OXFORD BULLETIN OF ECONOMICS AND STATISTICS, 85, 3 (2023) 0305-9049
doi: 10.1111/obes.12536

# Estimation of Panel Data Models with Mixed Sampling Frequencies*

YIMIN YANG,† FEI JIA‡ and HAORAN LI§

†*School of Finance, Renmin University of China, Beijing, 100872, China (e-mail: yimin.yang@ruc.edu.cn)*
‡*Department of Economics, Richard A. Chaifetz School of Business, Saint Louis University, St. Louis, 63108, Missouri, USA (e-mail: fei.jia@slu.edu)*
§*School of Applied Economics, Renmin University of China, Beijing, 100872, China (e-mail: haoranl@ruc.edu.cn)*

## Abstract

Standard panel models usually assume that data are available at the same frequency. Occasionally, researchers might work with variables sampled at different frequencies. A common practice is to aggregate all variables to the same frequency by an equal weighting scheme. We show that such a simple aggregation scheme results in biases for common estimators. We propose a data-driven method to determine weights for aggregation. We further demonstrate that, in contrast with single-frequency panel models, the Mundlak device and the Chamberlain's approach lead to different estimators for panels with mixed sampling frequencies. The proposed estimators have satisfying finite sample performances in various simulation designs. As an empirical illustration, we apply the new method to the estimation of the effects of temperature fluctuations on economic growth. The empirical evidence shows that the temperature shocks mainly work through the level effect instead of the growth effect for poor countries.

## I. Introduction

A standard panel data model typically involves data that are sampled at the same frequency. Many of the well-established methods are developed for balanced panel models. However, as emphasized by Wansbeek and Kapteyn (1989), the incompleteness should be the rule rather than exception for empirical researchers. Some recent progress has been made in panel data models with missing observations. For example, Abrevaya (2019) considers estimation of panel data models with missing dependent variables. Wooldridge (2019) shows how to accommodate unbalanced panels within the correlated random effects framework. Muris (2020) proposes an efficient GMM estimator with incomplete data.

JEL Classification numbers: C23, C51.
*The authors thank the editor Debopam Bhattacharya and two anonymous referees for their constructive comments. Yimin Yang gratefully acknowledges financial support from the National Natural Science Foundation of China (Grant No.72103191). This project is also supported by the Fundamental Research Funds for the Central Universities, and the Research Funds of Renmin University of China (Grant No. 202230182).

This paper considers panel data models with mixed sampling frequencies. Mixed data sampling (henceforth MIDAS) could arise in various contexts. For example, many macroeconomic data is available at quarterly or yearly frequencies while financial samples are collected at substantially higher frequencies (daily, hourly). At first glance, panel MIDAS models are similar to unbalanced panel models because the dependent variable and the covariates are not perfectly matched. However, the panel MIDAS models are fundamentally different from unbalanced panel models. The unbalanced panel arises when some of the dependent variables and/or the covariates are missing. On the other hand, there is no missing data problem in the MIDAS settings. The dependent variable and the covariates are not synchronized simply because they are available at different frequencies. For example, it is difficult or unreasonable to assign a country's GDP of a given quarter to a given day to match the daily financial data. The methods developed for unbalanced panels are not directly applicable to our MIDAS settings.

The MIDAS is first introduced for time series regression models. Some recent papers on time series models with MIDAS include Ghysels, Santa-Clara, and Valkanov (2006); Ghysels, Sinko, and Valkanov (2007), Ghysels and Wright (2009) and Andreou, Ghysels, and Kourtellos (2010) among others. See also Foroni and Marcellino (2013) for a comprehensive survey of recent progress. Another strand of the literature is related to temporal aggregation. Some seminal works include Sims (1971), Phillips (1972, 1974), Hsiao (1979) and Granger (1987) among many others. The comparisons of these two strands of the literatures can be found in Andreou *et al.* (2010).

The fixed effects estimator is one of the most popular estimators for empirical researchers. With the presence of MIDAS, a simple strategy is to adopt a flat weighting scheme, i.e. taking the simple arithmetic average of the high-frequency data up to the low-frequency level. It is of practical interests to derive the conditions under which the fixed effects estimator with a flat weighting scheme remains consistent. This partly motivates our paper. We make use of the decomposition technique developed by Andreou *et al.* (2010) to show that the equal weights aggregation scheme results in biases of the fixed effects estimator except for some special cases. Thereupon, we propose a data-driven method to assign aggregation weights to get consistent estimators.

A recent paper by Khalaf *et al.* (2021) studies the dynamic panel data model with MIDAS. They extend the GMM methods of Anderson and Hsiao (1982) and Arellano and Bond (1991) to dynamic panel models with MIDAS. The static model studied in our paper can been seen as a special case of the dynamic model in Khalaf *et al.* (2021) where the coefficient of the lagged dependent variable is set to zero. Their GMM estimator, utilizing more moment conditions, is potentially more efficient than our estimator. However, as will be seen later, our estimator is developed within a more general framework than theirs. Another difference is that our paper attempts to extend the Mundlak and Chamberlain approach to the panel MIDAS setting, which is not considered in Khalaf *et al.* (2021).

The Mundlak (1978) device has been a popular device to model the correlations between the fixed effects and the explanatory variables. It assumes that the time averages of the explanatory variables are good proxies of the individual fixed effects. Chamberlain (1982) proposes another approach that is more flexible than the Mundlak device by projecting the individual fixed effect onto the entire history of the explanatory variables. It is well known (see, e.g. Wooldridge (2010, Chapter 10)) that the fixed effects estimator can be obtained by a pooled ordinary least squares (OLS) estimation of the original linear panel model augmented by the time averages or the entire history of the explanatory variables. We show that, when the true weights are unknown in the panel MIDAS model, this equivalence result no longer holds for the use of Mundlak device but continues to be true for the Chamberlain's projection approach using high frequency series. To the best of our knowledge, this finding is new for the panel MIDAS model.

As an empirical illustration, we revisit the paper of Dell, Jones, and Olken (2012), which uses annual data to estimate the effects of temperature fluctuations within countries on economic outcomes from 1950–2003. Conforming to the old idea that climate may substantially influence the economic performance, they find that a 1°C increase in annual temperature reduces economic growth by 1.3 percentage points for poor countries. We note that they use an equal weighting scheme to transform the high frequency weather variables onto annual data. Yet, a 1°C increase in warm days and a 1°C increase in cold days, by their nature, can have very different impacts on economic activities. Thus, such a simple averaging strategy could lead to biased estimators. To tackle this issue, we use a data-driven approach to rebuild the weather variation measurement from high-frequency data instead of using simple annual averages. We re-examine various specifications of Dell *et al.* (2012) using our new method. Our estimates are broadly consistent with Dell *et al.* (2012) in the sense that the high temperature appears to reduce the economic growth rates for poor countries, although the mechanism is different from their finding.

The rest of the paper is organized as follows. Section II sets up the model and shows that the usual fixed effects estimator with an equal weighting scheme is generally inconsistent. We propose a modified fixed effects estimator and derive its asymptotic properties. Section III extends the Mundlak's device and Chamberlain's projection approach to the panel MIDAS regression model. Section IV presents a small simulation study to investigate the finite sample performance of the proposed estimators. An empirical application is provided in section V. Section VI concludes the paper.

