Business School / School of Economics

UNSW Business School Working Paper

UNSW Business School Research Paper No. 2016 ECON 15
First version: October 2016
UNSW Business School Research Paper No. 2016 ECON 15A
This version: February 2018

Multidimensional Slope Heterogeneity in Panel Data Models

Timothy Neal

Multidimensional Slope Heterogeneity in Panel Data Models

Timothy Neal *
University of New South Wales

February 26, 2018

This article introduces a technique to estimate static or dynamic panel data models that feature two dimensions of heterogeneity in the slope and intercept parameters. It is able to consistently estimate the marginal effect for each individual observation as well as the average over a sample, and allows for correlation between the heterogeneity and the regressors. Models with two-dimensional fixed-effects in the slope parameters have long been considered interesting to economists yet intractable to estimate. Asymptotic theory establishes the consistency and asymptotic normality of the proposed estimator 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, inefficient, and severely biased.

JEL: C13, C22, C23, C33

Keywords: Estimation; Dynamic Modelling; Parameter Heterogeneity; Varying Coefficients.

*The author acknowledges Michael Keane, Stéphane Bonhomme, Maurice Bun, Gerdie Everaert, Luke Hartigan, and Seojeong Lee for constructive comments that improved the article. Work on this project has been supported by Australian Research Council grant FL110100247.

1 Introduction

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, and misspecification. Panel Data estimators that allow for heterogeneity in the intercept term across both individuals and time periods have been extensively applied in economics, yet heterogeneity in the slope coefficients have only been estimated over one dimension in isolation. Modelling two dimensions of fixed effects in the slope coefficients is a highly desirable feature but considered intractable to solve. This article proposes an approach to consistently and efficiently estimate two dimensions of additive fixed effects in the slope coefficients, and thereby represents a significant contribution to the literature that is widely applicable to empirical research.

Consider a dynamic panel data model with minimal restrictions on the parameter heterogeneity:

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

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 is on heterogeneity in the intercept term α_{it}, where applied researchers use the one-way and two-way fixed effects, random effects, and first difference estimators to account for this heterogeneous intercept. Allowing the slope coefficients γ_{it} and β_{it} to vary across i is also a popular feature in panel data models. Pesaran and Smith (1995), which introduces a technique known as mean group OLS, considers estimation when β_{it} is restricted to $\beta_i = \beta + \lambda_i$ and γ_{it} is restricted to $\gamma_i = \gamma + \delta_i$. It shows that a consistent estimate of the average slope effect is possible by averaging individual-specific regression estimates. Explicit estimates of β_i and γ_i can also be of inherent economic interest, for instance when i represents a country or industry.

Meanwhile, time series econometrics contains techniques to model certain types of time heterogeneity in the slope coefficient, 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 $E(\epsilon_t) = 0$. There are some studies that have looked at time varying parameters in panel data as well, such as Degui et al. (2011), Lee (2015), and Liu and Hanssens (1981). For a useful survey of these panel and time series models see Hsiao and Pesaran (2004).

Very few studies consider slope heterogeneity across two dimensions, and the ones that do require strong restrictions on the structure of the heterogeneity. Hsiao (1975) considers a static version of (1) (i.e. $\gamma_{it} = 0 \ \forall \ i$ and t) where the individual effects λ_i and the time effects λ_t are both random processes such that $E(\lambda_i) = E(\lambda_t) = 0$. Importantly, the estimator requires that the heterogeneity is uncorrelated with the regressors, which makes it unsuitable for dynamic models and any static model with this correlation. Baltagi et al. (2016) allows for two dimensions of heterogeneity, but again for a static version of (1) and where the time heterogeneity is restricted to be in the form of a single 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... there [is] no point at which to pool the data, and 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 an estimate of the marginal effect associated with a single observation in the sample. It also proposes a procedure to remove any bias from the estimates that will emerge from correlation between the regressors and this parameter heterogeneity (whether in the intercept or slope coefficients). To the author's knowledge neither component has been considered before in the literature.

For the ease of explication the technique is developed in an environment that assumes cross-sectional independence (implying additive not interactive fixed effects in the intercept) 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.

The technique is potentially of significant interest to economists for several reasons. Since time varying parameters and individual varying parameters are important issues in time series and panel data applications respectively, a model that is robust to both forms of heterogeneity in both the intercept and slopes will have general applicability in empirical research. Since the estimator is robust to correlation between regressors and the intercept and slope heterogeneity, it is particularly useful for dynamic panel data estimation and complex economic relationships. Most importantly, even if an applied researcher is only interested in the average effect over the sample and not the nature of the multidimensional heterogeneity, fixed effects and mean group estimators will be inconsistent and (potentially severely) biased in this environment. Keane and Neal (2018) contains the first application of the estimator to the sensitivity of crop yields to climate change, and demonstrates the usefulness of the estimator for empirical research.

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 provides a consistent estimate of the average coefficient as well as the individual observation-level coefficients. Fixed effects and mean group estimators are found to be inconsistent and highly biased estimators of the average effect in this environment, and depending on the underlying structure of the heterogeneity the bias can be very severe. The results also suggest that the estimator can be applied without cost to any panel dataset, as it is no less efficient than fixed effects or mean group estimators in simpler environments that do not feature two dimensions of fixed effects in the slope parameters.

The rest of the article is organized in the following way. Section 2 presents the estimator in a dynamic environment for both the observation-level coefficients and average coefficient over the sample. Section 3 conducts a Monte Carlo simulation study that tests this approach against a number of alternatives under varying assumptions, while Section 4 provides some concluding remarks. The Appendix provides proofs of the asymptotic results presented in Section 2.

