this environment there are NT observations in the sample and $(2+K)(NT)$ unique coefficients, where $(2+K)(NT) > NT$ since $K \geq 1$. These coefficients are generated from $(2+K)(N+T)$ unknown parameters. Further lags of y_{it} or lags of $\boldsymbol{x}_{it}$ could be added to (2) without meaningfully altering any of the results of this article.

A standard OLS regression of (2) problem will yield:

$$y_{it} = \alpha + \gamma y_{it-1} + \boldsymbol{x}'_{it}\boldsymbol{\beta} + v_{it} \tag{7}$$

$$v_{it} = c_i + c_t + \delta_i y_{it-1} + \delta_t y_{it-1} + \boldsymbol{x}'_{it}\boldsymbol{\lambda}_i + \boldsymbol{x}'_{it}\boldsymbol{\lambda}_t + u_{it} \tag{8}$$

Simple examination of the pooled estimator reveals multiple sources of potential endogeneity that will lead to bias and inconsistency in the estimates. y_{it-1} and $\boldsymbol{x}_{it}$ may be correlated with c_i and c_t which represent the heterogeneity of the intercept term. The two-way fixed effects estimator (implemented through dummy variables or double-demeaning of the data) has been commonly used to control for this source of endogeneity. y_{it-1} will necessarily be correlated with δ_i and also $\boldsymbol{\lambda}_i$ if $\rho_x \neq 0$. Pesaran and Smith (1995) proposed mean group estimation (or 'MG-OLS') to deal with individual heterogeneity in the slope coefficients. Furthermore, y_{it-1} will be correlated with u_{it} if it possesses serial dependence, and $\boldsymbol{x}_{it}$ will be correlated with u_{it} if it is somehow endogenous. Instrumental variable estimation, such as Difference GMM or System GMM as seen in Arellano and Bond (1991) and Blundell and Bond (1998), can be used to control for the former as well as many forms of the latter. A particular form of endogeneity between $\boldsymbol{x}_{it}$ and u_{it} is called cross section dependence (often represented by unobserved common factors), and that has received attention in Pesaran (2006) and Bai (2009) for the case of static panel data models and Chudik and Pesaran (2015) for dynamic panel data models.

The remaining sources of endogeneity, that to the author's knowledge have not been addressed before in this environment,[1] is potential correlation between y_{it-1} and $\boldsymbol{x}_{it}$ with δ_t and $\boldsymbol{\lambda}_t$. y_{it-1} will be correlated with δ_t and $\boldsymbol{\lambda}_t$ if they contain serial dependence[2], while $\boldsymbol{x}_{it}$ can be correlated with δ_t and $\boldsymbol{\lambda}_t$ for a variety of reasons. For example, the heterogeneity may be correlated with an unobserved covariate which is, in turn, correlated with $\boldsymbol{x}_{it}$.

This article will abstract from sources of endogeneity that have already received a great deal of attention in the literature by making the following assumptions on (2) to ease explanation in the next section. See the fifth section for a brief discussion on weakening some of these assumptions along with others inherent in the formulation of (4) - (6).

Assumption 1: The elements of the regressor term $\boldsymbol{x}_{it}$ have a finite norm, $\|\boldsymbol{\mu}_i\| < R$ and $\|\boldsymbol{e}_{it}\| < R$ for all i and t, where $\|\boldsymbol{A}\|$ refers to the Frobenius norm

[1]Meaning a dynamic panel data model where heterogeneity also exists across panel units.
[2]The time series literature has often found it useful to model time varying heterogeneity with serial dependence, such as $\beta_t = \rho\beta_{t-1} + \epsilon_t$.

of matrix $\boldsymbol{A}$. Further assume that $-1 < \rho_x < 1$.

Assumption 2: The regressors are strictly exogenous with $\boldsymbol{e}_{it}$ distributed independently of $u_{jt'}$ for all i, j, t, and t'.

Assumption 3: The error term u_{it} is independently distributed across individuals and time (i.e. cross-sectionally independent with no serial dependence):

$$E(u_{it}u_{jt'}|\boldsymbol{x}_{it}) = 0, \text{ for all } i, j, t, \text{ and } t'$$

Assumption 4: The heterogeneous effects all have a finite norm where $||\boldsymbol{\lambda}_i|| < R$, $||\boldsymbol{\lambda}_t|| < R$, $||\delta_i|| < R$, $||\delta_t|| < R$, $||c_i|| < R$, and $||c_t|| < R$ for all i and t and some constant $R < \infty$. Furthermore, assume that the support of γ_{it} lies strictly within the unit circle.

2. A Consistent Estimate of the Individual Coefficients

The aim of this section is to obtain a consistent estimate of $\boldsymbol{\theta}_{it} = (\alpha_{it}, \gamma_{it}, \boldsymbol{\beta}_{it})$, while the next section will consider a consistent estimate of the average effect defined as $\bar{\boldsymbol{\theta}} = (\alpha + E(c_i) + E(c_t), \gamma + E(\delta_i) + E(\delta_t), \boldsymbol{\beta} + E(\boldsymbol{\lambda}_i) + E(\boldsymbol{\lambda}_t))$. Further define $\boldsymbol{\theta} = (\alpha, \gamma, \boldsymbol{\beta})$ as containing the constant effects, $\boldsymbol{\theta}_i = (c_i, \delta_i, \boldsymbol{\lambda}_i)$ as containing the individual effects, $\boldsymbol{\theta}_t = (c_t, \delta_t, \boldsymbol{\lambda}_t)$ as containing the time effects, and finally $\boldsymbol{z}_{it} = (1, y_{it-1}, \boldsymbol{x}_{it})$ as the set of regressors.[3] All of these vectors are $(K+2)$ x 1 in dimension. (2) can now be rewritten as:

$$y_{it} = \boldsymbol{z}'_{it}\boldsymbol{\theta} + v_{it}$$

$$v_{it} = \boldsymbol{z}'_{it}\boldsymbol{\theta}_i + \boldsymbol{z}'_{it}\boldsymbol{\theta}_t + u_{it}$$

Consider first the pooled OLS estimator of $\boldsymbol{\theta}$:

$$\hat{\boldsymbol{\theta}} = \left(\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\boldsymbol{z}_{it}\boldsymbol{z}'_{it}\right)^{-1}\left(\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\boldsymbol{z}_{it}y_{it}\right) \tag{9}$$

[3]Note that the constant term has been included in the set of regressors so that the fixed effects in the intercept will be treated together with the fixed effects of the slope coefficients.

Expanding on y_{it} and simplifying yields:

$$\begin{aligned}\hat{\boldsymbol{\theta}} =& \boldsymbol{\theta} + \boldsymbol{Q}_{zz,NT}^{-1}\left(\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\boldsymbol{z}_{it}\boldsymbol{z}_{it}'\boldsymbol{\theta}_i\right) \\ &+ \boldsymbol{Q}_{zz,NT}^{-1}\left(\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\boldsymbol{z}_{it}\boldsymbol{z}_{it}'\boldsymbol{\theta}_t\right) + \boldsymbol{Q}_{zz,NT}^{-1}\left(\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\boldsymbol{z}_{it}u_{it}\right)\end{aligned} \tag{10}$$

where $\boldsymbol{Q}_{zz,NT}^{-1} = \left(\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\boldsymbol{z}_{it}\boldsymbol{z}_{it}'\right)^{-1}$. Next, consider a series of regressions for each individual i:

$$y_{it} = \boldsymbol{z}_{it}'(\boldsymbol{\theta} + \boldsymbol{\theta}_i) + v_{it}$$

$$v_{it} = \boldsymbol{z}_{it}'\boldsymbol{\theta}_t + u_{it}$$

The resulting OLS estimates will yield:

$$\hat{\boldsymbol{\theta}}_i = \left(\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{z}_{it}\boldsymbol{z}_{it}'\right)^{-1}\left(\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{z}_{it}y_{it}\right) \tag{11}$$