## II. The fixed effects estimator of panel MIDAS models

A standard panel data model with an additive individual fixed effects is usually set up as

$$y_{it} = x_{it}\beta^* + c_i + u_{it}, \quad t = 1, ..., T; \quad i = 1, ..., N \tag{1}$$

where the dependent variable $y_{it}$ and the $1 \times p$ vector of explanatory variables $x_{it}$ are sampled at the same frequency. We are interested in estimating $\beta^*$ while allowing for arbitrary correlations between the individual fixed effect $c_i$ and the explanatory variables $x_{it}$.

This paper focuses on the case where the explanatory variables are sampled at a higher frequency than the dependent variables. Let $x_i^{(t)} = \left(x_{i1/m}^{(t)}, x_{i2/m}^{(t)}, ..., x_{im/m}^{(t)}\right)$, a $1 \times mp$ vector, denote the $m$ observations of the explanatory variables sampled between $t-1$ and $t$. [1] Define $y_i = (y_{i1}, ..., y_{iT})'$ and $x_i = \left(x_i^{(1)}, ..., x_i^{(T)}\right)$ as the collection of all covariates of

[1]Here we implicitly assume that all covariates are sampled at the same frequency. This assumption can be relaxed at the cost of complicating the notations without bringing us any new insights.

individual $i$. The asymptotics of this paper will be derived assuming $N \to \infty$ with $T$ and $m$ fixed. We leave extensions of large $T$ or large $m$ to future work. Our first assumption relates to the sampling process.

*Assumption 1. (random sampling)* $(y_i, x_i, c_i)$ is independently and identically distributed $(i.i.d)$ across $i = 1, \ldots, N$.

Note that Assumption 1 allows arbitrary temporal correlations of the covariates and the idiosyncratic errors. For static panel models, it is also common to impose the strict exogeneity assumption of the explanatory variables with respect to the idiosyncratic errors. Our exogeneity assumption takes a different form as we are conditioning on the high frequency explanatory variables instead of the aggregated variables.

*Assumption 2. (strict exogeneity)* $E(u_{it}|x_i, c_i) = 0$ for $t = 1, \ldots, T$.

One might wonder if the above strict exogeneity assumption can be weakened by replacing the high frequency covariates $x_i$ by the aggregated covariates. As to be shown later, we need the high frequency strict exogeneity to ensure the identification of the underlying parameters. While we impose the conditional mean independence between the explanatory variables and the idiosyncratic errors, we do not restrict the dependences between the covariates and the fixed effects, which is important in many empirical analysis. The true generating process of the aggregated variable $x_{it}$ is characterized as follows.

*Assumption 3. (true weighting schemes)* For $k = 1, \ldots, p$, $x_{k,it}(\alpha_k^*) = \sum_{j=1}^m a_{jk}^* x_{k,ij/m}^{(t)} + \varepsilon_{k,it}^x$, where $\alpha_k^* = (a_{1k}^*, \ldots, a_{mk}^*)'$ with $\sum_{j=1}^m a_{jk}^* = 1$, and $\varepsilon_{k,it}^x$ is the stochastic error such that $E(\varepsilon_{k,it}^x|x_i) = 0$.

Assumption 3 specifies the regressors $x_{it}(\alpha^*)$ as the sum of a weighted average of the $m$ high frequency series and the stochastic error. The main results in this paper remain unaltered if the number of high frequency series is larger than $m$. The normalization assumption that $\sum_{j=1}^m a_{jk}^* = 1$ is imposed to give identification of $\beta$. Instead of using a flat weighting scheme by setting $a_{jk} = \frac{1}{m}$ for all $j$ and $k$, we allow different regressors to have different aggregation schemes. Given our fixed-$T$ framework, a possible generalization of the data generating process in assumption 3 is to make the weights specific to each period so that the true aggregating schemes could vary across time. Here we choose not to make such a generalization to avoid further notational complications.

An implicit constraint imposed by assumption 3 is that the aggregation schemes are homogeneous across all individuals. This could be rather restrictive within the fixed effects estimation framework as it imposes strong assumptions on the joint distribution of the regressors and the unobserved heterogeneity.[2] We note that the data generating process of the aggregated variables can depend on the individual fixed effects by incorporating $c_i$ in $\varepsilon_{k,it}^x$ in an additive form, which is ultimately subsumed into the fixed effects in the main regression model. In fact, our current framework can also accommodate the case that all aggregation weights are unit-specific under the random effects assumption. Let $\left\{a_{i,jk}^*\right\}_{j,k}$ be the weights associated with individual $i$, which can be further decomposed as $a_{i,jk}^* =$

[2] We thank an anonymous referee for pointing out this.

$a^*_{jk} + \varepsilon^a_{i,jk}$ with $a^*_{jk} = E(a^*_{i,jk})$. We assume that $\varepsilon^a_{i,jk}$ is independent of all other variables. Then $x_{k,it}(\alpha^*_k) = \sum_{j=1}^m a^*_{i,jk} x^{(t)}_{k,ij/m} + \varepsilon^x_{k,it} = \sum_{j=1}^m a^*_{jk} x^{(t)}_{k,ij/m} + \sum_{j=1}^m \varepsilon^a_{i,jk} x^{(t)}_{k,ij/m} + \varepsilon^x_{k,it}$. It can be easily verified that the new composed error $\sum_{j=1}^m \varepsilon^a_{i,jk} x^{(t)}_{k,ij/m} + \varepsilon^x_{k,it}$ is still mean independent of all the high frequency covariates. Thus our assumption (3) can also accommodate the case that the unit-specific weights are independent of all covariates. Allowing more general unit-specific weights could be an interesting topic for future researches.

From the above discussion we can see that our assumption 3 is more general than the usual aggregation scheme in the literature. Our aggregation process allows for stochastic errors as well as independent random weighting coefficients, while many other researches are built on the deterministic weighting scheme (e.g. Ghysels *et al.* (2006) and Khalaf *et al.* (2021)). Thus our estimator is developed within a more general framework than the GMM estimator proposed by Khalaf *et al.* (2021).[3]

Let $\alpha^* = (\alpha^{*'}_1, \dots, \alpha^{*'}_p)'$. We can then rewrite equation (1) as

$$y_{it} = x_{it}(\alpha^*)\beta^* + c_i + u_{it}, \quad t = 1, \dots, T; \quad i = 1, \dots, N \tag{2}$$

Starting from equation (2), we let $x_{it}(\alpha^*)$ denote the linear combination of high frequency regressors without the stochastic errors, which are subsumed into $u_{it}$ and/or $c_i$. With some algebra it can be shown that $x_{it}(\alpha^*) = x^{(t)}_i A(\alpha^*)$ where $A(\alpha^*)$ is a $mp \times p$ matrix obtained by vertically stacking $p \times p$ diagonal matrices such as $\text{diag}\{a^*_{j1}, a^*_{j2}, \dots, a^*_{jp}\}$ for $j = 1, \dots, m$. Let $x^E_{it}$ be the explanatory variables generated by the equal weighting scheme, i.e. $a_{jk} = \frac{1}{m}$ for all $j$ and $k$. If we run a fixed effects estimation for the following model

$$y_{it} = x^E_{it}\beta^* + c_i + u_{it} + x^B_{it}\beta^*, \tag{3}$$

where $x^B_{it} = x_{it}(\alpha^*) - x^E_{it}$ is the omitted term resulted from using a misspecified weighting scheme, the naive FE estimator with equal weights can be written as

$$\hat{\beta}^E_{\text{FE}} = \left(\sum_{i=1}^N \sum_{t=1}^T \ddot{x}^{E'}_{it} \ddot{x}^E_{it}\right)^{-1} \left(\sum_{i=1}^N \sum_{t=1}^T \ddot{x}^{E'}_{it} \ddot{y}_{it}\right) = \left(\sum_{i=1}^N X^{E'}_i Q_T X^E_i\right)^{-1} \left(\sum_{i=1}^N X^{E'}_i Q_T y_i\right), \tag{4}$$

where $\ddot{x}^E_{it} = x^E_{it} - T^{-1}\sum_{t=1}^T x^E_{it}$, $X^E_i$ is a $T \times p$ matrix stacked by $x^E_{it}$, and $Q_T$ is a time-demeaning matrix that transforms $x^E_{it}$ into $\ddot{x}^E_{it}$.