2 The Estimator

2.1 Description of the Environment

Consider the following estimation problem:

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

for individuals $i = 1, 2, \cdots, N$ and time periods $t = 1, 2, \cdots, T$, where $\boldsymbol{x}_{it} = (x_{1it}, x_{2it}, \cdots, x_{Kit})$ is a K x 1 vector of regressors, $\boldsymbol{\beta}_{it} = (\beta_{1it},\ \beta_{2it},\ \cdots,\ \beta_{Kit})$ is a K x 1 vector of coefficients that vary across individuals and over time, γ_{it} is the heterogeneous 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 $\boldsymbol{\beta} = (\beta_1,\ \beta_2,\ \cdots,\ \beta_K)$, individual effects that vary across every unit in the panel c_i, δ_i, and $\boldsymbol{\lambda_i} = (\lambda_{1i},\ \lambda_{2i},\ \cdots,\ \lambda_{Ki})$, and finally time effects that vary between each time period c_t, δ_t, and $\boldsymbol{\lambda_t} = (\lambda_{1t},\ \lambda_{2t},\ \cdots,\ \lambda_{Kt})$. Accordingly, in this environment there are NT observations in the sample and $(2+K)(NT)$ unique coefficients. 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) will yield:

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

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

Examining the pooled estimator reveals multiple sources of potential endogeneity that will lead to bias and inconsistency of the parameter estimates. y_{it-1} and $\boldsymbol{x}_{it}$ may be correlated with c_i and c_t which represent the fixed effects of the intercept term. y_{it-1} will necessarily be correlated with δ_i and also $\boldsymbol{\lambda}_i$ if $\rho_x \neq 0$. Pesaran and Smith (1995) proposes mean group estimation (or 'MG-OLS') to deal with heterogeneity across i 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.[1]

The remaining sources of endogeneity, which to the author's knowledge has not been addressed before in this environment, is 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 propose an estimator that is able to account for fixed effects in the intercept and slope coefficients across both N and T dimensions, and is robust to any correlation between $y_{i,t-1}$ and $\boldsymbol{x}_{it}$ with c_i, c_t, δ_i, δ_t, $\boldsymbol{\lambda}_i$, and $\boldsymbol{\lambda}_t$. To ease explanation in the next section, it abstracts from sources of endogeneity that have already received attention in the literature by making the following four assumptions on (2). Weakening these assumptions, along with others inherent in the formulation of (4) - (6), is left for future work and briefly discussed in the conclusion to the paper.

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 and some constant $R < \infty$, where $\|\boldsymbol{A}\|$ refers to the Frobenius norm 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'.

[1] A particular form of endogeneity between $\boldsymbol{x}_{it}$ and u_{it} is called cross section dependence, 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.

[2] Indeed, the time series literature has often found it useful to model time varying heterogeneity with serial dependence, such as $\beta_t = \rho\beta_{t-1} + \epsilon_t$

Assumption 3: The error term u_{it} is independently distributed across individuals and time (i.e. no cross-section or 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 $-1 < \gamma_{it} < 1$ for all i and t.

2.2 A Consistent Estimate of the Observation Coefficients

The aim of this subsection is to obtain a consistent estimate of $\boldsymbol{\theta}_{it} = (\alpha_{it}, \gamma_{it}, \boldsymbol{\beta}_{it})$, while the next subsection 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 (note that it contains the constant term so that the fixed effects in the intercept will be treated together with the slope coefficients). 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 $\boldsymbol{\theta}$:

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

Expanding on y_{it} and simplifying yields:

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

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 time-specific regressions yield:

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

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 $\boldsymbol{\theta}_{it}$ we combine (10), (12), and (13) as follows:

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

This simplifies to:

$$\hat{\boldsymbol{\theta}}_{it}^{Prel} = \boldsymbol{\theta} + \boldsymbol{\theta}_i + \boldsymbol{\theta}_t + (\boldsymbol{R}_N - \boldsymbol{R}_{i,NT}) + (\boldsymbol{R}_T - \boldsymbol{R}_{t,NT}) + \left(\boldsymbol{Q}_{zu,N} + \boldsymbol{Q}_{zu,T} - \boldsymbol{Q}_{zu,NT}\right) \quad (15)$$

where $\boldsymbol{R}_N = \boldsymbol{Q}_{zz,N}^{-1}\left(\frac{1}{N}\sum_{i=1}^{N} \boldsymbol{z}_{it}\boldsymbol{z}_{it}'\boldsymbol{\theta}_i\right)$, $\boldsymbol{R}_{t,NT} = \boldsymbol{Q}_{zz,NT}^{-1}\left(\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} \boldsymbol{z}_{it}\boldsymbol{z}_{it}'\boldsymbol{\theta}_t\right)$ and similarly for $\boldsymbol{R}_T$ and $\boldsymbol{R}_{i,NT}$, and also $\boldsymbol{Q}_{zu,N} = \boldsymbol{Q}_{zz,N}^{-1}\left(\frac{1}{N}\sum_{i=1}^{N} \boldsymbol{z}_{it}u_{it}\right)$ and similarly for $\boldsymbol{Q}_{zu,T}$ and $\boldsymbol{Q}_{zu,NT}$.

The expression in (15) can be decomposed into three parts. First, there are the true observation-level coefficients $\boldsymbol{\theta} + \boldsymbol{\theta}_i + \boldsymbol{\theta}_t$, then the biases originating from any correlation between the regressors and the heterogeneity (including the fixed effects in the intercept) $(\boldsymbol{R}_N - \boldsymbol{R}_{i,NT}) + (\boldsymbol{R}_T - \boldsymbol{R}_{t,NT})$, and terms involving the errors $\left(\boldsymbol{Q}_{ze,N} + \boldsymbol{Q}_{ze,T} - \boldsymbol{Q}_{ze,NT}\right)$. It is impossible to remove the biases relating to the heterogeneity by pooling the data in different ways (as MG-OLS is able to do when the heterogeneity is only over the N dimension), as here there are two dimensions of heterogeneity in a two-dimension panel.

Nevertheless, the bias terms relating to the heterogeneity can be calculated to arbitrary accuracy and eliminated from the observation-level coefficients by using a certain procedure. To start, it is possible to use $\hat{\boldsymbol{\theta}}_i$ as a sample approximation for $\boldsymbol{\theta}_i$ in $(\boldsymbol{R}_N - \boldsymbol{R}_{i,NT})$ to form $\hat{\boldsymbol{R}}_N$ and $\hat{\boldsymbol{R}}_{i,NT}$, and also using $\hat{\boldsymbol{\theta}}_t$ as a sample approximation for $\boldsymbol{\theta}_t$ in $(\boldsymbol{R}_T - \boldsymbol{R}_{t,NT})$ to form $\hat{\boldsymbol{R}}_T$ and $\hat{\boldsymbol{R}}_{t,NT}$. Inserting (12) and (13) into these parts of (15) yields:

$$\begin{aligned}
&(\hat{\boldsymbol{R}}_N - \hat{\boldsymbol{R}}_{i,NT}) + (\hat{\boldsymbol{R}}_T - \hat{\boldsymbol{R}}_{t,NT}) = (\boldsymbol{R}_N - \boldsymbol{R}_{i,NT}) + (\boldsymbol{R}_T - \boldsymbol{R}_{t,NT}) + \\
&\boldsymbol{Q}_{zz,N}^{-1}\frac{1}{N}\sum_{i=1}^{N}\left(\boldsymbol{z}_{it}\boldsymbol{z}_{it}'\boldsymbol{R}_T + \boldsymbol{z}_{it}\boldsymbol{z}_{it}'\boldsymbol{Q}_{zu,N}\right) + \boldsymbol{Q}_{zz,T}^{-1}\frac{1}{T}\sum_{t=1}^{T}\left(\boldsymbol{z}_{it}\boldsymbol{z}_{it}'\boldsymbol{R}_N + \boldsymbol{z}_{it}\boldsymbol{z}_{it}'\boldsymbol{Q}_{zu,T}\right) \\
&- \boldsymbol{Q}_{zz,NT}^{-1}\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\boldsymbol{z}_{it}\boldsymbol{z}_{it}'\boldsymbol{R}_T + \boldsymbol{z}_{it}\boldsymbol{z}_{it}'\boldsymbol{R}_N + \boldsymbol{z}_{it}\boldsymbol{z}_{it}'\boldsymbol{Q}_{zu,N} + \boldsymbol{z}_{it}\boldsymbol{z}_{it}'\boldsymbol{Q}_{zu,T}\right)
\end{aligned} \quad (16)$$

While this correction contains the original heterogeneity bias $(\boldsymbol{R}_N - \boldsymbol{R}_{i,NT}) + (\boldsymbol{R}_T - \boldsymbol{R}_{t,NT})$, it also introduce further bias terms, some of which relate to the idiosyncratic error term while others the heterogeneity. However, it is possible to show that these new bias terms are smaller in magnitude to the previous bias terms, and can again be approximated using $\hat{\boldsymbol{R}}_N$, $\hat{\boldsymbol{R}}_{i,NT}$, $\hat{\boldsymbol{R}}_T$, and $\hat{\boldsymbol{R}}_{t,NT}$ in the additional biases of (16). This will in turn produce additional biases (that are again smaller in magnitude), which can again be approximated. In fact, this process can be repeated L times to render the heterogeneity biases arbitrarily small and form the final estimates:

$$\hat{\boldsymbol{\theta}}_{it} = \hat{\boldsymbol{\theta}}_i + \hat{\boldsymbol{\theta}}_t - \hat{\boldsymbol{\theta}} + \sum_{\ell=0}^{L}(-1)^{\ell+1}\Bigg(\boldsymbol{Q}_{zz,N}^{-1}\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{z}_{it}\boldsymbol{z}_{it}'\Gamma_{1,\ell} + \boldsymbol{Q}_{zz,T}^{-1}\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{z}_{it}\boldsymbol{z}_{it}'\Gamma_{2,\ell} - \boldsymbol{Q}_{zz,NT}^{-1}\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\boldsymbol{z}_{it}\boldsymbol{z}_{it}'\Gamma_{1,\ell} + \boldsymbol{z}_{it}\boldsymbol{z}_{it}'\Gamma_{2,\ell}\right)\Bigg) \tag{17}$$

where $\Gamma_{1,\ell} = \boldsymbol{Q}_{zz,T}^{-1}(\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{z}_{it}\boldsymbol{z}_{it}'\Gamma_{2,\ell-1})$ and $\Gamma_{2,\ell} = \boldsymbol{Q}_{zz,N}^{-1}(\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{z}_{it}\boldsymbol{z}_{it}'\Gamma_{1,\ell-1})$ when $\ell > 0$, $\Gamma_{1,0} = \hat{\boldsymbol{\theta}}_i$, and finally $\Gamma_{2,0} = \hat{\boldsymbol{\theta}}_t$. The procedure will approximate the true heterogeneity bias when L is sufficiently large. This is a Cauchy sequence in ℓ, so a suitable L can be determined endogenously by programming the sum to stop once the procedure converges to a given level of tolerance. In practice, in all examples we have considered the bias becomes negligible for reasonable values of L.

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 $\hat{\boldsymbol{\theta}}_{it}$

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:

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

2.3 A Consistent Estimate of the Average Coefficient

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

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

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 $\boldsymbol{\Sigma}_{MO} = T^{-1}Var(\boldsymbol{\theta}_i) + N^{-1}Var(\boldsymbol{\theta}_t)$. The asymptotic variance can be consistently estimated nonparametrically by:

$$\hat{\boldsymbol{\Sigma}}_{MO} = \frac{1}{NT-1}\sum_{i=1}^{N}\sum_{t=1}^{T}\left((\hat{\boldsymbol{\theta}}_{it} - \hat{\boldsymbol{\theta}}_{\bar{t}})(\hat{\boldsymbol{\theta}}_{it} - \hat{\boldsymbol{\theta}}_{\bar{t}})' + (\hat{\boldsymbol{\theta}}_{it} - \hat{\boldsymbol{\theta}}_{\bar{i}})(\hat{\boldsymbol{\theta}}_{it} - \hat{\boldsymbol{\theta}}_{\bar{i}})'\right) \tag{22}$$

where $\hat{\boldsymbol{\theta}}_{\bar{i}} = \frac{1}{T}\sum_{t=1}^{T}\hat{\boldsymbol{\theta}}_{it}$ and $\hat{\boldsymbol{\theta}}_{\bar{t}} = \frac{1}{N}\sum_{i=1}^{N}\hat{\boldsymbol{\theta}}_{it}$.Restrictions on the relative rate of convergence of N and T are required due to the presence of a small sample time series bias $O(T^{-1})$ (first documented in Hurwicz 1950) and originates from the inclusion of a lagged dependent variable. Accordingly, the estimator is not appropriate for panels with small T.

3 Monte Carlo Simulations

This section conducts a Monte Carlo simulation study to determine the finite sample performance of the MO-OLS estimate of the average coefficient vector $\bar{\boldsymbol{\theta}}$, as well as the set of observation-level coefficients $\boldsymbol{\theta}_{it}$, that were proposed in the previous section. A number of scenarios are formulated from a data generating process that features a large panel data structure with multidimensional slope heterogeneity and fixed effects. The performance of the average coefficient will be compared with the one-way and two-way fixed effects estimators and mean group OLS ('MG-OLS'), where each estimator will be tested according to its mean bias and empirical standard deviation. To demonstrate the consistency of the observation-level coefficients, the distribution of the bias over the set of coefficients will be analyzed at each sample size.