Expanding on y_{it} and noting that $\boldsymbol{\theta}_i$ is now a scalar vector:

$$\hat{\boldsymbol{\theta}}_i = \boldsymbol{\theta} + \boldsymbol{\theta}_i + \boldsymbol{Q}_{zz,T}^{-1}\left(\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{z}_{it}\boldsymbol{z}_{it}'\boldsymbol{\theta}_t\right) + \boldsymbol{Q}_{zz,T}^{-1}\left(\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{z}_{it}u_{it}\right) \tag{12}$$

where $\boldsymbol{Q}_{zz,T}^{-1} = \left(\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{z}_{it}\boldsymbol{z}_{it}'\right)^{-1}$. Next, consider a series of regressions for each time period t:

$$y_{it} = \boldsymbol{z}_{it}'(\boldsymbol{\theta} + \boldsymbol{\theta}_t) + v_{it}$$

$$v_{it} = \boldsymbol{z}_{it}'\boldsymbol{\theta}_i + u_{it}$$

The estimates will be a mirror of those found in (12):

$$\hat{\boldsymbol{\theta}}_t = \boldsymbol{\theta} + \boldsymbol{\theta}_t + \boldsymbol{Q}_{zz,N}^{-1}\left(\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{z}_{it}\boldsymbol{z}_{it}'\boldsymbol{\theta}_i\right) + \boldsymbol{Q}_{zz,N}^{-1}\left(\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{z}_{it}u_{it}\right) \tag{13}$$

where $\boldsymbol{Q}_{zz,N}^{-1} = \left(\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{z}_{it}\boldsymbol{z}_{it}'\right)^{-1}$. In order to obtain a preliminary estimate of $\boldsymbol{\theta}_{it}$, simply exploit the relations found in (10), (12), and (13) and apply them to each observation in the sample:

$$\begin{aligned}
\hat{\boldsymbol{\theta}}_{it} =& \hat{\boldsymbol{\theta}}_i + \hat{\boldsymbol{\theta}}_t - \hat{\boldsymbol{\theta}} = \\
& \boldsymbol{\theta} + \boldsymbol{\theta}_i + \boldsymbol{Q}_{zz,T}^{-1}\left(\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{z}_{it}\boldsymbol{z}_{it}'\boldsymbol{\theta}_t\right) + \boldsymbol{Q}_{zz,T}^{-1}\left(\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{z}_{it}u_{it}\right) + \\
& \boldsymbol{\theta} + \boldsymbol{\theta}_t + \boldsymbol{Q}_{zz,N}^{-1}\left(\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{z}_{it}\boldsymbol{z}_{it}'\boldsymbol{\theta}_i\right) + \boldsymbol{Q}_{zz,N}^{-1}\left(\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{z}_{it}u_{it}\right) - \\
& \boldsymbol{\theta} - \boldsymbol{Q}_{zz,NT}^{-1}\left(\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\boldsymbol{z}_{it}\boldsymbol{z}_{it}'\boldsymbol{\theta}_i\right) \\
& - \boldsymbol{Q}_{zz,NT}^{-1}\left(\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\boldsymbol{z}_{it}\boldsymbol{z}_{it}'\boldsymbol{\theta}_t\right) - \boldsymbol{Q}_{zz,NT}^{-1}\left(\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\boldsymbol{z}_{it}u_{it}\right)
\end{aligned} \tag{14}$$

This simplifies to:

$$\begin{aligned}
\hat{\boldsymbol{\theta}}_{it} =& \boldsymbol{\theta} + \boldsymbol{\theta}_i + \boldsymbol{\theta}_t + (\boldsymbol{R}_N\boldsymbol{\theta}_i - \boldsymbol{R}_{NT}\boldsymbol{\theta}_i) + (\boldsymbol{R}_T\boldsymbol{\theta}_t - \boldsymbol{R}_{NT}\boldsymbol{\theta}_t) + \\
& \left(\boldsymbol{Q}_{zz,N}^{-1}\boldsymbol{Q}_{zu,N} + \boldsymbol{Q}_{zz,T}^{-1}\boldsymbol{Q}_{zu,T} - \boldsymbol{Q}_{zz,NT}^{-1}\boldsymbol{Q}_{zu,NT}\right)
\end{aligned} \tag{15}$$

where $\boldsymbol{R}_N = \boldsymbol{Q}_{zz,N}^{-1}\boldsymbol{Q}_{zz,N}$ and similarly for $\boldsymbol{R}_{NT}$ and $\boldsymbol{R}_T$, and also $\boldsymbol{Q}_{zu,N} = \frac{1}{N}\sum_{i=1}^{N}\boldsymbol{z}_{it}u_{it}$ and similarly for $\boldsymbol{Q}_{zu,T}$ and $\boldsymbol{Q}_{zu,NT}$. In addition to the biases related to the error term, there is a bias originating from correlation between the regressors and the individual or time heterogeneity (this includes the fixed effects of the intercept term). Because there are two dimensions of heterogeneity in a two-dimension panel, it is impossible to remove this bias by pooling the data in different ways (as MG-OLS is able to do when the heterogeneity is only over the i dimension).

In order to remove the heterogeneity bias this article proposes using $\hat{\boldsymbol{\theta}}_i$ as a sample approximation for $\boldsymbol{\theta}_i$ in $(\boldsymbol{R}_N\boldsymbol{\theta}_i - \boldsymbol{R}_{NT}\boldsymbol{\theta}_i)$ and $\hat{\boldsymbol{\theta}}_t$ as a sample approximation for $\boldsymbol{\theta}_t$ in $(\boldsymbol{R}_T\boldsymbol{\theta}_t - \boldsymbol{R}_{NT}\boldsymbol{\theta}_t)$. Inserting (12) and (13) into these parts of (15) yields:

$$
\begin{aligned}
&(\boldsymbol{R}_N\hat{\boldsymbol{\theta}}_i - \boldsymbol{R}_{NT}\hat{\boldsymbol{\theta}}_i) + (\boldsymbol{R}_T\hat{\boldsymbol{\theta}}_t - \boldsymbol{R}_{NT}\hat{\boldsymbol{\theta}}_t) = (\boldsymbol{R}_N\boldsymbol{\theta}_i - \boldsymbol{R}_{NT}\boldsymbol{\theta}_i) + (\boldsymbol{R}_T\boldsymbol{\theta}_t - \boldsymbol{R}_{NT}\boldsymbol{\theta}_t)\\
&+ (\boldsymbol{R}_N\boldsymbol{R}_T\boldsymbol{\theta}_t - \boldsymbol{R}_{NT}\boldsymbol{R}_T\boldsymbol{\theta}_t) + (\boldsymbol{R}_T\boldsymbol{R}_N\boldsymbol{\theta}_i - \boldsymbol{R}_{NT}\boldsymbol{R}_N\boldsymbol{\theta}_i)+\\
&(\boldsymbol{R}_N\boldsymbol{Q}_{zz,N}^{-1}\boldsymbol{Q}_{zu,N} - \boldsymbol{R}_{NT}\boldsymbol{Q}_{zz,NT}^{-1}\boldsymbol{Q}_{zu,NT}) + (\boldsymbol{R}_T\boldsymbol{Q}_{zz,T}^{-1}\boldsymbol{Q}_{zu,T} - \boldsymbol{R}_{NT}\boldsymbol{Q}_{zz,NT}^{-1}\boldsymbol{Q}_{zu,NT})
\end{aligned}
\tag{16}
$$