Applying the weak law of large numbers, we have $\text{plim}\hat{\beta}^E_{\text{FE}} = \beta + (E[X^{E'}_i Q_T X^E_i])^{-1} (E[X^{E'}_i Q_T X^B_i])$, where $X^B_i$ is a $T \times p$ matrix stacked by row vectors $x^B_{it}$. Andreou *et al.* (2010) obtain a similar result for the time series MIDAS regression model. We extend their framework to the panel setting. While the bias term follows from the standard omitted variable bias formula, there are two distinct features as pointed out by Andreou *et al.* (2010). Firstly, the omitted variable $x^B_{it}$ has the same coefficient as the regressor $x^E_{it}$. Secondly, the omitted variable depends on the true weighting scheme $\alpha$.

[3]It should be mentioned that the GMM estimator developed by Khalaf *et al.* (2021) can also accommodate the same random weighting coefficients setting, but our paper is the first one to explicitly consider such generalizations.

An immediate implication is that if the true weights are equal, then $X_i^B$ is a matrix of zeros and there will be no bias for $\hat{\beta}^E_{\text{FE}}$. Intuitively, if the true weights are close to equal weights, the inconsistency of the usual FE estimator is small. A typical row of $Q_T X_i^B$ is $(x_{it}(\alpha^*) - \bar{x}_i(\alpha^*)) - \left(x_{it}^E - \bar{x}_i^E\right)$, where $\bar{x}_i(\alpha^*)$ and $\bar{x}_i^E$ are the sample averages of $x_{it}(\alpha^*)$ and $x_{it}^E$, respectively. This implies that if the fluctuations of the regressors around their time averages under the true weighting scheme are the same as the deviations of the regressors from their averages under the equal weights, the FE estimator based on equal weights is consistent.

*Remark 1.* In the time series MIDAS model, when the high frequency regressors $x_{ij/m}^{(t)}$ are *i.i.d.* across time, it is shown by Andreou *et al.* (2010) that the OLS estimator with flat weighting scheme is consistent. Here we obtain a similar result for the fixed effects estimator with *i.i.d.* high frequency regressors. Without loss of generality, consider an *i.i.d.* sequence of scalar variables $x_{ij/m}^{(t)}$ with zero means and finite variance $\sigma^2$. Let $\alpha^* = (a_1^*, \ldots, a_m^*)$ be the true weights with the constraint that $\sum_{j=1}^m a_j^* = 1$. It follows that

$$
\begin{aligned}
E\left(X_i^E Q_T X_i^B\right) &= \sum_{t=1}^T E\left[\ddot{x}_{it}^E\left(x_{it}(\alpha^*) - x_{it}^E\right)\right] \\
&= \sum_{t=1}^T E\left[\left(\frac{1}{m}\sum_{j=1}^m x_{ij/m}^{(t)} - \bar{x}_i^E\right)\left(\sum_{j=1}^m\left(a_j^* - \frac{1}{m}\right)x_{ij/m}^{(t)}\right)\right] \\
&= \frac{1}{m}\sum_{t=1}^T\sum_{j=1}^m\left(a_j^* - \frac{1}{m}\right)E\left(x_{ij/m}^{(t)2}\right) + \sum_{t=1}^T\sum_{j=1}^m\left(a_j^* - \frac{1}{m}\right)E\left(\bar{x}_i^E x_{ij/m}^{(t)}\right) = 0.
\end{aligned}
$$

The above analysis indicates that the usual FE estimator of panel MIDAS model is generally inconsistent except for some special cases. When $m$ and $p$ are small, we can proceed to estimate $(\alpha', \beta')'$ by a fixed effects nonlinear least squares estimation (FE-NLS) method.

$$
\begin{aligned}
(\hat{\alpha}_{FE-NLS}, \hat{\beta}_{FE-NLS}) &= \underset{\alpha,\beta}{\operatorname{argmin}}\sum_{i=1}^N (y_i - X_i(\alpha)\beta)' Q_T (y_i - X_i(\alpha)\beta) \\
\text{s.t.}\sum_{j=1}^m a_{jk} &= 1, \quad \text{for } k = 1, 2, \ldots, p,
\end{aligned} \tag{5}
$$

where $X_i(\alpha)$ is a $T \times p$ matrix stacked by the row vector $x_{it}(\alpha)$. We can easily transform the above restricted NLS estimation to an unrestricted version by setting $a_{mk} = 1 - \sum_{j=1}^{m-1} a_{jk}$. Hereafter, we only consider the unrestricted minimization problem transformed from equation (5). Define $\theta = (\alpha', \beta')'$ and $q(W_i; \theta) = (y_i - X_i(\alpha)\beta)' Q_T (y_i - X_i(\alpha)\beta)$. We summarize the asymptotic properties of the FE-NLS estimator by the following theorem.

*Theorem 1.* *(Asymptotic properties of the FE-NLS estimator)* For the panel MIDAS model specified in equation (2), if assumptions (1)–(3) hold, $\Theta$ is compact, and $\theta^* \in \Theta$ is the unique minimizer of $E[q(W_i;\theta)]$, then $\hat{\theta}_{\text{FE-NLS}} = (\hat{\alpha}'_{\text{FE-NLS}}, \hat{\beta}'_{\text{FE-NLS}})'$ is consistent for $\theta^* = (\alpha^{*'}, \beta^{*'})'$. In addition, if $\theta^*$ is in the interior of $\Theta$, $B^* = E[\nabla^2_\theta q(W;\theta^*)]$ is positive definite, and the gradient function $\nabla_\theta q(W;\theta)$ has finite second moment, then as $N \to \infty$,

$$\sqrt{N}(\hat{\theta}_{FE-NLS} - \theta^*) \xrightarrow{d} \text{Normal}(\mathbf{0}, B^{*-1}C^*B^{*-1}), \tag{6}$$

where $C^* = E[\nabla_\theta q(W;\theta^*)'\nabla_\theta q(W;\theta^*)]$.

The proof of the above theorem follows directly from the standard results in the literature (see, e.g. Wooldridge (2010, chapter 12)) and is thus omitted. The calculations of the score functions and the Hessian matrix are left to the appendix. We note that the asymptotic variance matrix $B^{*-1}C^*B^{*-1}$ is robust to arbitrary serial correlations and heteroskedasticity of the composed error $c_i + u_{it}$. The above asymptotic variance formula can also accommodate the case that there are stochastic errors in the aggregation process as in assumption (3) because $\{\varepsilon^x_{k,it}\}_{k=1}^p$ are ultimately subsumed into the composed error. Consistent estimates of the asymptotic variances can be obtained by calculating $\hat{B} = N^{-1}\sum_{i=1}^N \nabla^2_\theta q(W_i;\hat{\theta}_{\text{FE-NLS}})$ and $\hat{C} = N^{-1}\sum_{i=1}^N \nabla_\theta q(W_i;\hat{\theta}_{\text{FE-NLS}})'\nabla_\theta q(W_i;\hat{\theta}_{\text{FE-NLS}})$. Following standard linear panel models, we refer to SEs obtained from $\hat{B}^{-1}\hat{C}\hat{B}^{-1}$ as 'cluster robust standard errors'.