3.1 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 unbiased estimate of the average β_{it} over all Monte Carlo repetitions will be approximately equal to 1, while an unbiased 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 $c = 1$ and $c_i \sim c_t \sim N(0, 0.353)$. 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 $\alpha_1 = 1$, whereas α_2 will vary by scenario.

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 and $\sigma^2 = 0.353$ for the regressor heterogeneity λ_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:

$$\delta_t = \begin{cases} -0.104 & \text{if } t < T/2 \\ 0.104 & \text{if } t \geq T/2 \end{cases}$$

and:

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

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

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.

3.2 Results for the Average Coefficient

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 and the four estimators under consideration. Since the average true 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 the econometric estimators. N is held at 50 in all scenarios while T varies between 30 and 200, since it is primarily variation in T that affects performance (due to the presence of the $O(T^{-1})$ bias). The mean and standard deviation of iterations in the MO-OLS bias removal procedure is also reported.[3] This scenario presents a static panel data model that does not contain correlation between the multidimensional slope heterogeneity and the regressor (i.e. $\alpha_2 = 0$), but does possess fixed effects across both dimensions that are correlated with the regressor.

The results show that ignoring the time effects in the intercept term 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. The final MO-OLS estimates, utilizing the procedure defined in (17), successfully removes all bias, as does using the two-way FE estimator. Moreover, MO-OLS is slightly more efficient than FE-OLS, particularly in small samples. This result suggests that the procedure introduced in this article to remove heterogeneity bias may be superior to standard econometric estimators even when there is only correlation between the regressors and the heterogeneous intercept.

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
Final	0.998	1.002	0.999	0.998	0.999	0.087	0.074	0.066	0.062	0.058
Iterations	14	12	11	10	9	2	2	1	1	1

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

The second scenario introduces correlation between the slope heterogeneity and the single regressor, and the results can be seen in Table 3. The results show that the one-way

[3] This is the value of L sufficient for the procedure to achieve convergence.

FE and MG-OLS estimates are severely biased and do not improve as T increases. It's clear that ignoring the time heterogeneity in the slope coefficients can have dramatic implications for statistical inference. The MO-OLS estimates removes all of the heterogeneity bias as does using the two-way FE estimator. Unpublished simulation results, which are available upon request, show that increasing the value of α_2 in this scenario introduces bias to the two-way FE estimator and also significantly worsens its efficiency relative to MO-OLS.

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
Final	0.998	1.002	1.001	0.998	1.002	0.087	0.074	0.069	0.061	0.057
Iterations	17	15	14	14	14	2	2	1	1	1

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 in environments where MG-OLS will perform adequately. 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 procedure to remove heterogeneity bias does not introduce any additional inefficiency when it is unnecessarily 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.

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
Final	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
Final	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.

The results for the fourth scenario are presented in Table 5, which features a lagged dependent variable with a heterogeneous slope coefficient. The heterogeneity in γ_{it} and β_{it} is random across both dimensions and there is correlation between the regressor 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 removal procedure to the final estimates successfully eliminates 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 $T > 50$. Also worth noting is that the number of iterations (i.e. the value of L) required to achieve convergence in the bias removal procedure has significantly increased, with $L > 300$ required in most MC iterations. Since each iteration is computationally cheap at these sample sizes, the number of required iterations has little significance in practice.

The results for the fifth scenario are presented in Table 6. In this scenario, the slope heterogeneity across t is generated with dependence (see Section 3.1 for details). The results show very severe bias for all estimators considered here save the final MO-OLS estimates. The autoregressive parameter $\bar{\gamma}$ is significantly overestimated, while $\bar{\beta}$ is more than double its true value in one-way FE and MG-OLS. The MO-OLS estimator successfully removes all bias at moderate to large T, and is able to estimate both parameters relatively efficiently. The value of L required for the bias removal procedure to converge increases as well.

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 Section 3.1 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 unity for both FE estimators, MG-OLS, and also the preliminary MO-OLS estimates. Fortunately, the final MO-OLS estimates are consistent and perform similarly to the preceding scenarios.

To summarize, the Monte Carlo simulation study has successfully demonstrated that the MO-OLS estimator is able to consistently and efficiently estimate the mean coefficient

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
Final	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
Final	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.

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
Final	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
Final	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.

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
Final	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
Final	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.

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 lead to severely overestimated and highly inefficient estimates.

3.3 Results for the Observation Coefficients

Even if MO-OLS is a consistent estimate of the average coefficients, it is possible for the estimates of observation-level coefficients proposed in Section 2.2 to remain inconsistent. This section tests the consistency of (17) in the second and fourth scenario.[4] Since there are numerous observation-level coefficients in a panel dataset, and the number increases with the sample size, the overall distribution of bias across the sample will give an indication on the consistency and bias of each estimate. Bias for an observation-level coefficient is simply defined as:

$$B_{it} = S^{-1} \sum_{s=1}^{S} (\hat{\boldsymbol{\theta}}_{its} - \boldsymbol{\theta}_{its}) \tag{25}$$

where S is set to $1,000$ for this study as above.

While the mean value of B_{it} will be approximately zero wherever we find no bias in the MO-OLS estimate of the average coefficient in Tables 2 - 7, the standard deviation and percentiles of the distribution will determine whether there exists specific observation coefficients that remain biased even after S repetitions. The proportion of coefficients that lie within 1% of bias will also indicate whether the tails of the distribution are noteworthy at a specific sample size. In order to establish consistency, the standard deviation should decline and the proportion of coefficients within 1% of bias should converge to one as the sample size increases.

Table 8 presents the results of this exercise. It separates the results first by parameter (as the fourth scenario includes a lagged dependent variable), and then by sample size where N and T increase together in this case. In regards to the second scenario, the simulations report that the mean bias of the observation coefficients over the sample is approximately zero at all sample sizes, which is consistent with the results of the average

[4]Other scenarios are excluded for the sake of brevity, and the results (which are comparable) are available upon request.

coefficient in Table 3. Nevertheless, specific observation parameters over or undershoot the true observation-level coefficient (which will not be exactly equal to one, it will vary over the sample and across simulations). When $N, T = 30$, we find that 80% of the sample contain estimates that are within 1% bias of the true coefficient, and the 25th to 75th percentile range lies within -0.003 and 0.008 of bias.