While this correction will contain the original bias, it will also introduce further biases, some of which relate to the idiosyncratic error term while others the slope heterogeneity. In order to approximate the additional biases relating to the slope heterogeneity, this article proposes inserting $\hat{\boldsymbol{\theta}}_i$ into $(\boldsymbol{R}_T\boldsymbol{R}_N\boldsymbol{\theta}_i - \boldsymbol{R}_{NT}\boldsymbol{R}_N\boldsymbol{\theta}_i)$ and $\hat{\boldsymbol{\theta}}_t$ into $(\boldsymbol{R}_N\boldsymbol{R}_T\boldsymbol{\theta}_t - \boldsymbol{R}_{NT}\boldsymbol{R}_T\boldsymbol{\theta}_t)$. In fact, this process can be repeated L times to form the bias corrected estimates:

$$
\begin{aligned}
\hat{\boldsymbol{\theta}}_{it}^{BC} = \hat{\boldsymbol{\theta}}_i + \hat{\boldsymbol{\theta}}_t - \hat{\boldsymbol{\theta}} + \sum_{\ell=0}^{L}(-1)^{\ell+1}((\boldsymbol{R}_N\Gamma_{1,\ell} - \boldsymbol{R}_{NT}\Gamma_{1,\ell})+\\
(\boldsymbol{R}_T\Gamma_{2,\ell} - \boldsymbol{R}_{NT}\Gamma_{2,\ell}))
\end{aligned}
\tag{17}
$$

where $\Gamma_{1,\ell} = \boldsymbol{R}_T\Gamma_{2,\ell-1}$ and $\Gamma_{2,\ell} = \boldsymbol{R}_N\Gamma_{1,\ell-1}$ when $\ell > 0$, $\Gamma_{1,0} = \hat{\boldsymbol{\theta}}_i$, and finally $\Gamma_{2,0} = \hat{\boldsymbol{\theta}}_t$. The proof of Theorem 1 in the Appendix will demonstrate that when L is sufficiently large, this bias correction will approximate the true heterogeneity bias. In practice, since the magnitude of the adjustment declines as ℓ increases, a suitable L can be determined endogenously by programming the sum to stop once the bias adjustment converges to a given level of tolerance. Section 4 will show its finite sample performance and the number of 'iterations' required to achieve convergence in a number of scenarios.

Theorem 1 provides the result of asymptotic consistency for the individual observation coefficients as L first goes to infinity, and then both N and T jointly go to infinity. The proof can be found in the Appendix.

Theorem 1: Consistency of $\hat{\boldsymbol{\theta}}_{it}^{BC}$
For the panel model outlined in (2) - (6) where Assumptions 1-4 hold, when $L \to \infty$ and then $(N,T) \xrightarrow{j} \infty$ it is true that:

$$\hat{\boldsymbol{\theta}}_{it}^{BC} - \boldsymbol{\theta}_{it} \xrightarrow{p} 0 \tag{18}$$

3. A Consistent Estimate of the Average Coefficient

Applied researchers may be exclusively interested in the average relationship between variables over the sample: $\bar{\boldsymbol{\theta}} = (\alpha + E(c_i) + E(c_t), \gamma + E(\delta_i) + E(\delta_t), \boldsymbol{\beta} + E(\boldsymbol{\lambda}_i) + E(\boldsymbol{\lambda}_t))$. Possessing a consistent estimate for each individual coefficient, the average coefficient can be easily constructed by taking the simple average of these individual coefficients:

$$\hat{\boldsymbol{\theta}}_{MO} = \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \hat{\boldsymbol{\theta}}_{it}^{BC} \tag{19}$$

For ease of reference this average is referred to as Mean Observation OLS (or 'MO-OLS'). The following two theorems provide the results for asymptotic consistency and also asymptotic normality.

Theorem 2: Consistency of $\hat{\boldsymbol{\theta}}_{MO}$
For the panel model outlined in (2) - (6) where Assumptions 1-4 hold, when $L \to \infty$ and then $(N,T) \xrightarrow{j} \infty$ it is true that:

$$\hat{\boldsymbol{\theta}}_{MO} - \bar{\boldsymbol{\theta}} \xrightarrow{p} 0 \tag{20}$$

Theorem 3: Asymptotic Normality of $\hat{\boldsymbol{\theta}}_{MO}$

For the panel model outlined in (2) - (6) where Assumptions 1-4 hold, when $L \to \infty$ and then $(N,T) \xrightarrow{j} \infty$ such that $N/T \to \chi$ and $\chi > 0$ it is true that:

$$\sqrt{NT}(\hat{\boldsymbol{\theta}}_{MO} - \bar{\boldsymbol{\theta}}) \xrightarrow{d} N(0, \boldsymbol{\Sigma}_{MO}) \tag{21}$$

where $\boldsymbol{\Sigma}_{MO}$ is defined in (A14). The asymptotic variance can be consistently estimated nonparametrically by:

$$\begin{aligned}\hat{\boldsymbol{\Sigma}}_{MO} = \frac{1}{NT-1}\sum_{i=1}^{N}\sum_{t=1}^{T}\Big((\hat{\boldsymbol{\theta}}_{\bar{i}} - \hat{\boldsymbol{\theta}}_{MO})(\hat{\boldsymbol{\theta}}_{\bar{i}} - \hat{\boldsymbol{\theta}}_{MO})' + (\hat{\boldsymbol{\theta}}_{\bar{t}} - \hat{\boldsymbol{\theta}}_{MO})(\hat{\boldsymbol{\theta}}_{\bar{t}} - \hat{\boldsymbol{\theta}}_{MO})' \\ + 2(\hat{\boldsymbol{\theta}}_{\bar{i}} - \hat{\boldsymbol{\theta}}_{MO})(\hat{\boldsymbol{\theta}}_{\bar{t}} - \hat{\boldsymbol{\theta}}_{MO})'\Big)\end{aligned} \tag{22}$$

where $\hat{\boldsymbol{\theta}}_{\bar{i}} = \frac{1}{T}\sum_{t=1}^{T}\hat{\boldsymbol{\theta}}_{it}^{BC}$ and $\hat{\boldsymbol{\theta}}_{\bar{t}} = \frac{1}{N}\sum_{i=1}^{N}\hat{\boldsymbol{\theta}}_{it}^{BC}$. Restrictions on the relative rate of convergence of N and T are required due to the presence of a small sample time series bias $O(T^{-1})$ that has been well documented in the literature (starting with Hurwicz (1950)) and originates from the inclusion of a lagged dependent variable. Accordingly, the estimator is not appropriate for panels with very large N and small T, although adopting existent small T bias corrections for this estimator is an area for future research.

4. Monte Carlo Simulations

This section conducts a Monte Carlo simulation study to determine the finite sample performance of the MO-OLS estimator of the average coefficient vector $\bar{\boldsymbol{\theta}}$ proposed in the third section against the closest alternatives that are currently available: the one-way and two-way fixed effects estimators and mean group OLS ('MG-OLS'). A number of scenarios are constructed from a data generating process that features a large panel data structure and multidimensional slope heterogeneity and fixed effects.

The chief differentiation between scenarios will be the inclusion of a lagged dependent variable, the degree of correlation between the unobserved heterogeneity and the regressor, and lastly the generating process of the slope heterogeneity. Each estimator will be tested according to its mean bias and empirical standard deviation, while the number of simulation repetitions has been set to 1,000 for this study.

4.1. Data Generating Process

The dependent variable is defined by:

$$y_{it} = c_{it} + \gamma_{it} y_{it-1} + \beta_{it} x_{it} + \epsilon_{it} \tag{23}$$

where $i = 1, 2, ..., N$ and $t = -10, ..., 0, 1, ..., T$ with the first 10 observations of each i discarded prior to estimation. In all scenarios, we generate heterogeneous coefficients $\beta_{it} = \beta + \lambda_i + \lambda_t$ where $\beta = 1$, $\lambda_i \sim N(0, 0.353)$ and λ_t will vary between scenarios. The heterogeneous autoregressive term will be $\gamma_{it} = \gamma + \delta_i + \delta_t$, where $\gamma = 0$ in the static scenarios and $\gamma = 0.5$ in the dynamic scenarios. δ_i and δ_t will vary between scenarios. In all scenarios an unbias estimate of the average β_{it} over all Monte Carlo repetitions will be approximately equal to 1, while an unbias estimate of the average γ_{it} will be approximately equal to 0.5 (in the scenarios featuring a lagged dependent variable).