With these SEs, a variety of statistics, such as the Wald statistic and the LM statistic, can be constructed for hypothesis testing. One hypothesis that is of particular interest is whether the aggregation weights are flat, i.e. $a^*_{jk} = \frac{1}{m}$ for all $j$ and $k$. Under the null that the true aggregation scheme is flat, the Wald statistic, formulated as $\left(\hat{\alpha}_{\text{FE-NLS}} - \frac{1}{m}\iota_{(m-1)p}\right)'(\hat{\Sigma}_\alpha/N)^{-1}\left(\hat{\alpha}_{\text{FE-NLS}} - \frac{1}{m}\iota_{(m-1)p}\right)$, converges in distribution to $\chi^2_{(m-1)p}$, where $\iota_l$ is a $l \times 1$ vector of unity and $\hat{\Sigma}_\alpha$ is the consistent estimate of the asymptotic variance of $\hat{\alpha}_{\text{FE-NLS}}$. We note that the degrees of freedom are $(m-1)p$ instead of $mp$ because of the constraint that $\sum_{j=1}^m a_{jk} = 1$ for all $k = 1, ..., p$.

*Remark 2.* The assumption that $\theta^*$ is the unique minimizer of $E[q(W_i;\theta)]$, along with other maintained assumptions, is a sufficient condition for the identification of $\theta^*$. The uniqueness of $\theta^*$ turns out to be intrinsically related to the positive definite property of $B^*$. Specifically, the assumption that $B^*$ is positive definite entails a rank condition on the matrix of time demeaned high-frequency explainable variables as well as a nonzero restriction for $\beta$. [4] As an illustrative example, consider the case where $m = 2$ and $p = 1$, $\theta = (a_1, \beta_1)'$ and $q(w_i, \theta) = \frac{1}{2}\sum_{t=1}^T \left[\ddot{y}_{it} - \ddot{x}^{(t)}_{i1/2}a_1\beta_1 - \ddot{x}^{(t)}_{i2/2}(1-a_1)\beta_1\right]^2$. Assuming exchangeability of expectation and differentiation, the expected Hessian is

$$B^* = \begin{bmatrix} \beta^*_1 & -\beta^*_1 \\ a^*_1 & 1-a^*_1 \end{bmatrix} \left\{ \sum_{t=1}^T \begin{bmatrix} E(\ddot{x}^{(t)2}_{i1/2}), & E(\ddot{x}^{(t)}_{i1/2}\ddot{x}^{(t)}_{i2/2}) \\ E(\ddot{x}^{(t)}_{i1/2}\ddot{x}^{(t)}_{i2/2}), & E(\ddot{x}^{(t)2}_{i2/2}) \end{bmatrix} \right\} \begin{bmatrix} \beta^*_1 & a^*_1 \\ -\beta^*_1 & 1-a^*_1 \end{bmatrix}.$$

[4]We thank an anonymous referee for suggesting this exploration. The proof of the general case is available upon request. The fact that $\alpha^*$ is not point identified when $\beta^* = 0$ is also called the Davies problem (Davies, 1977, Davies, 1987). Khalaf *et al.* (2021) provide detailed discussion of the identification issue in the dynamic panel MIDAS models.

It is now straightforward to see that the semi-definite positive matrix $B^*$ is of full rank if and only if

$$H_r \equiv \sum_{t=1}^{T} \begin{bmatrix} E(\ddot{x}_{i1/2}^{(t)2}), & E(\ddot{x}_{i1/2}^{(t)}\ddot{x}_{i2/2}^{(t)}) \\ E(\ddot{x}_{i1/2}^{(t)}\ddot{x}_{i2/2}^{(t)}), & E(\ddot{x}_{i2/2}^{(t)2}) \end{bmatrix},$$

is of full rank and $\beta_1^* \neq 0$. $H_r$ is of full rank if and only if there is no perfect collinearity between $\sum_{t=1}^{T} \ddot{x}_{i1/2}^{(t)}$ and $\sum_{t=1}^{T} \ddot{x}_{i2/2}^{(t)}$, which is exactly the same as the usual linear FE estimation. Here the new constraint that $\beta_1^* \neq 0$ is specific to our FE-NLS estimation. It is easy to see that $\beta_1^* = 0$ causes the true model to depend on fewer parameters than specified, which is an example of a poorly identified model. For the purpose of identification of $\theta^*$, we need to rule out $\beta^* = 0$.

The above theorem is also useful to quantify the efficiency loss from estimating the aggregation parameter vector $\alpha$. When $\alpha^*$ is unknown, the asymptotic variance of $\hat{\beta}_{\text{FE}-\text{NLS}}$ is the right lower block matrix of $B^{*-1}C^*B^{*-1}$. When $\alpha^*$ is known, the asymptotic variance matrix of the standard FE estimator is $(E[X_i(\alpha^*)'Q_TX_i(\alpha^*)])^{-1}(E[X_i(\alpha^*)'Q_Tu_iu_i'Q_TX_i(\alpha^*)])(E[X_i(\alpha^*)'Q_TX_i(\alpha^*)])^{-1}$. The difference between the two asymptotic matrices measures the efficiency loss from estimating the unknown aggregation weights. See Newey and McFadden (1994, section 9) for a formal proof of the efficiency comparison.

It is worth emphasizing that the strict exogeneity maintained in assumption 2 is crucial for identification. In our context, one can easily verify that the score evaluated at the true values does not necessarily have zero expectation under the low-frequency strict exogeneity assumption. For example, if we differentiate the population objective function with respect to $a_{jk}$, the corresponding first-order condition $E\left[\sum_{t=1}^{T} x_{k,ij/m}^{(t)}\ddot{u}_{it}\right]$ does not equal to zero if we only assume $E[u_{it}|x_{i1}(\alpha^*), ..., x_{iT}(\alpha^*)] = 0$. Thus the identification may fail if we use the low-frequency strict exoegeneity assumption.

*Remark 3.* In many empirical applications, some covariates are nonlinear function of others, such as quadratics and interactions. A natural question is how to assign weights for these nonlinear covariates. The consistency of our FE-NLS estimator requires correct specification of $E(y_{it}|x_i, c_i)$. However this requirement alone does not provide any useful guidance about assigning weights for nonlinear terms. To clarify this point, consider the scalar covariate case with $m = 2$ and we conjecture that quadratic terms should be included in the conditional mean. The conditional mean could be $E(y_{it}|x_i, c_i) = \left(x_{i1/2}^{(t)}a_1 + x_{i2/2}^{(t)}a_2\right)\beta_1 + \left(x_{i1/2}^{(t)}a_1 + x_{i2/2}^{(t)}a_2\right)^2\beta_2 + c_i$, or $E(y_{it}|x_i, c_i) = \left(x_{i1/2}^{(t)}a_{11} + x_{i2/2}^{(t)}a_{21}\right)\beta_1 + \left[(x_{i1/2}^{(t)})^2a_{12} + (x_{i2/2}^{(t)})^2a_{22}\right]\beta_2 + c_i$. At first glance, the specification $(x_{i1/2}^{(t)}a_1 + x_{i2/2}^{(t)}a_2)^2$ seems to be more restrictive than $(x_{i1/2}^{(t)})^2a_{12} + (x_{i2/2}^{(t)})^2a_{22}$ as the the latter has additional free weighting parameters. However, if we expand $\left(x_{i1/2}^{(t)}a_1 + x_{i2/2}^{(t)}a_2\right)^2$, there is an interaction term $x_{i1/2}^{(t)}x_{i2/2}^{(t)}$ that is not included in the latter specification. Without additional information we usually do not know which specification is correct. For empirical applications, the choice is at the researcher's discretion. It is worth noting though that most applications may start from specifying a model on low-frequency covariates, and in that case it makes more sense to introduce nonlinear terms through the former way. A practical suggestion is that, when the sample size is moderate and $m$ is relatively large, it is reasonable to use the nonlinear function of the aggregated variables to conserve the degrees of freedom without sacrificing too much flexibility.