Table 8: Simulation Results for $\hat{\boldsymbol{\theta}}_{it}$

(N,T)	**Scenario 2**					**Scenario 4**				
	30	50	100	200	350	30	50	100	200	350
Results for $\hat{\boldsymbol{\beta}}_{it}$										
Mean	0.002	-0.000	0.000	-0.000	0.000	0.013	0.007	0.004	0.002	0.001
Std. Dev.	0.007	0.004	0.003	0.002	0.002	0.008	0.005	0.004	0.003	0.002
75th pctile.	0.008	0.003	0.002	0.002	0.001	0.019	0.011	0.007	0.004	0.003
25th pctile.	-0.003	-0.003	-0.002	-0.002	-0.001	0.008	0.004	0.002	0.000	0.000
Within 1% bias	0.794	0.987	0.999	1.000	1.000	0.346	0.698	1.000	1.000	1.000
Results for $\hat{\boldsymbol{\gamma}}_{it}$										
Mean						-0.025	-0.014	-0.007	-0.004	-0.002
Std. Dev.						0.001	0.001	0.000	0.000	0.000
75th pctile.						-0.025	-0.014	-0.007	-0.003	-0.002
25th pctile.						-0.026	-0.015	-0.008	-0.004	-0.002
Within 1% bias						0.000	0.000	0.936	0.998	1.000

Note: 1,000 Monte Carlo Simulations with varied N and T. 'Within 1% bias' refers to the proportion of parameter estimates across all i and t that have a bias within 1% of the true parameter value.

Importantly, the range of bias compresses as the sample size increases. When $N, T = 100$ approximately 100% of the sample coefficient estimates are within 1% of bias, and the 25th to 75th percentile lie within -0.002 and 0.002 of bias. This indicates that at this sample size statistical inference on the observation-level coefficients is very reliable. It compresses further still when the sample continues to grow.

The results for the fourth scenario are reported on the right panel of Table 8. Consistent with the results in Table 5, the mean bias becomes insignificant when $N, T > 50$ for $\hat{\boldsymbol{\gamma}}_{it}$ and when $N, T \geq 50$ for $\hat{\boldsymbol{\beta}}_{it}$. The range of bias across observation coefficients for $\hat{\boldsymbol{\gamma}}_{it}$ is very narrow, and accordingly the proportion of coefficients within 1% of bias shifts suddenly from none to over 90% when $N, T > 50$ (or 100% when $N, T > 100$). For $\hat{\boldsymbol{\beta}}_{it}$, the range of bias is wider yet declines more quickly as the sample size increases. The distribution of observation coefficients fits entirely within 1% of bias between $N, T = 50$ and $N, T = 100$.

In summary, the results demonstrate that the estimate of the observation-level coefficients outlined in (17) is consistent, and that even when the mean parameter is not biased it is possible for an observation-level coefficient to be biased. It also provides information specific to this data generating process on the reliability of statistical inference as N and T jointly increase. Indeed, it demonstrates that larger sample sizes are needed to ensure the unbiasedness of an individual observation-level coefficient, relative to the overall average coefficient. This is intuitive as the overall sample average coefficient is able to average out random biases in the observation-level coefficients that originate from additional error terms.

4 Conclusion

Slope heterogeneity is a significant issue in both panel data and time series econometrics. In the past, researchers have assumed that modelling slope heterogeneity across multiple dimensions without imposing strict assumptions on that heterogeneity (such as assuming that the heterogeneity is idiosyncratic and uncorrelated with the regressors) is an intractable problem. The purpose of this article is to demonstrate that it is in fact possible to consistently estimate the observation-level coefficients and the average coefficient, even if the heterogeneity is correlated with the regressors, assuming an additive structure is imposed on the heterogeneity and the panel is moderate to large in both N and T. Since either time varying parameters or individual varying parameters are applied extensively in economics, an estimator that is capable of modelling both types of heterogeneity has the potential to be significantly useful to applied researchers.

The capability of the estimator was tested in both an asymptotic and finite sample setting. The Monte Carlo simulation study demonstrates that the MO-OLS estimator is able to consistently and efficiently estimate both the mean coefficient and the observation-level coefficients 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.

In comparison, a number of existent estimators produce severely biased estimates of the mean coefficients.

Future work may be able to extend MO-OLS in a number of directions using techniques adapted from extensions to MG-OLS. Perhaps the most important of these is to allow for cross-section dependence in the data, which may appear through interactive fixed effects (as in Bai 2009) or unobserved common factors (as in Pesaran 2006 and Chudik and Pesaran 2015). A further extension would allow for serial dependence in the error term and also endogeneity in the regressors by using 2SLS or GMM in each of the regressions that form the MO-OLS estimates.

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, 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.

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, P. (1996). Introduction to linear models for panel data. In L. Mátyás and P. Sevestre (Eds.), *The Econometrics of Panel Data*, Volume 33 of *Advanced Studies in Theoretical and Applied Econometrics*, pp. 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.

Bonhomme, S. and E. Manresa (2015). Grouped patterns of heterogeneity in panel data. *Econometrica 83*(3), 1147–1184.

Chudik, A. and M. 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. and E. 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.

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. Pesaran (2004). Random coefficient panel data models. *IZA Discussion Paper No. 1236*.

Hurwicz, L. (1950). Least squares bias in time series. In T. Koopman (Ed.), *Statistical Inference in Dynamic Economic Models*, pp. 365–383. New York: Wiley.

Keane, M. and T. Neal (2018). The impact of climate change on u.s. agriculture: The roles of adaptation techniques and emissions reductions. *Working Paper, UNSW*.

Lee, M. (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.

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 multi-factor 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.