Since complexity in the idiosyncratic error term is not a focus of this article, it will be set simply to $\epsilon_{it} \sim N(0, 1)$. The fixed effects are generated by $c_{it} = c + c_i + c_t$ where $c = 1$, $c_i \sim N(0, 0.353)$ and c_t will vary between scenarios. The regressor takes the following form:

$$x_{it} = \rho x_{it-1} + \alpha_1 c_{it} + \alpha_2(\gamma_{it} + \beta_{it}) + e_{it} \tag{24}$$

where $\rho = 0.5$, α_1 measures the degree of correlation between the fixed effects c_{it} and the regressor, α_2 measures the degree of correlation between the slope coefficients γ_{it} and β_{it} with the regressor, and $e_{it} \sim N(0,1)$. In all scenarios $\alpha_1 = 1$ while α_2 will vary between scenarios.

Six scenarios are considered and are outlined in Table 1. 'Random' heterogeneity in the table refers to the random process $N(0, \sigma^2)$, where $\sigma^2 = 0.104$ for the autoregressive heterogeneity δ_i and δ_t, while $\sigma^2 = 0.353$ for the regressor heterogeneity λ_t and also for the intercept heterogeneity c_t. 'Dependent' heterogeneity adds serial dependence to the random heterogeneity, with $\delta_t = 0.5(\delta_{t-1}) + N(0, 0.104)$ and $\lambda_t = 0.5(\lambda_{t-1}) + N(0, 0.353)$. 'Fixed' heterogeneity refers to a fixed process where:

$$\gamma_i = \begin{cases} -0.104 & \text{if } i < N/2 \\ 0.104 & \text{if } i \geq N/2 \end{cases}$$

and:

$$\lambda_t = \begin{cases} -0.353 & \text{if } t < T/2 \\ 0.353 & \text{if } t \geq T/2 \end{cases}$$

The first scenario is a static panel data model that will serve as a benchmark for relative efficiency in a simple setting. The second scenario introduces correlation between the regressor and the slope heterogeneity through α_2. The third scenario is a dynamic panel data model that features heterogeneity in the slope coefficients and fixed effects across the i dimension, but not along the t dimension. The purpose of this scenario is again to test the efficiency losses from using the technique proposed in this article on data that is simpler than its intended purpose. The fourth scenario features heterogeneity across both the i and t dimensions. Both dimensions of heterogeneity are randomly generated but are also correlated with the regressor term through α_2. The fifth scenario is identical to the fourth except that the time heterogeneity of the slope coefficients possess serial dependence, and the sixth scenario features time heterogeneity that is fixed and different.

Table 1—Scenario Design

Scenario	1	2	3	4	5	6
γ	0	0	0.5	0.5	0.5	0.5
α_2	0	1	1	1	1	1
δ_i	N/A	N/A	Random	Random	Random	Random
δ_t	N/A	N/A	N/A	Random	Dependent	Fixed
λ_t	Random	Random	N/A	Random	Dependent	Fixed
c_t	Random	Random	N/A	Random	Random	Random

4.2. Results

The results for the first scenario are presented in Table 2. The mean coefficient is reported in the left panel, while the right panel lists the empirical standard deviation for both the true values[4] and the four estimators under consideration. N is held at 50 in all scenarios while T varies between 30 and 200, since it was found that variation in T is the main driver of variation in performance due to the presence of the $O(T^{-1})$ bias. This scenario presents a static panel data model that does not contain correlation between the multidimensional slope heterogeneity and the regressor, but does possess fixed effects across both dimensions that are correlated with the regressor.

The results show that ignoring the time effects introduces some bias in one-way FE and MG-OLS. The preliminary MO-OLS estimates (defined in (15)) are also biased due to the correlation with the fixed effects. Performing the bias correction on the MO-OLS estimates successfully removes all bias as does using the two-way FE estimator. The second scenario introduces correlation between the slope heterogeneity and the single regressor, and the results can be seen in Table 3. The bias found in the one-way FE estimator and MG-OLS becomes very severe. It demonstrates that ignoring heterogeneity in the slope coefficients can have dramatic implications for statistical inference. The performance of the MO-OLS estimator is identical to the previous scenario, with the bias correction

[4]Since the average slope coefficients vary between simulation repetitions due to them being a function of random variables, the empirical standard deviation of the true coefficients serve as a natural benchmark for the efficiency of other estimators.

procedure requiring an average of three additional iterations to converge. Unpublished simulation results show that increasing the value of α_2 in this scenario introduces some bias to the two-way FE estimator and also significantly worsens the standard deviation relative to MO-OLS.[5]

Table 2—Simulation Results - Scenario 1

$(N = 50, T)$	**Mean**					**Std. Dev.**				
	30	50	70	100	200	30	50	70	100	200
Results for $\bar{\beta}$										
True Values	0.998	1.003	0.999	0.999	0.998	0.082	0.071	0.064	0.060	0.057
Pooled OLS										
One-way FE	1.077	1.087	1.087	1.080	1.080	0.208	0.167	0.144	0.122	0.095
Two-way FE	0.999	1.003	0.999	0.998	0.999	0.090	0.076	0.067	0.063	0.058
MG-OLS	1.079	1.088	1.087	1.081	1.080	0.211	0.168	0.144	0.122	0.095
MO-OLS										
Prelim.	1.039	1.040	1.038	1.034	1.033	0.112	0.092	0.078	0.074	0.065
Bias Corr.	0.998	1.002	0.999	0.998	0.999	0.087	0.074	0.066	0.062	0.058

Note: 1,000 Monte Carlo Simulations with N=50 and varied T.

The results for the third scenario are presented in Table 4. The third scenario features a lagged dependent variable, but restricts heterogeneity in the slope coefficient and fixed effects to be across the i dimension alone. The purpose of doing this is to determine whether using MO-OLS is inefficient when MG-OLS will perform fine. Indeed, the results show that MG-OLS is unbiased when T is large (the typical small T time series bias in dynamic panel data models is still present). The one-way and two-way FE models are both biased and inconsistent as the sample size grows. MO-OLS performs virtually identically to MG-OLS in terms of bias and efficiency. Furthermore, the results show that the bias correction procedure does not introduce any additional inefficiency when it is unnecessarily

[5]The results of this exercise are available upon request.

Table 3—Simulation Results - Scenario 2

$(N = 50, T)$	**Mean**					**Std. Dev.**				
	30	50	70	100	200	30	50	70	100	200
Results for $\bar{\beta}$										
True Values	0.998	1.003	1.000	0.998	1.002	0.082	0.071	0.066	0.059	0.056
Pooled OLS										
One-way FE	1.981	1.988	1.977	1.970	1.981	0.302	0.236	0.201	0.167	0.130
Two-way FE	0.999	1.003	1.001	0.998	1.002	0.090	0.077	0.071	0.062	0.058
MG-OLS	2.012	2.007	1.991	1.980	1.985	0.310	0.240	0.204	0.168	0.130
MO-OLS										
Prelim.	1.623	1.604	1.590	1.572	1.572	0.192	0.156	0.133	0.120	0.103
Bias Corr.	0.998	1.002	1.001	0.998	1.002	0.087	0.074	0.069	0.061	0.057

Note: 1,000 Monte Carlo Simulations with N=50 and varied T.

applied to a model. This suggests that MO-OLS may be able to be applied to models at virtually no cost when the underlying structure of the heterogeneity is difficult to ascertain.