When $m$ and $p$ are small, a natural alternative of our proposed FE-NLS estimator is the fixed effects estimator obtained from high frequency panel models.[5] For the purpose of illustration we consider the scalar covariate case. Our FE-NLS estimator can be obtained by solving min $\sum_{i=1}^{N}\sum_{t=1}^{T}(\ddot{y}_{it}-\ddot{x}_i^{(t)}\alpha\beta)^2$ under the constraint $\sum_{j=1}^{m}a_j=1$. The FE estimator based on high frequency panel models is obtained by minimizing $\sum_{i=1}^{N}\sum_{t=1}^{T}(\ddot{y}_{it}-\ddot{x}_i^{(t)}\eta)^2$. Under the condition that rank$E[\sum_{t=1}^{T}\ddot{x}_i^{(t)'}\ddot{x}_i^{(t)}]=m$, the high-frequency FE estimator $\hat{\eta}=(\hat{\eta}_1,\ldots,\hat{\eta}_m)'$ is uniquely determined with probability approaching one. Given the constraint that $\sum_{j=1}^{m}a_j=1$, there is a one-to-one mapping from $\hat{\eta}$ to $\hat{\beta}_{\text{FE}-\text{NLS}}$. Specifically, it can be shown that $\hat{\beta}_{\text{FE}-\text{NLS}}=\sum_{j=1}^{m}\hat{\eta}_j$. If $\hat{\beta}_{\text{FE}-\text{NLS}}$ differs from $\sum_{j=1}^{m}\hat{\eta}_j$, then the FE-NLS estimators are not optimal and the criterion function can be further reduced by resetting $\hat{\beta}_{\text{FE}-\text{NLS}}=\sum_{j=1}^{m}\hat{\eta}_j$ and $\hat{\alpha}_{\text{FE}-\text{NLS}}=\left(\frac{\hat{\eta}_1}{\sum_{j=1}^{m}\hat{\eta}_j},\ldots,\frac{\hat{\eta}_m}{\sum_{j=1}^{m}\hat{\eta}_j}\right)'$. The algebraic equivalence[6] between the FE-NLS estimator and the high-frequency FE estimator also highlights the importance of assuming the high-frequency strict exogeneity. If the strict exogeneity assumption is conditioning on the aggregated covariates, the high frequency FE estimator, hence the FE-NLS estimator, would be inconsistent.

When $m$ and $p$ are moderate, the dimension of $\alpha$ could be very large, which may lead to the problem of parameter proliferation. To avoid such circumstances, our proposed FE-NLS estimation method could use flexible parametric functions with a low-dimension parameter vector, e.g. $a_{jk}=f_{j,k}(\xi_k)$, to model these aggregation weights. See Ghysels *et al.* (2007) for a variety of popular choices of these parametric weighting functions. We introduce the Almon lags polynomial weighting functions in the appendix. It should be mentioned that, when parametric functions are used to model the unknown aggregation weights, the FE estimator from the high-frequency panel model is no longer identical to our FE-NLS estimator. When these parametric weighting functions are correctly specified, it is likely that the FE-NLS estimator using the parametric weighting functions are more efficient than the FE-NLS estimator with unrestricted weighting parameters as defined in equation (5). This is because NLS estimator with correctly specified constraints is asymptotically more efficient than unrestricted NLE estimator. See, for example, Newey and McFadden (1994, section 9). However, when the parametric weighting functions are misspecified, the FE-NLS estimator based on the parametric aggregation scheme would be inconsistent.

Another case that the equivalence between the high-frequency FE estimator and the FE-NLS estimator breaks down is when some covariates are nonlinear functions of other aggregated variables. Consider again the scalar covariate example in remark 2 and suppose that $E(y_{it}|x_i,c_i)=\left(x_{i1/2}^{(t)}a_1+x_{i2/2}^{(t)}a_2\right)\beta_1+(x_{i1/2}^{(t)}a_1+x_{i2/2}^{(t)}a_2)^2\beta_2+$

[5]We thank an anonymous referee for suggesting this comparison.

[6]We also present two important cases that such equivalence result no longer holds for these two estimators in the following discussions.

$c_i$. The objective function of the FE-NLS estimation is to minimize $\sum_{i=1}^{N}\sum_{t=1}^{T}\left[y_{it}-\left(x_{i1/2}^{(t)}a_1+x_{i2/2}^{(t)}a_2\right)\beta_1-\left(x_{i1/2}^{(t)}a_1+x_{i2/2}^{(t)}a_2\right)^2\beta_2-c_i\right]^2$, while the objective function of the high-frequency FE estimator is $\sum_{i=1}^{N}\sum_{t=1}^{T}\left[y_{it}-x_{i1/2}^{(t)}\eta_{11}-x_{i2/2}^{(t)}\eta_{21}-\left(x_{i1/2}^{(t)}\right)^2\eta_{12}-\left(x_{i2/2}^{(t)}\right)^2\eta_{22}-x_{i1/2}^{(t)}x_{i2/2}^{(t)}\eta_{32}-c_i\right]^2$. In this case the high-frequency FE estimator is generally different from our FE-NLS estimator, unless additional constraints, such as $\eta_{12}/\eta_{22}=(\eta_{11}/\eta_{21})^2$, are imposed for the high-frequency FE estimation.

## III. The Chamberlain and Mundlak approaches to the panel MIDAS models

The main advantage of the panel data models over the cross section models is that we can explicitly account for the time-constant individual fixed effects, which are usually correlated with the explanatory variables. As an alternative to the fixed effects estimator, Mundlak (1978) develops a novel modelling device for the individual fixed effects. For the balanced panel model, Mundlak (1978) projects the individual fixed effects onto the time averages of the explanatory variables $L(c_i|x_{i1},...,x_{iT})=\bar{x}_i\gamma^*$. Make substitution of the Mundlak device into the original linear panel model in equation (1)

$$y_{it}=x_{it}\beta^*+\bar{x}_i\gamma^*+a_i+u_{it}, \tag{7}$$

where $a_i=c_i-\bar{x}_i\gamma^*$ is the projection error. It is well-known that the pooled OLS estimator of $\beta^*$ is identical to the fixed effects estimator. Wooldridge (2019) obtains a similar equivalence result for the unbalanced panel models when the selection indicator is conditionally independent with $u_{it}$.