Appendix: Mathematical Proofs

Lemma 1

Consider a MxN square matrix $\boldsymbol{B}$ and a Mx1 column vector $\boldsymbol{\omega}$:

$$\boldsymbol{B} = \begin{pmatrix} b_{1,1} & b_{1,2} & \cdots & b_{1,n} \\ b_{2,1} & b_{2,2} & \cdots & b_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ b_{m,1} & b_{m,2} & \cdots & b_{m,n} \end{pmatrix}, \qquad \boldsymbol{\omega} = \begin{pmatrix} \omega_1 \\ \omega_2 \\ \vdots \\ \omega_m \end{pmatrix}$$

where $b_{m,n} > 0 \ \forall \ m$ and n. Then the Mx1 vector sequence:

$$a_\ell = \begin{cases} \left(\frac{1}{N}\sum_{n=1}^{N} b_{mn}\right)^{-1} \frac{1}{N}\sum_{n=1}^{N} b_{mn} a_{\ell-1} & \text{if } \ell \text{ is odd} \\ \left(\frac{1}{M}\sum_{m=1}^{M} b_{mn}\right)^{-1} \frac{1}{M}\sum_{m=1}^{M} b_{mn} a_{\ell-1} & \text{if } \ell \text{ is even} \end{cases}$$

where $a_0 = \left(\frac{1}{M}\sum_{m=1}^{M} b_{mn}\right)^{-1} \frac{1}{M}\sum_{m=1}^{M} b_{mn}\omega_m$ is a convergent sequence that has the following pointwise limit in ℓ:

$$\lim_{\ell\to\infty}(a_\ell) = \bar{\boldsymbol{\omega}} = \left(\frac{1}{MN}\sum_{m=1}^{M}\sum_{n=1}^{N} b_{m,n}\right)^{-1} \frac{1}{MN}\sum_{m=1}^{M}\sum_{n=1}^{N} b_{m,n}\omega_m \quad \textit{pointwise}$$

Proof of Lemma 1

a_0 represents an average of $\boldsymbol{\omega}$ over m for each n which is weighted by $\boldsymbol{B}$:

$$a_0 = \left[\bar{\omega}_0^{n=1}, \bar{\omega}_0^{n=2}, \cdots, \bar{\omega}_0^{n=N}\right]$$

where $\bar{\omega}_0^n = \left(\frac{1}{M}\sum_{m=1}^M b_{mn}\right)^{-1} \frac{1}{M}\sum_{m=1}^M b_{mn}\omega_m$. Each element of a_1, in turn, represents a weighted average of all the elements in a_0 over n for each m:

$$a_1 = [\bar{\omega}_1^{m=1}, \bar{\omega}_1^{m=2}, \cdots, \bar{\omega}_1^{m=M}]$$

where $\bar{\omega}_1^m = \left(\frac{1}{N}\sum_{n=1}^N b_{m,n}\right)^{-1} \left[b_{m,n=1}\bar{\omega}_0^{n=1} + b_{m,n=2}\bar{\omega}_0^{n=2} + \cdots + b_{m,n=N}\bar{\omega}_0^{n=N}\right]$.

Since $b_{m,n} > 0 \;\forall\; m$ and n, it follows that $inf\{a_0\} \leq inf\{a_1\}$ and $sup\{a_0\} \geq sup\{a_1\}$. If $\exists$ an i, j pair such that $\bar{\omega}_0^{n=i} \neq \bar{\omega}_0^{n=j}$ and $i \neq j$, then it follows that $inf\{a_0\} < inf\{a_1\}$ and $sup\{a_0\} > sup\{a_1\}$. Only if $\bar{\omega}_0^{n=i} = \bar{\omega}_0^{n=j} \;\forall\, i, j$ will $inf\{a_0\} = inf\{a_1\}$ and $sup\{a_0\} = sup\{a_1\}$. The same argument applies for a_2, which is a weighted average of all elements of a_1 over n for each m, and indeed all subsequent values of ℓ in a_ℓ.

Accordingly, for every positive real number $\epsilon > 0$ there is a positive integer N such that for all positive integers $i, j > N$, the distance $d(a_i, a_j) < \epsilon$ (i.e. the sequence is convergent). The sequences condenses until all elements are equal, at which point it rests at the converged value.

To demonstrate that $\lim_{\ell\to\infty}(a_\ell) = \bar{\boldsymbol{\omega}}$ pointwise, first note that:

$$sup\{a_0\} \geq \bar{\boldsymbol{\omega}} \geq inf\{a_0\}$$

as it is impossible for:

$$\left(\frac{1}{M}\sum_{m=1}^M b_{mn}\right)^{-1} \frac{1}{M}\sum_{m=1}^M b_{mn}\omega_m > \left(\frac{1}{MN}\sum_{m=1}^M\sum_{n=1}^N b_{m,n}\right)^{-1} \frac{1}{MN}\sum_{m=1}^M\sum_{n=1}^N b_{m,n}\omega_m \quad \forall n$$

or:

$$\left(\frac{1}{M}\sum_{m=1}^M b_{mn}\right)^{-1} \frac{1}{M}\sum_{m=1}^M b_{mn}\omega_m < \left(\frac{1}{MN}\sum_{m=1}^M\sum_{n=1}^N b_{m,n}\right)^{-1} \frac{1}{MN}\sum_{m=1}^M\sum_{n=1}^N b_{m,n}\omega_m \quad \forall n$$

when $b_{m,n} > 0 \;\forall\; m$ and n.

Then:

$$sup\{a_1\} \geq \bar{\boldsymbol{\omega}} \geq inf\{a_1\}$$

since it is impossible for:

$$\left(\frac{1}{N}\sum_{n=1}^N b_{m,n}\right)^{-1} \frac{1}{N}\sum_{n=1}^N b_{m,n}\bar{\omega}_0^n > \left(\frac{1}{MN}\sum_{m=1}^M\sum_{n=1}^N b_{m,n}\right)^{-1} \frac{1}{MN}\sum_{m=1}^M\sum_{n=1}^N b_{m,n}\bar{\omega}_0^n \quad \forall m$$

or:

$$\left(\frac{1}{N}\sum_{n=1}^{N} b_{m,n}\right)^{-1}\frac{1}{N}\sum_{n=1}^{N} b_{m,n}\bar{\omega}_0^n < \left(\frac{1}{MN}\sum_{m=1}^{M}\sum_{n=1}^{N} b_{m,n}\right)^{-1}\frac{1}{MN}\sum_{m=1}^{M}\sum_{n=1}^{N} b_{m,n}\bar{\omega}_0^n \quad \forall m$$

when $b_{m,n} > 0 \ \forall \ m$ and n, and:

$$\left(\frac{1}{MN}\sum_{m=1}^{M}\sum_{n=1}^{N} b_{m,n}\right)^{-1}\frac{1}{MN}\sum_{m=1}^{M}\sum_{n=1}^{N} b_{m,n}\bar{\omega}_0^n =$$
$$\left(\frac{1}{MN}\sum_{m=1}^{M}\sum_{n=1}^{N} b_{m,n}\right)^{-1}\frac{1}{MN}\sum_{m=1}^{M}\sum_{n=1}^{N} b_{m,n}\left[\left(\frac{1}{MN}\sum_{m=1}^{M}\sum_{n=1}^{N} b_{m,n}\right)^{-1}\frac{1}{MN}\sum_{m=1}^{M}\sum_{n=1}^{N} b_{m,n}\omega_m\right] =$$
$$\left(\frac{1}{MN}\sum_{m=1}^{M}\sum_{n=1}^{N} b_{m,n}\right)^{-1}\frac{1}{MN}\sum_{m=1}^{M}\sum_{n=1}^{N} b_{m,n}\omega_m = \bar{\boldsymbol{\omega}}$$

The same argument can be applied to $a_\ell \ \forall \ell > 0$, so that $sup\{a_\ell\} \geq \bar{\boldsymbol{\omega}} \geq inf\{a_\ell\} \ \forall \ell$. Therefore, a_ℓ is a convergent sequence of vectors that always contains within it $\bar{\boldsymbol{\omega}}$, which is accordingly the pointwise limit of the sequence.