The results for the fourth scenario are presented in Table 5. In this scenario, heterogeneity in the intercept term and slope coefficients are extended to vary across both the i and t dimensions. The heterogeneity is random across both dimensions and there is correlation between the regressors and the heterogeneity. Both forms of the FE estimator and MG-OLS show strong bias (that does not decline as T increases, with the exception of the autoregressive term for MG-OLS) and inefficiency. The preliminary MO-OLS estimates are also strongly biased. Applying the bias correction successfully removes all of the bias when T is moderate to large, and also significantly decreases the standard deviation of the estimate. It is unbiased in $\bar{\gamma}$ when $T > 70$, and unbiased in $\bar{\beta}$ when $T > 30$. Also worth noting is that the number of iterations required to achieve convergence in the bias correction has significantly increased, with over 100 iterations required in most MC iterations. Since each iteration is computationally cheap, this has

Table 4—Simulation Results - Scenario 3

$(N = 50, T)$	Mean					Std. Dev.				
	30	50	70	100	200	30	50	70	100	200
Results for $\bar{\gamma}$										
True Values	0.500	0.501	0.501	0.500	0.501	0.015	0.015	0.015	0.015	0.015
Pooled OLS										
One-way FE	0.531	0.543	0.545	0.549	0.552	0.045	0.033	0.027	0.031	0.026
Two-way FE	0.531	0.542	0.545	0.549	0.552	0.045	0.033	0.027	0.031	0.026
MG-OLS	0.458	0.477	0.484	0.488	0.495	0.023	0.019	0.017	0.017	0.016
MO-OLS										
Prelim.	0.459	0.478	0.485	0.489	0.496	0.023	0.019	0.017	0.017	0.016
Bias Corr.	0.457	0.476	0.483	0.487	0.495	0.023	0.019	0.017	0.017	0.016
Results for $\bar{\beta}$										
True Values	1.000	1.001	1.000	1.001	0.999	0.050	0.049	0.048	0.050	0.051
Pooled OLS										
One-way FE	0.983	0.979	0.977	0.974	0.971	0.059	0.054	0.053	0.052	0.052
Two-way FE	0.983	0.979	0.977	0.974	0.971	0.060	0.054	0.053	0.053	0.052
MG-OLS	1.019	1.013	1.009	1.007	1.003	0.056	0.053	0.051	0.052	0.051
MO-OLS										
Prelim.	1.006	1.001	0.998	0.996	0.992	0.055	0.053	0.051	0.052	0.051
Bias Corr.	1.020	1.013	1.009	1.007	1.003	0.056	0.053	0.051	0.052	0.051
Iterations	26	22	20	20	19	14	10	9	10	10

Note: 1,000 Monte Carlo Simulations with N=50 and varied T.

little significance in applied situations (at least at these sample sizes).

Table 5—Simulation Results - Scenario 4

$(N = 50, T)$	Mean					Std. Dev.				
	30	50	70	100	200	30	50	70	100	200
Results for $\bar{\gamma}$										
True Values	0.500	0.500	0.500	0.500	0.501	0.025	0.022	0.020	0.018	0.016
Pooled OLS										
One-way FE	0.412	0.450	0.462	0.477	0.491	0.146	0.116	0.104	0.085	0.067
Two-way FE	0.597	0.629	0.639	0.653	0.664	0.098	0.085	0.078	0.068	0.063
MG-OLS	0.356	0.393	0.407	0.420	0.434	0.138	0.107	0.094	0.075	0.055
MO-OLS										
Prelim.	0.526	0.560	0.574	0.583	0.599	0.123	0.094	0.084	0.071	0.052
Bias Corr.	0.474	0.485	0.490	0.493	0.497	0.030	0.024	0.022	0.019	0.017
Results for $\bar{\beta}$										
True Values	0.998	1.003	1.000	0.998	1.002	0.082	0.071	0.066	0.059	0.056
Pooled OLS										
One-way FE	2.381	2.370	2.343	2.325	2.327	0.465	0.361	0.304	0.250	0.194
Two-way FE	0.949	0.932	0.920	0.906	0.900	0.109	0.095	0.083	0.073	0.066
MG-OLS	2.541	2.492	2.452	2.421	2.410	0.493	0.372	0.312	0.255	0.191
MO-OLS										
Prelim.	1.139	1.089	1.056	1.043	1.016	0.274	0.210	0.192	0.170	0.145
Bias Corr.	1.010	1.010	1.007	1.002	1.004	0.087	0.075	0.069	0.061	0.057
Iterations	356	335	325	320	313	82	62	56	47	39

Note: 1,000 Monte Carlo Simulations with N=50 and varied T.

The results for the fifth scenario are presented in Table 6. In this scenario, the slope heterogeneity across t is generated with dependence (see Table 1 for details). The results show very severe bias for all estimators considered here save the bias corrected MO-OLS estimates. The autoregressive parameter $\bar{\gamma}$ is significantly overestimated, while the regressor coefficient is also overestimated by more than 100% (save for the two-way FE estimator). The bias corrected MO-OLS estimates successfully removes all bias at moderate to large T, and is able to estimate both parameters very efficiently. The number of iterations required for the bias correction procedure to converge increased as well.

Table 6—Simulation Results - Scenario 5

$(N = 50, T)$	**Mean**					**Std. Dev.**				
	30	50	70	100	200	30	50	70	100	200
Results for $\bar{\gamma}$										
True Values	0.499	0.500	0.500	0.499	0.500	0.042	0.033	0.030	0.026	0.022
Pooled OLS										
One-way FE	0.668	0.706	0.717	0.728	0.747	0.109	0.086	0.074	0.061	0.048
Two-way FE	0.763	0.800	0.809	0.818	0.839	0.097	0.079	0.070	0.061	0.048
MG-OLS	0.616	0.654	0.666	0.678	0.694	0.104	0.080	0.069	0.055	0.040
MO-OLS										
Prelim.	0.746	0.778	0.789	0.800	0.812	0.088	0.064	0.055	0.044	0.033
Bias Corr.	0.474	0.485	0.490	0.492	0.497	0.046	0.035	0.031	0.027	0.022
Results for $\bar{\beta}$										
True Values	0.997	1.004	1.001	0.999	1.003	0.137	0.110	0.097	0.084	0.073
Pooled OLS										
One-way FE	2.356	2.306	2.272	2.244	2.224	0.586	0.445	0.376	0.307	0.231
Two-way FE	0.862	0.836	0.819	0.805	0.788	0.148	0.120	0.099	0.085	0.071
MG-OLS	2.609	2.513	2.460	2.415	2.387	0.628	0.466	0.393	0.315	0.233
MO-OLS										
Prelim.	1.332	1.247	1.199	1.166	1.134	0.298	0.212	0.191	0.154	0.134
Bias Corr.	1.008	1.010	1.007	1.003	1.005	0.140	0.113	0.099	0.085	0.074
Iterations	457	434	422	415	409	125	101	88	77	72

Note: 1,000 Monte Carlo Simulations with N=50 and varied T.

The results for the sixth scenario are presented in Table 7. In this scenario, the slope heterogeneity across t is fixed and different over the sample (see Table 1 for details). More than any other, this scenario demonstrates how badly existent estimators can overestimate the parameter values when the underlying data contains heterogeneity over time that is correlated with one of the regressors. The estimates for $\bar{\gamma}$ are very close to 1 for both FE estimators, MG-OLS, and also the preliminary MO-OLS estimates. Fortunately, the bias corrected MO-OLS estimates are consistent and perform as well as in preceding scenarios. This suggests that the performance of the estimator is almost independent of the underlying complexity of the slope heterogeneity, nor the presence of correlation between the regressors and that heterogeneity.

To summarise, the Monte Carlo simulation study has successfully demonstrated that the bias corrected MO-OLS estimator is able to consistently and efficiently estimate the mean coefficient in data containing intercept and slope heterogeneity over both dimensions of the panel. It is able to achieve this even when the heterogeneity is correlated with the regressors, and when it is not randomly generated. Relying on existent estimators in these environments can potentially lead to severely overestimated autoregressive parameters and also slope coefficients. Given that the results also suggest that MO-OLS can potentially be applied with virtually no cost (relative to MG-OLS) when the data features heterogeneity across i alone, it would appear that MO-OLS is suitable for most empirical applications using dynamic panel data models provided that both N and T is moderate to large in size.