This paper attempts to extend the Mundlak device to the panel MIDAS regression model. Define $\bar{x}_i(\alpha^*)=T^{-1}\sum_{t=1}^{T}x_{it}(\alpha^*)$. If we make linear projection of $c_i$ onto the entire history of $x_{it}(\alpha^*)$ and assume that $L(c_i|x_{i1}(\alpha^*),...,x_{iT}(\alpha^*))=\bar{x}_i(\alpha^*)\gamma^{*L}$, then the augmented regression is given by the following *low-frequency Mundlak equation*

$$y_{it}=x_{it}(\alpha^*)\beta^*+\bar{x}_i(\alpha^*)\gamma^{*L}+\varepsilon_{it}^{L}, \tag{8}$$

where $\varepsilon_{it}^{L}$ is composed of the projection error and $u_{it}$. At first glance one might conjecture that the NLS estimator of $\beta^*$ in equation (8) is still numerically equivalent to the FE-NLS estimator. However, as to be shown later, this equivalence holds only if the true weights $\alpha^*$ are known. While the Frisch–Waugh theorem still applies to the estimation of the slope parameters, the NLS estimator of $\beta^*$ in the augmented regression (8) is no longer identical to the FE-NLS estimator because of the presence of the unknown parameters $\alpha^*$.

Here the interesting case is that we project the individual fixed effects onto the high-frequency series $x_i$ instead of the history of the aggregated explanatory variables $x_{it}(\alpha)$. An obvious reason is that the projection error of a long projection has a smaller variance, which follows from the property of linear projection. If we use $\bar{x}_i=(Tm)^{-1}\sum_{t=1}^{T}\sum_{j=1}^{m}x_{ij/m}^{(t)}$ as proxy variables for $c_i$, then the *high-frequency Mundlak equation* can be written as

$$y_{it}=x_{it}(\alpha^*)\beta^*+\bar{x}_i\gamma^{*H}+\varepsilon_{it}^{H}. \tag{9}$$

With some straightforward algebra, one can show that the pooled OLS estimator of $\beta^*$ in equation (9) is no longer identical to the FE-NLS estimator even when $\alpha^*$ is known but is not flat.

A more flexible alternative to the Mundlak device is proposed by Chamberlain (1982). Instead of assuming the linear projection of the individual fixed effects depends only on the time averages of the explanatory variables, Chamberlain (1982) considers an unrestricted linear projection of the individual fixed effects. For the single-frequency panel model, the linear projection is written as follows

$$L(c_i|x_{i1},...,x_{iT}) = \psi^* + x_{i1}\lambda_1^* + \cdots + x_{iT}\lambda_T^*. \quad (10)$$

Making use of equation (10), we obtain the following estimating equation for single-frequency panel models

$$y_{it} = x_{it}\beta^* + \psi^* + x_{i1}\lambda_1^* + \cdots + x_{iT}\lambda_T^* + \varepsilon_{it}. \quad (11)$$

Since the composed error $\varepsilon_{it}$ is uncorrelated with all the regressors, the pooled OLS estimator of $\beta^*$ in equation (11) is consistent and can be shown to be identical to the fixed effects estimator.

Extensions of the Chamberlain's approach for unbalanced panel data models have been investigated by Abrevaya (2013). Define the selection indicator $s_{it} = 1$ if $(y_{it}, x_{it})$ is observed and $s_{it} = 0$ otherwise. As demonstrated by Abrevaya (2013), a simple linear projection of the individual fixed effects onto $(s_{i1}x_{i1},...,s_{iT}x_{iT})$ results in inconsistent estimation. Abrevaya (2013) further provides a modified Chamberlain projection approach within the GMM estimation framework. Here we provide a comprehensive study of the Chamberlain's approach for the panel MIDAS regression model. Specifically we consider the following two possible specifications.

$$L(c_i|x_{i1}(\alpha^*),...,x_{iT}(\alpha^*)) = \psi^{*L} + x_{i1}(\alpha^*)\lambda_1^{*L} + \cdots + x_{iT}(\alpha^*)\lambda_T^{*L}, \quad (12)$$

$$L(c_i|x_i) = \psi^{*H} + x_i\lambda^{*H}, \quad (13)$$

where equation (12) gives the projection of the individual fixed effects onto the aggregated explanatory variables and equation (13) represents the linear projection onto the high frequency series. We use the superscripts $L$ and $H$ to distinguish the parameters associated with the linear projections onto the low-frequency and the high-frequency series, respectively. Substituting equations (12) and (13) into the original regression model yields the *low-frequency Chamberlain equation* and the *high-frequency Chamberlain equation*

$$y_{it} = x_{it}(\alpha^*)\beta^* + \psi^{*L} + x_{i1}(\alpha^*)\lambda_1^{*L} + \cdots + x_{iT}(\alpha^*)\lambda_T^{*L} + e_{it}^L. \quad (14)$$

$$y_{it} = x_{it}(\alpha^*)\beta^* + \psi^{*H} + x_i\lambda^{*H} + e_{it}^H. \quad (15)$$

Since the regression errors are uncorrelated with the covariates in these two equations and assumption (3) holds, it follows that the pooled OLS estimators of $\beta^*$ in equation

(14) and (15) are all consistent for $\beta^*$ for the same reason as in theorem 1. However, only the pooled OLS estimator of $\beta^*$ in equation (15) is numerically identical to the FE-NLS estimator when $\alpha^*$ is unknown.

Here we make a few instructive comparisons. For single-frequency panel data models, the pooled OLS estimator of the original panel model augmented with the time averages or the entire history of the explanatory variables are both numerically identical to the fixed effects estimator. For the panel MIDAS regression model, the high-frequency Mundlak estimator of $\beta^*$ is not identical to the fixed effects estimator unless the true aggregation weights of the high frequency variables are flat. The use of low-frequency Mundlak device no longer delivers the fixed effects estimator unless the true aggregation weights are known. The Chamberlain's projection technique still delivers the fixed effects estimator when the individual fixed effects is projected onto the high frequency series. Define $x_i(\alpha) = (x_{i1}(\alpha), ..., x_{iT}(\alpha))$. We summarize these algebraic equivalence results in the following proposition.

*Proposition 1. (Equivalence results for panels MIDAS regression models)* When the true aggregation weights $\alpha^*$ are known, the following estimators of $\beta^*$ are numerically identical. (a1) the FE-NLS estimator (NLS of $\ddot{y}_{it}$ on $\ddot{x}_{it}(\alpha)$); (b1) the low-frequency Mundlak regression estimator (NLS of $y_{it}$ on 1 and $x_{it}(\alpha)$ and $\bar{x}_i(\alpha)$); (c1) the low-frequency Chamberlain regression estimator (NLS of $y_{it}$ on $x_{it}(\alpha)$, 1 and $x_i(\alpha)$); (d1) the high-frequency Chamberlain regression estimator (NLS of $y_{it}$ on 1, $x_{it}(\alpha)$ and $x_i$). When the true aggregation weights $\alpha^*$ are unknown, only the high-frequency Chamberlain regression estimator (d1) is numerically equivalent to the FE-NLS estimator (a1).

All proofs are delegated to the appendix. Remarkably, the above proposition remains to be true when we use smooth parametric weighting functions $\{f_{jk}(\xi_k)\}_{j=1,k=1}^{m,p}$, such as the Almon polynomial functions, to model the aggregation weights $\{a_{jk}\}_{j=1,k=1}^{m,p}$.[7] Define $\xi = (\xi_1', ..., \xi_p')'$, $x_{k,it}(\xi_k) = \sum_{j=1}^{m} f_{jk}(\xi_k) x_{k,ij/m}^{(t)}$, $\bar{x}_i(\xi) = T^{-1} \sum_{t=1}^{T} x_{it}(\xi)$ and $x_i(\xi) = (x_{i1}(\xi), ..., x_{iT}(\xi))$. The following proposition summarizes the equivalence results when parametric weighting functions are used in the panel MIDAS models.