Proof of Theorem 1

From (10), (12), (13), and (17) it is true that:

$$\begin{aligned}\hat{\boldsymbol{\theta}}_{it} - \boldsymbol{\theta}_{it} = \left(\boldsymbol{Q}_{zu,N} + \boldsymbol{Q}_{zu,T} - \boldsymbol{Q}_{zu,NT}\right) + (-1)^L\Bigg(&\boldsymbol{Q}_{zz,N}^{-1}\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{z}_{it}\boldsymbol{z}_{it}'\Theta_{1,L} \\ &+ \boldsymbol{Q}_{zz,T}^{-1}\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{z}_{it}\boldsymbol{z}_{it}'\Theta_{2,L} - \boldsymbol{Q}_{zz,NT}^{-1}\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\boldsymbol{z}_{it}\boldsymbol{z}_{it}'\Theta_{1,L} + \boldsymbol{z}_{it}\boldsymbol{z}_{it}'\Theta_{2,L}\right)\Bigg) \\ &+ \sum_{\ell=0}^{L}(-1)^{\ell+1}\Bigg(\boldsymbol{Q}_{zz,N}^{-1}\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{z}_{it}\boldsymbol{z}_{it}'\Lambda_{1,\ell} + \boldsymbol{Q}_{zz,T}^{-1}\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{z}_{it}\boldsymbol{z}_{it}'\Lambda_{2,\ell} - \\ &\boldsymbol{Q}_{zz,NT}^{-1}\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\boldsymbol{z}_{it}\boldsymbol{z}_{it}'\Lambda_{1,\ell} + \boldsymbol{z}_{it}\boldsymbol{z}_{it}'\Lambda_{2,\ell}\right)\Bigg)\end{aligned} \tag{A.1}$$

where $\Theta_{1,\ell} = \boldsymbol{Q}_{zz,T}^{-1}(\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{z}_{it}\boldsymbol{z}_{it}'\Theta_{2,\ell-1})$ and $\Theta_{2,\ell} = \boldsymbol{Q}_{zz,N}^{-1}(\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{z}_{it}\boldsymbol{z}_{it}'\Theta_{1,\ell-1})$ for $\ell > 0$, $\boldsymbol{\Lambda}_{1,\ell} = \boldsymbol{Q}_{zz,T}^{-1}(\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{z}_{it}\boldsymbol{z}_{it}'\boldsymbol{\Lambda}_{2,\ell-1})$ and $\boldsymbol{\Lambda}_{2,\ell} = \boldsymbol{Q}_{zz,N}^{-1}(\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{z}_{it}\boldsymbol{z}_{it}'\boldsymbol{\Lambda}_{1,\ell-1})$ for $\ell > 0$, $\Theta_{1,0} = \boldsymbol{\theta}_i$, $\Theta_{2,0} = \boldsymbol{\theta}_t$, $\boldsymbol{\Lambda}_{1,0} = \boldsymbol{Q}_{zu,N}$, and finally $\boldsymbol{\Lambda}_{2,0} = \boldsymbol{Q}_{zu,T}$. First, using Lemma 1, (16), and (17) it is true that:

$$\lim_{L\to\infty}\Bigg(\boldsymbol{Q}_{zz,N}^{-1}\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{z}_{it}\boldsymbol{z}_{it}'\Theta_{1,L}+\boldsymbol{Q}_{zz,T}^{-1}\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{z}_{it}\boldsymbol{z}_{it}'\Theta_{2,L} - \boldsymbol{Q}_{zz,NT}^{-1}\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}(\boldsymbol{z}_{it}\boldsymbol{z}_{it}'\Theta_{1,L}+\boldsymbol{z}_{it}\boldsymbol{z}_{it}'\Theta_{2,L})\Bigg)=0 \tag{A.2}$$

To see this, exchange $b_{m,n}$ and ω_m in Lemma 1 for $\boldsymbol{z}_{it}\boldsymbol{z}_{it}'$ and either $\boldsymbol{\theta}_i$ or $\boldsymbol{\theta}_t$. Since Lemma 1 showed the sequence a_ℓ converges pointwise to $\bar{\boldsymbol{\omega}}$ in ℓ, then also the vector sequence:

$$q_\ell=\boldsymbol{Q}_{zz,N}^{-1}\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{z}_{it}\boldsymbol{z}_{it}'\Theta_{1,\ell}+\boldsymbol{Q}_{zz,T}^{-1}\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{z}_{it}\boldsymbol{z}_{it}'\Theta_{2,\ell}$$

must converge to $\boldsymbol{Q}_{zz,NT}^{-1}\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}(\boldsymbol{z}_{it}\boldsymbol{z}_{it}'\Theta_{1,L}+\boldsymbol{z}_{it}\boldsymbol{z}_{it}'\Theta_{2,L})$ in ℓ which gives us the result in (A.2).