5. Future Work

MG-OLS has received a number of extensions in order to generalise the estimator, perhaps most importantly to incorporate cross-section dependence in the data. This may appear in the data through interactive fixed effects (as in Bai (2009)) or unobserved common factors (as in Pesaran (2006) and Chudik and Pesaran (2015)). Further extensions also allow for serial dependence in the error term and also endogeneity in the regressors as in Harding and Lamarche (2011).

Table 7—Simulation Results - Scenario 6

$(N = 50, T)$	Mean					Std. Dev.				
	30	50	70	100	200	30	50	70	100	200
Results for $\bar{\gamma}$										
True Values	0.499	0.500	0.501	0.500	0.499	0.016	0.016	0.015	0.015	0.015
Pooled OLS										
One-way FE	0.950	0.956	0.961	0.963	0.967	0.017	0.012	0.012	0.011	0.009
Two-way FE	0.991	0.983	0.982	0.981	0.981	0.020	0.015	0.014	0.013	0.011
MG-OLS	0.920	0.930	0.936	0.941	0.947	0.016	0.010	0.008	0.006	0.004
MO-OLS										
Prelim.	0.944	0.943	0.943	0.943	0.943	0.011	0.008	0.007	0.006	0.005
Bias Corr.	0.461	0.478	0.484	0.488	0.493	0.023	0.020	0.018	0.018	0.016
Results for $\bar{\beta}$										
True Values	0.999	0.999	0.998	0.998	1.003	0.051	0.051	0.051	0.049	0.049
Pooled OLS										
One-way FE	1.310	1.098	0.992	0.919	0.829	0.213	0.150	0.128	0.107	0.080
Two-way FE	0.784	0.755	0.740	0.727	0.718	0.069	0.055	0.050	0.042	0.039
MG-OLS	1.571	1.309	1.186	1.094	0.978	0.221	0.156	0.117	0.089	0.058
MO-OLS										
Prelim.	1.170	1.044	0.992	0.948	0.901	0.118	0.093	0.085	0.071	0.058
Bias Corr.	1.010	1.008	1.006	1.003	1.006	0.059	0.054	0.054	0.050	0.050
Iterations	820	839	881	905	932	232	214	222	210	214

Note: 1,000 Monte Carlo Simulations with N=50 and varied T.

Future work may be able to extend MO-OLS in similar directions, using modified versions of the techniques introduced in the papers referenced above.

Relaxing the assumptions that are inherent in (4) - (6) is another potential avenue for future work. One possibility would be to allow for interactive fixed effects in the slope coefficients, where for instance $\boldsymbol{\theta}_{it} = \boldsymbol{\theta} + \boldsymbol{\theta}_i \cdot \boldsymbol{\theta}_t$. Lastly, it may be possible to relax the structure of the time heterogeneity in the slope coefficients to vary between latent groups in the data, as Bonhomme and Manresa (2015) did for time heterogeneity in the intercept term. For instance, it might be possible to set $\boldsymbol{\theta}_{it} = \boldsymbol{\theta} + \boldsymbol{\theta}_i + \boldsymbol{\theta}_{gt}$ where g is the group indicator.

6. Conclusion

Slope heterogeneity has been a significant issue for several decades in the econometrics literature. Researchers have assumed in the past that it is impossible to model slope heterogeneity across multiple dimensions without imposing strict assumptions,[6] noting that it would require a consistent estimate of at least as many coefficients as there are observations in the sample. The purpose of this article is to demonstrate that it is in fact possible to consistently estimate these individual coefficients, assuming an additive structure is imposed on the heterogeneity and the panel is moderate to large in both N and T. To the author's knowledge this has not been attempted before in the literature. Since time varying parameters and individual varying parameters are often studied in the time series and panel data literature respectively, an estimator that is capable of accounting for both types of heterogeneity has the potential to be of significant interest to economists.

The capability of the estimator was tested in both an asymptotic and finite sample setting. The Monte Carlo simulation study showed that the MO-OLS estimator is able to consistently and efficiently estimate the mean coefficient in a dynamic panel data model where the individual and time heterogeneity is correlated with the regressors, and where the heterogeneity is not randomly generated.

[6]Such as restricting the time heterogeneity to only vary during a discrete structural break, or assuming that the heterogeneity is random and uncorrelated with the regressors.

In comparison, a number of existent estimators provided very severely biased estimates of the mean coefficients. The results suggest that MO-OLS has wide applicability and benefits for empirical applications. The paper also discussed the potential to generalise estimator in several directions, the most important being to allow for cross-section dependence in the data.

REFERENCES

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: 277–297.

Bai, J. 2009. "Panel Data Models with Interactive Fixed Effects." *Econometrica*, 77: 1229–1279.

Balestra, Pietro. 1996. "Introduction to Linear Models for Panel Data." In *The Econometrics of Panel Data.* Vol. 33 of *Advanced Studies in Theoretical and Applied Econometrics*, , ed. László Mátyás and Patrick Sevestre, 25–33. Springer Netherlands.

Baltagi, B., Q. Feng, and C. Kao. 2016. "Estimation of heterogeneous panels with structural breaks." *Journal of Econometrics*, 191: 176–195.

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

Boneva, L., O. Linton, and M. Vogt. 2015. "A semiparametric model for heterogeneous panel data with fixed effects." *Journal of Econometrics*, 188: 327–345.

Bonhomme, S., and E. Manresa. 2015. "Grouped Patterns of Heterogeneity in Panel Data." *Econometrica*, 83(3): 1147–1184.

Chudik, A., and M.H. Pesaran. 2015. "Common correlated effects estimation of heterogeneous dynamic panel data models with weakly exogenous regressors." *Journal of Econometrics*, 188(2): 393–420.

Cooley, T.F., and E.C. Prescott. 1976. "Estimation in the Presence of Stochastic Parameter Variation." *Econometrica*, 44: 167–184.

Degui, L., J. Chen, and S. Gao. 2011. "Non-parametric time-varying coefficient panel data models with fixed effects." *Econometrics Journal*, 14: 387–408.

Harding, M., and C. Lamarche. 2011. "Least squares estimation of a panel data model with multifactor error structure and endogenous covariates." *Economic Letters*, 111(3): 197–199.

Hsiao, C. 1974. "Statistical Inference for a Model with Both Random Cross-Sectional and Time Effects." *International Economic Review*, 15(1): 12–30.

Hsiao, C. 1975. "Some Estimation Methods for a Random Coefficient Model." *Econometrica*, 43(2): 305–325.

Hsiao, C., and M.H. Pesaran. 2004. "Random Coefficient Panel Data Models." *IZA Discussion Paper No. 1236*.

Hurwicz, L. 1950. "Least squares bias in time series." In *Statistical Inference in Dynamic Economic Models*. , ed. T.C. Koopman, 365–383. New York:Wiley.

Lee, M.J. 2015. "Panel conditional and multinomial logit with time-varying parameters." *Studies in Nonlinear Dynamics & Econometrics*, 19: 317–337.

Liu, L., and D. Hanssens. 1981. "A bayesian approach to time-varying cross-sectional regression models." *Journal of Econometrics*, 15: 341–356.

Mesters, G., and S.J. Koopman. 2014. "Generalized dynamic panel data models with random effects for cross-section and time." *Journal of Econometrics*, 180: 127–140.

Pagan, A. 1980. "Some Identification and Estimation Results for Regression Models with Stochastically Varying Coefficients." *Journal of Econometrics*, 13: 341–363.

Pesaran, M. 2006. "Estimation and inference in large heterogeneous panels with a multifactor error structure." *Econometrica*, 74(4): 967–1012.

Pesaran, M. H. 2015. *Time series and panel data econometrics.* Oxford University Press.

Pesaran, M. H., and R. Smith. 1995. "Estimating long-run relationships from dynamic heterogeneous panels." *Journal of Econometrics*, 68: 79–113.