*Proposition 2. (Equivalence results for panels MIDAS regression models with parametric aggregation functions)* Let $\{f_{jk}(\xi_k)\}_{j=1,k=1}^{m,p}$ be a set of differentiable functions such that $\sum_{j=1}^{m} f_{jk}(\xi_k) = 1$ for $k = 1, ..., p$. When $\xi$ is known, the following estimators of $\beta$ are numerically identical. (a2) the FE-NLS estimator (NLS of $\ddot{y}_{it}$ on $\ddot{x}_{it}(\xi)$); (b2) the low-frequency Mundlak regression estimator (NLS of $y_{it}$ on 1 and $x_{it}(\xi)$ and $\bar{x}_i(\xi)$); (c2) the low-frequency Chamberlain regression estimator (NLS of $y_{it}$ on $x_{it}(\xi)$, 1 and $x_i(\xi)$); (d2) the high-frequency Chamberlain regression estimator (NLS of $y_{it}$ on 1, $x_{it}(\xi)$ and $x_i$). When $\xi$ is unknown and is estimated along with $\beta$, only the high-frequency Chamberlain regression estimator (d2) is numerically equivalent to the FE-NLS estimator (a2).

It should be emphasized that the focus of this section is on algebraic equivalences. We examine the equivalence relationship among various estimators when the (parametric) aggregation parameters are known/unknown. When these estimators differ from each

[7]This finding is motivated by the comment of an anonymous referee. We thank the referee for this insightful suggestion.

other, it is interesting to compare the relative efficiency of these estimators. In principle, the projection error has a smaller variance when the unobserved fixed effect is projected onto a longer list of variables, which would lead to more efficient estimator of $\beta^*$. On the other hand, adding more unknown parameters to the regression models usually reduces the precision of the estimates of $\beta^*$. We compare the efficiency of these estimators via a small simulation study in the next section.

## IV. Simulation results

In this section we investigate the finite sample properties of the proposed estimators through various simulation designs. Specifically we evaluate the small sample performances of the following six estimators (i) the naive fixed effects (FE-Naive) estimator, which is the fixed effects estimator with an equal weighting scheme; (ii) the fixed effects nonlinear least squares (FE-NLS) estimator as expressed in equation (5); (iii) the low-frequency Mundlak (LF-Mundlak) estimator as defined in equation (8); (iv) the high-frequency Mundlak (HF-Mundlak) estimator as defined in equation (9); (v) the low-frequency Chamberlain (LF-Chamberlain) estimator obtained from equation (14); (vi) the high-frequency Chamberlain (HF-Chamberlain) estimator as described in equation (15). To economize the space, the simulation results of the FE-NLS estimator are reported together with the high-frequency Chamberlain estimator because they are numerically identical as shown in proposition 1. Throughout this section, the panel sample size is set as $(N, T) = (500, 3)$ and the replication number for each design is 1,000. The mixed frequency panel data model is specified as the following

$$y_{it} = \beta^* x_{it}(\alpha^*) + c_i + u_{it}. \tag{16}$$

$$x_{it}(\alpha^*) = x_{i1/4}^{(t)} a_1^* + x_{i2/4}^{(t)} a_2^* + x_{i3/4}^{(t)} a_3^* + x_{i4/4}^{(t)} a_4^*. \tag{17}$$

We consider the estimation of the scalar $\beta^* = 1$ when the explanatory variable is sampled at four times more often (i.e. $m = 4$) than the dependent variable. In what follows we consider simulation designs with four different weighting schemes: $\alpha^* = (0.25, 0.25, 0.25, 0.25)$, $\alpha^* = (0.1, 0.2, 0.4, 0.3)$, $\alpha^* = (0.2, 0.3, 0.2, 0.3)$ and $\alpha^* = (0.1, 0.4, 0.1, 0.4)$.

We examine the small sample performances of these estimators based on four different data generating processes. DGP1: $x_{ij/4}^{(t)}$ are *i.i.d.* Normal(0,1) across $j$ and $t$. $c_i = x_i \lambda + \varepsilon_i$, where $\lambda$ is a $12 \times 1$ vector of ones and $\varepsilon_i$ is a standard normal random variable that is independent of all other variables. The idiosyncratic term $u_{it}$ is *i.i.d.* Normal(0,9) and is independent of everything. DGP2: $(x_i, c_i)$ are drawn from a joint multivariate normal distribution with zero means and covariance matrix $R$. All diagonal elements of $R$ are set to be one. The correlation coefficient among the high-frequency variables is 0.6 and the correlation between the high-frequency variable and the fixed effects is 0.4. $u_{it}$ is independent of everything and is *i.i.d.* Normal(0,9). DGP3: Define $x_i^{(t)} = [x_{i1/4}^{(t)}, x_{i2/4}^{(t)}, x_{i3/4}^{(t)}, x_{i4/4}^{(t)}]'$ and $e_i^{(t)} = \left[e_{i1/4}^{(t)}, e_{i2/4}^{(t)}, e_{i3/4}^{(t)}, e_{i4/4}^{(t)}\right]'$. $x_i^{(t)} = Hx_i^{(t-1)} + 0.5c_i\iota + e_i^{(t)}$, where $\iota$ is a $4 \times 1$ vector of ones, $e_{ij/4}^{(t)}$ and $c_i$ are standard normal and independent of each other. The initial values $x_{ij/4}^{(1)}$ are drawn from Normal(1

$+\ 0.5c_i$,1). The idiosyncratic errors $u_{it}$ follow Normal(0,9). The $4 \times 4$ coefficient matrix $H = [\rho_1, \rho_2, \rho_3, \rho_4]$, with $\rho_1 = [0.6, 0.2, 0.3, 0.4]'$, $\rho_2 = [0.9, 0.2, 0.3, 0.3]'$, $\rho_3 = [0.6, 0.1, 0.1, 0.4]'$ and $\rho_4 = [0.5, 0.3, 0.3, 0.4]'$. DGP4 is exactly the same as DGP3 except that we use a different $H$ matrix by resetting $\rho_1 = [0.6, 0.6, 0.6, 0.6]'$, $\rho_2 = [0.9, 0.9, 0.9, 0.9]'$, $\rho_3 = [0.1, 0.1, 0.1, 0.1]'$ and $\rho_4 = [0.5, 0, 0, 0]'$.

From the previous discussion in section II, our FE-NLS estimation continues to produce reliable estimates under random weighting coefficients as long as the randomness of these coefficients are independent of all covariates. Here we consider several simulation designs with random weighting coefficients. Specifically, the data generating process of $(x_i, c_i, u_i)$ is the same as in DGP3 and DGP4 except that there are randomness in the weighting coefficients. For example, now $\alpha^* = (0.1, 0.2, 0.4, 0.3)$ stands for the mean value of all the random coefficients. For a given individual $i$, the associated weighting parameters are given as $a_{i,1} = 0.1 + v_{i,1}$, $a_{i,2} = 0.2 + v_{i,2}$, $a_{i,3} = 0.4 + v_{i,3}$ and $a_{i,4} = 1 - a_{i1} - a_{i,2} - a_{i,3}$, where $(v_{i,1}, v_{i,2}, v_{i,3})$ are independent of all covariates and individual fixed effects. Each is independently drawn from $Uniform(-0.1, 0.1)$. We will call the combination of DGP3 with random weighting coefficients as DGP5, and the combination of DGP4 with random weighting as DGP6.