Furthermore, from the Weak Law of Large Numbers, the Continuous Mapping Theorem, and Assumption 2 (exogenous regressors) it is true that as $N\to\infty$:

$$\boldsymbol{Q}_{zz,N}^{-1}\left(\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{z}_{it}u_{it}\right)\xrightarrow{p}E(\boldsymbol{Q}_{zz,N}^{-1})E(\boldsymbol{z}_{it}u_{it})=E(\boldsymbol{Q}_{zz,N}^{-1})0=0 \tag{A.3}$$

Furthermore, as $T\to\infty$:

$$\boldsymbol{Q}_{zz,T}^{-1}\left(\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{z}_{it}u_{it}\right)\xrightarrow{p}E(\boldsymbol{Q}_{zz,T}^{-1})E(\boldsymbol{z}_{it}u_{it})=E(\boldsymbol{Q}_{zz,T}^{-1})0=0 \tag{A.4}$$

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

$$\boldsymbol{Q}_{zz,NT}^{-1}\left(\frac{1}{NT}\sum_{t=1}^{T}\sum_{i=1}^{N}\boldsymbol{z}_{it}u_{it}\right)\xrightarrow{p}E(\boldsymbol{Q}_{zz,NT}^{-1})E(\boldsymbol{z}_{it}u_{it})=E(\boldsymbol{Q}_{zz,NT}^{-1})0=0 \tag{A.5}$$

Given (A.3) - (A.5) and the Continuous Mapping Theorem it is also true that:

$$\sum_{\ell=0}^{L}(-1)^{\ell+1}\Bigg(\boldsymbol{Q}_{zz,N}^{-1}\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{z}_{it}\boldsymbol{z}_{it}'\Lambda_{1,\ell}+\boldsymbol{Q}_{zz,T}^{-1}\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{z}_{it}\boldsymbol{z}_{it}'\Lambda_{2,\ell}- \boldsymbol{Q}_{zz,NT}^{-1}\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}(\boldsymbol{z}_{it}\boldsymbol{z}_{it}'\Lambda_{1,\ell}+\boldsymbol{z}_{it}\boldsymbol{z}_{it}'\Lambda_{2,\ell})\Bigg)\xrightarrow{p}0 \tag{A.6}$$

Therefore, as required for Theorem 1:

$$\hat{\boldsymbol{\theta}}_{it}-\boldsymbol{\theta}_{it}\xrightarrow{p}0 \tag{A.7}$$

Proof of Theorem 2

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

$$\hat{\boldsymbol{\theta}}_{MO} = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\hat{\boldsymbol{\theta}}_{it} \xrightarrow{p} E(\boldsymbol{\theta}_{it}) \tag{A.8}$$

which implies Theorem 2.

Proof of Theorem 3

From (A.1), (A.2), (19), and $\bar{\boldsymbol{\theta}} = \boldsymbol{\theta} + E(\boldsymbol{\theta}_i) + E(\boldsymbol{\theta}_t)$ when $L \to \infty$ it is true that:

$$\sqrt{NT}(\hat{\boldsymbol{\theta}}_{MO} - \bar{\boldsymbol{\theta}}) = \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}((\boldsymbol{\theta}_i - E(\boldsymbol{\theta}_i)) + (\boldsymbol{\theta}_t - E(\boldsymbol{\theta}_t)) + \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}(\Psi_{it} + \Xi_{it}) \tag{A.9}$$

where $\boldsymbol{\Psi}_{it} = \left(\boldsymbol{Q}_{zu,N} + \boldsymbol{Q}_{zu,T} - \boldsymbol{Q}_{zu,NT}\right)$ and furthermore $\boldsymbol{\Xi}_{it} = \sum_{\ell=0}^{L}(-1)^{\ell+1}\Bigg(\boldsymbol{Q}_{zz,N}^{-1}\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{z}_{it}\boldsymbol{z}_{it}'\Lambda_{1,\ell} + \boldsymbol{Q}_{zz,T}^{-1}\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{z}_{it}\boldsymbol{z}_{it}'\Lambda_{2,\ell} - \boldsymbol{Q}_{zz,NT}^{-1}\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}(\boldsymbol{z}_{it}\boldsymbol{z}_{it}'\Lambda_{1,\ell} + \boldsymbol{z}_{it}\boldsymbol{z}_{it}'\Lambda_{2,\ell})\Bigg)$.

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:

$$\sqrt{NT}(\hat{\boldsymbol{\theta}}_{MO} - \bar{\boldsymbol{\theta}}) \xrightarrow{d} N(0, \Sigma_{MO}) \tag{A.10}$$

where

$$\Sigma_{MO} = \frac{Var(\boldsymbol{\theta}_i)}{T} + \frac{Var(\boldsymbol{\theta}_t)}{N} \tag{A.11}$$

since $\boldsymbol{\theta}_i$ and $\boldsymbol{\theta}_t$ are independent by construction.

Now consider the nonparametric estimate of Σ_{MO} that was proposed in (22):

$$\hat{\boldsymbol{\Sigma}}_{MO} = \frac{1}{NT-1}\sum_{i=1}^{N}\sum_{t=1}^{T}\left((\hat{\boldsymbol{\theta}}_{it} - \hat{\boldsymbol{\theta}}_{\bar{t}})(\hat{\boldsymbol{\theta}}_{it} - \hat{\boldsymbol{\theta}}_{\bar{t}})' + (\hat{\boldsymbol{\theta}}_{it} - \hat{\boldsymbol{\theta}}_{\bar{i}})(\hat{\boldsymbol{\theta}}_{it} - \hat{\boldsymbol{\theta}}_{\bar{i}})'\right)$$

From (A.1) and (A.2) it is true that:

$$\hat{\boldsymbol{\theta}}_{it} = \boldsymbol{\theta} + \boldsymbol{\theta}_i + \boldsymbol{\theta}_t + \Psi_{it} + \Xi_{it} \tag{A.12}$$

and accordingly:

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

and using a symmetric argument:

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

where $\hat{\boldsymbol{\theta}}_{\bar{i}} = \frac{1}{T}\sum_{t=1}^{T} \hat{\boldsymbol{\theta}}_{it}$ and $\hat{\boldsymbol{\theta}}_{\bar{t}} = \frac{1}{N}\sum_{i=1}^{N} \hat{\boldsymbol{\theta}}_{it}$. Therefore it is true that:

$$\frac{1}{NT-1}\sum_{i=1}^{N}\sum_{t=1}^{T}\left((\hat{\boldsymbol{\theta}}_{it} - \hat{\boldsymbol{\theta}}_{\bar{t}})(\hat{\boldsymbol{\theta}}_{it} - \hat{\boldsymbol{\theta}}_{\bar{t}})' + (\hat{\boldsymbol{\theta}}_{it} - \hat{\boldsymbol{\theta}}_{\bar{i}})(\hat{\boldsymbol{\theta}}_{it} - \hat{\boldsymbol{\theta}}_{\bar{i}})'\right) \xrightarrow{p} \frac{Var(\boldsymbol{\theta}_i)}{T} + \frac{Var(\boldsymbol{\theta}_t)}{N} \quad \text{(A.13)}$$

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