Swamy, P.A.V. 1970. "Efficient inference in a random coefficient regression model." *Econometrica*, 38: 311–323.

Appendix

Postulate 1

Consider two bounded random variables A_{it} and B_{it}. It can be easily verified that the sequence:

$$a_\ell = \boldsymbol{I}[\ell \text{ is odd}]\frac{1}{T}\sum_{t=1}^{T} A_{it}^{-1}A_{it}(a_{\ell-1}) + \boldsymbol{I}[\ell \text{ is even}]\frac{1}{N}\sum_{i=1}^{N} A_{it}^{-1}A_{it}(a_{\ell-1}) \tag{A1}$$

is a Cauchy sequence in ℓ, holding N and T constant where $a_0 = \frac{1}{N}\sum_{i=1}^{N} A_{it}^{-1}B_{it}$. The sequence will converge to the overall sample average:

$$\lim_{\ell\to\infty}(a_\ell) = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} A_{it}^{-1}B_{it} \tag{A2}$$

Intuitively, when ℓ increases the sequence utilises more information from the entire NT sample for each element in the sum and will eventually begin to approximate the simple average of the entire sample arbitrarily well. This postulate will be useful when proving Theorem 1.

Proof of Theorem 1

From (12), (13), and (17) it is seen that the bias correction constitutes:

$$\sum_{\ell=0}^{L}(-1)^{\ell+1}((\boldsymbol{R}_N\Gamma_{1,\ell}-\boldsymbol{R}_{NT}\Gamma_{1,\ell})+(\boldsymbol{R}_T\Gamma_{2,\ell}-\boldsymbol{R}_{NT}\Gamma_{2,\ell}))=$$
$$(\boldsymbol{R}_N\boldsymbol{\theta}_i-\boldsymbol{R}_{NT}\boldsymbol{\theta}_i)+(\boldsymbol{R}_T\boldsymbol{\theta}_t-\boldsymbol{R}_{NT}\boldsymbol{\theta}_t)+$$
$$((\boldsymbol{R}_N\boldsymbol{\Theta}_{1,L}-\boldsymbol{R}_{NT}\boldsymbol{\Theta}_{1,L})+(\boldsymbol{R}_T\boldsymbol{\Theta}_{2,L}-\boldsymbol{R}_{NT}\boldsymbol{\Theta}_{2,L}))+ \tag{A3}$$
$$\sum_{\ell=0}^{L}(-1)^{\ell+1}\left((\boldsymbol{R}_N\boldsymbol{\Lambda}_{1,\ell}-\boldsymbol{R}_{NT}\boldsymbol{\Lambda}_{1,\ell})+(\boldsymbol{R}_T\boldsymbol{\Lambda}_{2,\ell}-\boldsymbol{R}_{NT}\boldsymbol{\Lambda}_{2,\ell})\right)$$

where $\Gamma_{1,\ell}=\boldsymbol{R}_T\Gamma_{2,\ell-1}$ and $\Gamma_{2,\ell}=\boldsymbol{R}_N\Gamma_{1,\ell-1}$ for $\ell>0$, $\Theta_{1,\ell}=\boldsymbol{R}_T\Theta_{2,\ell-1}$ and $\Theta_{2,\ell}=\boldsymbol{R}_N\Theta_{1,\ell-1}$ for $\ell>0$, $\boldsymbol{\Lambda}_{1,\ell}=\boldsymbol{R}_T\boldsymbol{\Lambda}_{2,\ell-1}$ and $\boldsymbol{\Lambda}_{2,\ell}=\boldsymbol{R}_N\boldsymbol{\Lambda}_{1,\ell-1}$ for $\ell>0$, $\Gamma_{1,0}=\hat{\boldsymbol{\theta}}_i$, $\Gamma_{2,0}=\hat{\boldsymbol{\theta}}_t$, $\Theta_{1,0}=\boldsymbol{\theta}_i$, $\Theta_{2,0}=\boldsymbol{\theta}_t$, $\boldsymbol{\Lambda}_{1,0}=\boldsymbol{Q}_{zz,N}^{-1}\boldsymbol{Q}_{zu,N}$, and finally $\boldsymbol{\Lambda}_{2,0}=\boldsymbol{Q}_{zz,T}^{-1}\boldsymbol{Q}_{zu,T}$. The first term contains the original heterogeneity bias, which the bias correction is seeking to approximate. The second term contains a second-order heterogeneity bias that was introduced from the bias correction procedure. The third term contains biases relating to the error term and was also introduced from the bias correction procedure.

First, using Postulate 1 it is true that:

$$\lim_{L\to\infty}\Big((\boldsymbol{R}_N\boldsymbol{\Theta}_{1,L}-\boldsymbol{R}_{NT}\boldsymbol{\Theta}_{1,L})+(\boldsymbol{R}_T\boldsymbol{\Theta}_{2,L}-\boldsymbol{R}_{NT}\boldsymbol{\Theta}_{2,L})\Big)=0 \tag{A4}$$

where $\boldsymbol{z}_{it}\boldsymbol{z}_{it}'$ is used in place of A_{it} (in Postulate 1), while $\boldsymbol{z}_{it}\boldsymbol{z}_{it}'\boldsymbol{\theta}_i$ and $\boldsymbol{z}_{it}\boldsymbol{z}_{it}'\boldsymbol{\theta}_t$ are used in place of B_{it}. Given this, when $L\to\infty$:

$$\hat{\boldsymbol{\theta}}_{it}^{BC}-\boldsymbol{\theta}_{it}=\Big(\boldsymbol{Q}_{zz,N}^{-1}\boldsymbol{Q}_{zu,N}+\boldsymbol{Q}_{zz,T}^{-1}\boldsymbol{Q}_{zu,T}-\boldsymbol{Q}_{zz,NT}^{-1}\boldsymbol{Q}_{zu,NT}\Big)+$$
$$\sum_{\ell=0}^{L}(-1)^{\ell+1}\left((\boldsymbol{R}_N\boldsymbol{\Lambda}_{1,\ell}-\boldsymbol{R}_{NT}\boldsymbol{\Lambda}_{1,\ell})+(\boldsymbol{R}_T\boldsymbol{\Lambda}_{2,\ell}-\boldsymbol{R}_{NT}\boldsymbol{\Lambda}_{2,\ell})\right) \tag{A5}$$

From the Weak Law of Large Numbers and Assumption 2 (exogenous regressors) it is true that as $N \to \infty$:

$$\boldsymbol{Q}_{zu,N} \xrightarrow{p} E(\boldsymbol{Q}_{zu,N}) = 0$$

$$\boldsymbol{Q}^{-1}_{zz,N} \xrightarrow{p} E(\boldsymbol{Q}^{-1}_{zz,N})$$

Furthermore, as $T \to \infty$:

$$\boldsymbol{Q}_{zu,T} \xrightarrow{p} E(\boldsymbol{Q}_{zu,T}) = 0$$

$$\boldsymbol{Q}^{-1}_{zz,T} \xrightarrow{p} E(\boldsymbol{Q}^{-1}_{zz,T})$$

and lastly as $(N, T) \xrightarrow{j} \infty$:

$$\boldsymbol{Q}_{zu,NT} \xrightarrow{p} E(\boldsymbol{Q}_{zu,NT}) = 0$$

$$\boldsymbol{Q}^{-1}_{zz,NT} \xrightarrow{p} E(\boldsymbol{Q}^{-1}_{zz,NT})$$

From the Continuous Mapping Theorem we know that:

$$\boldsymbol{Q}^{-1}_{zz,N}\boldsymbol{Q}_{zu,N} \xrightarrow{p} E(\boldsymbol{Q}^{-1}_{zz,N})0 = 0 \tag{A6}$$

$$\boldsymbol{Q}^{-1}_{zz,T}\boldsymbol{Q}_{zu,T} \xrightarrow{p} E(\boldsymbol{Q}^{-1}_{zz,T})0 = 0 \tag{A7}$$

$$\boldsymbol{Q}^{-1}_{zz,NT}\boldsymbol{Q}_{zu,NT} \xrightarrow{p} E(\boldsymbol{Q}^{-1}_{zz,NT})0 = 0 \tag{A8}$$

Given (A6) - (A8) and the Continuous Mapping Theorem it is also true that:

$$\begin{aligned} &\sum_{\ell=0}^{L}(-1)^{\ell+1}\left((\boldsymbol{R}_N\boldsymbol{\Lambda}_{1,\ell} - \boldsymbol{R}_{NT}\boldsymbol{\Lambda}_{1,\ell}) + (\boldsymbol{R}_T\boldsymbol{\Lambda}_{2,\ell} - \boldsymbol{R}_{NT}\boldsymbol{\Lambda}_{2,\ell})\right) \xrightarrow{p} \\ &\sum_{\ell=0}^{L}(-1)^{\ell+1}\left((\boldsymbol{R}_N 0 - \boldsymbol{R}_{NT}0) + (\boldsymbol{R}_T 0 - \boldsymbol{R}_{NT}0)\right) = 0 \end{aligned} \tag{A9}$$

Therefore, as required for Theorem 1:

$$\hat{\boldsymbol{\theta}}_{it}^{BC} - \boldsymbol{\theta}_{it} \xrightarrow{p} 0 \tag{A10}$$

Proof of Theorem 2

Since $\bar{\boldsymbol{\theta}} = \boldsymbol{\theta} + E(\boldsymbol{\theta}_i) + E(\boldsymbol{\theta}_t) = E(\boldsymbol{\theta}_{it})$, $\hat{\boldsymbol{\theta}}_{MO} = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\hat{\boldsymbol{\theta}}_{it}^{BC}$, and the result from Theorem 1 that $\hat{\boldsymbol{\theta}}_{it}^{BC} - \boldsymbol{\theta}_{it} \xrightarrow{p} 0$ when $L \to \infty$ and then $(N,T) \xrightarrow{j} \infty$, the Weak Law of Large Numbers shows that:

$$\hat{\boldsymbol{\theta}}_{MO} = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\hat{\boldsymbol{\theta}}_{it} \xrightarrow{p} E(\boldsymbol{\theta}_{it}) \tag{A11}$$

which implies Theorem 2.

Proof of Theorem 3

From (A5), (19), and $\bar{\boldsymbol{\theta}} = \boldsymbol{\theta} + E(\boldsymbol{\theta}_i) + E(\boldsymbol{\theta}_t)$ when $L \to \infty$ it is true that:

$$\sqrt{NT}(\hat{\boldsymbol{\theta}}_{MO} - \bar{\boldsymbol{\theta}}) = \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}((\boldsymbol{\theta}_i - E(\boldsymbol{\theta}_i)) + (\boldsymbol{\theta}_t - E(\boldsymbol{\theta}_t)) + \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}(\Psi_{it} + \Xi_{it}) \tag{A12}$$

where $\boldsymbol{\Psi}_{it} = \left(\boldsymbol{Q}_{zz,N}^{-1}\boldsymbol{Q}_{zu,N} + \boldsymbol{Q}_{zz,T}^{-1}\boldsymbol{Q}_{zu,T} - \boldsymbol{Q}_{zz,NT}^{-1}\boldsymbol{Q}_{zu,NT}\right)$ and furthermore $\boldsymbol{\Xi}_{it} = \sum_{l=1}^{L}(-1)^l\left((\boldsymbol{R}_N\Lambda_{1,l-1} - \boldsymbol{R}_{NT}\Lambda_{1,l-1}) + (\boldsymbol{R}_T\Lambda_{2,l-1} - \boldsymbol{R}_{NT}\Lambda_{2,l-1})\right)$.

Consider now the asymptotics where $(N,T) \xrightarrow{j} \infty$, Assumption 2 and the WLLN implies that both $\boldsymbol{\Psi}_{it} \xrightarrow{p} 0$ and $\boldsymbol{\Xi}_{it} \xrightarrow{p} 0$ (as shown in Theorem 1). Accordingly, it is true that:

$$\sqrt{NT}(\hat{\boldsymbol{\theta}}_{MO} - \bar{\boldsymbol{\theta}}) \xrightarrow{d} N\left(0, \Sigma_{MO}\right) \tag{A13}$$

where

$$\Sigma_{MO} = \frac{Var(\boldsymbol{\theta}_i)}{T} + \frac{Var(\boldsymbol{\theta}_t)}{N} + 2Cov(\boldsymbol{\theta}_i, \boldsymbol{\theta}_t) \tag{A14}$$

Now consider the nonparameteric estimate of Σ_{MO} that was proposed in (22). From (A5) it is true that:

$$\hat{\boldsymbol{\theta}}_{it}^{BC} = \boldsymbol{\theta} + \boldsymbol{\theta}_i + \boldsymbol{\theta}_t + \Psi_{it} + \Xi_{it} \tag{A15}$$

and accordingly:

$$\begin{aligned}\left(\hat{\boldsymbol{\theta}}_{\bar{i}} - \hat{\boldsymbol{\theta}}_{MO}\right) =& (\boldsymbol{\theta} - \boldsymbol{\theta}) + \left(\boldsymbol{\theta}_i - \frac{1}{N}\sum_{i=1}^{N}\boldsymbol{\theta}_i\right) + \left(\frac{1}{T}\sum_{t=1}^{T}(\boldsymbol{\theta}_t - \boldsymbol{\theta}_t)\right) + \\ & \left(\frac{1}{T}\sum_{t=1}^{T}\Psi_{it} - \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\Psi_{it}\right) + \left(\frac{1}{T}\sum_{t=1}^{T}\Xi_{it} - \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\Xi_{it}\right) \\ & \xrightarrow{p} (\boldsymbol{\theta}_i - E(\boldsymbol{\theta}_i))\end{aligned}$$

and using a symmetric argument:

$$\left(\hat{\boldsymbol{\theta}}_{\bar{t}} - \hat{\boldsymbol{\theta}}_{MO}\right) \xrightarrow{p} (\boldsymbol{\theta}_t - E(\boldsymbol{\theta}_t))$$

where $\hat{\boldsymbol{\theta}}_{\bar{i}} = \frac{1}{T}\sum_{t=1}^{T}\hat{\boldsymbol{\theta}}_{it}^{BC}$ and $\hat{\boldsymbol{\theta}}_{\bar{t}} = \frac{1}{N}\sum_{i=1}^{N}\hat{\boldsymbol{\theta}}_{it}^{BC}$.

Therefore it is true that:

$$\begin{aligned}\frac{1}{NT-1}\sum_{i=1}^{N}\sum_{t=1}^{T} & \big((\hat{\boldsymbol{\theta}}_{\bar{i}} - \hat{\boldsymbol{\theta}}_{MO})(\hat{\boldsymbol{\theta}}_{\bar{i}} - \hat{\boldsymbol{\theta}}_{MO})' + (\hat{\boldsymbol{\theta}}_{\bar{t}} - \hat{\boldsymbol{\theta}}_{MO})(\hat{\boldsymbol{\theta}}_{\bar{t}} - \hat{\boldsymbol{\theta}}_{MO})' \\ & + 2(\hat{\boldsymbol{\theta}}_{\bar{i}} - \hat{\boldsymbol{\theta}}_{MO})(\hat{\boldsymbol{\theta}}_{\bar{t}} - \hat{\boldsymbol{\theta}}_{MO})'\big) \xrightarrow{p} \frac{Var(\boldsymbol{\theta}_i)}{T} + \frac{Var(\boldsymbol{\theta}_t)}{N} + 2Cov(\boldsymbol{\theta}_i, \boldsymbol{\theta}_t)\end{aligned} \tag{A16}$$

and $\hat{\boldsymbol{\Sigma}}_{MO} \xrightarrow{p} \boldsymbol{\Sigma}_{MO}$ as required.