For all the above data generating processes, the explanatory variables are correlated with the fixed effects. DGP1 restricts that the high-frequency variables are independent across time. Based on the previous analysis in section II, we would expect that the FE-Naive estimator is also consistent. DGP2 allows for exchangeable correlations among the high frequency variables $x^{(t)}_{ij/m}$, i.e. $\mathrm{Cov}(x^{(t)}_{ij/m}, x^{(s)}_{il/m}) = 0.6$ if $t \neq s$ or $j \neq l$. DGP3 and DGP4 generate the high frequency variables from a VAR(1) process with additive individual fixed effects.

We compare the biases, SD and the actual coverage rates (CR) of 95% confidence intervals of these estimators via various simulation plots. The simulation results for deterministic weighting designs are collected in Tables 1 and 2. Firstly, we note that when $\alpha = (0.25, 0.25, 0.25, 0.25)$, the naive fixed effects estimator, as expected, has better finite sample performance than the LF-Mundlak, HF-Mundlak and LF-Chamberlain estimators. It is worth mentioning that, even under the equal weighting scheme, the HF-Chamberlain, or equivalently, the FE-NLS estimator, provides satisfying small sample performances in terms of bias and SD, although the actual coverage rates of HF-Chamberlain are not as good as the FE-Naive estimator.

Secondly, when the true weighting scheme deviates from the equal aggregation weights, the inconsistency of the FE-Naive estimator varies across different simulation designs. The bias of the FE-Naive estimator is negligible in DGP1 since the high-frequency variables are independent across time, which is in line with the discussion of the remark in section II. Surprisingly, the FE-Naive estimator continues to perform well in DGP2 although its coverage rate is now slightly larger than the nominal level. Moving to DGP3 and DGP4, we can see that the FE-NLS estimator has larger biases than the other competing estimators. The actual coverage rate of the naive fixed effects estimator deteriorates in DGP4, while other proposed estimators continue to provide reliable inferences.

We can compare the four proposed estimators across different simulation plots. Our simulation results show that, in most cases, the HF-Chamberlain estimator and

TABLE 1
*Simulation results for DGP1 and DGP2*

| Weighting scheme | FE-Naive | | | LF-Mundlak | | | HF-Munlak | | | LF-Chamberlain | | | HF-Chamberlain/ FE-NLS | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Bias | SD | CR | Bias | SD | CR | Bias | SD | CR | Bias | SD | CR | Bias | SD | CR |
| DGP1 (0.25, 0.25, 0.25, 0.25) | −0.008 | 0.191 | 0.949 | −0.009 | 0.191 | 0.951 | −0.008 | 0.190 | 0.949 | −0.009 | 0.191 | 0.952 | −0.007 | 0 .190 | 0.948 |
| DGP1 (0.1, 0.2, 0.4, 0.3) | −0.009 | 0.192 | 0.948 | −0.007 | 0.192 | 0.946 | −0.001 | 0.184 | 0.945 | 0.007 | 0.192 | 0.946 | −0.000 | 0 .184 | 0.944 |
| DGP1 (0.2, 0.3, 0.2, 0.3) | −0.008 | 0.191 | 0.947 | −0.006 | 0.191 | 0.947 | −0.007 | 0.189 | 0.946 | −0.006 | 0.191 | 0.947 | −0.006 | 0 .189 | 0.947 |
| DGP1 (0.1, 0.4, 0.1, 0.4) | −0.008 | 0.191 | 0.946 | 0.020 | 0.192 | 0.942 | 0.005 | 0.179 | 0.944 | 0.020 | 0.192 | 0.942 | 0.007 | 0 .180 | 0.942 |
| DGP2 (0.25, 0.25, 0.25, 0.25) | −0.005 | 0.307 | 0.950 | −0.027 | 0.318 | 0.924 | 0.004 | 0.297 | 0.946 | −0.039 | 0.321 | 0.922 | 0.006 | 0 .295 | 0.949 |
| DGP2 (0.1, 0.2, 0.4, 0.3) | −0.005 | 0.307 | 0.952 | 0.015 | 0.303 | 0.935 | 0.025 | 0.284 | 0.949 | 0.008 | 0.303 | 0.936 | 0.026 | 0 .286 | 0.954 |
| DGP2 (0.2, 0.3, 0.2, 0.3) | −0.005 | 0.307 | 0.952 | −0.020 | 0.315 | 0.930 | 0.006 | 0.295 | 0.947 | −0.032 | 0.318 | 0.922 | 0.009 | 0 .294 | 0.954 |
| DGP2 (0.1, 0.4, 0.1, 0.4) | −0.005 | 0.307 | 0.954 | 0.037 | 0.291 | 0.938 | 0.039 | 0.275 | 0.954 | 0.028 | 0.293 | 0.937 | 0.039 | 0 .278 | 0.954 |

TABLE 2

*Simulation results for DGP3 and DGP4*

| | FE-Naive | | | LF-Mundlak | | | HF-Munlak | | | LF-Chamberlain | | | HF-Chamberlain/ FE-NLS | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Weighting scheme | Bias | SD | CR | Bias | SD | CR | Bias | SD | CR | Bias | SD | CR | Bias | SD | CR |
| DGP3 (0.25, 0.25, 0.25, 0.25) | 0.001 | 0.076 | 0.949 | −0.007 | 0.080 | 0.935 | −0.003 | 0.081 | 0.937 | −0.004 | 0.081 | 0.939 | 0.002 | 0 .082 | 0.941 |
| DGP3 (0.1, 0.2, 0.4, 0.3) | −0.021 | 0.076 | 0.938 | −0.010 | 0.080 | 0.937 | 0.002 | 0.078 | 0.940 | −0.008 | 0.081 | 0.935 | 0.003 | 0 .081 | 0.943 |
| DGP3 (0.2, 0.3, 0.2, 0.3) | 0.026 | 0.076 | 0.933 | −0.003 | 0.079 | 0.936 | −0.002 | 0.080 | 0.938 | 0.000 | 0.080 | 0.936 | 0.002 | 0 .082 | 0.941 |
| DGP3 (0.1, 0.4, 0.1, 0.4) | 0.075 | 0.076 | 0.826 | 0.011 | 0.075 | 0.943 | 0.002 | 0.077 | 0.939 | 0.011 | 0.076 | 0.939 | 0.005 | 0 .079 | 0.943 |
| DGP4 (0.25, 0.25, 0.25, 0.25) | 0.001 | 0.055 | 0.947 | 0.071 | 0.102 | 0.908 | 0.062 | 0.113 | 0.908 | −0.002 | 0.114 | 0.899 | 0.004 | 0 .123 | 0.938 |
| DGP4 (0.1, 0.2, 0.4, 0.3) | −0.287 | 0.055 | 0.001 | 0.034 | 0.108 | 0.934 | 0.063 | 0.113 | 0.906 | −0.007 | 0.131 | 0.894 | 0.005 | 0 .122 | 0.942 |
| DGP4 (0.2, 0.3, 0.2, 0.3) | 0.050 | 0.055 | 0.831 | 0.078 | 0.101 | 0.900 | 0.062 | 0.113 | 0.908 | 0.016 | 0.110 | 0.917 | 0.004 | 0 .122 | 0.936 |
| DGP4 (0.1, 0.4, 0.1, 0.4) | 0.149 | 0.055 | 0.207 | 0.095 | 0.099 | 0.878 | 0.065 | 0.110 | 0.904 | 0.056 | 0.104 | 0.910 | 0.011 | 0 .115 | 0.934 |

HF-Mundlak estimator have smaller biases compared to the LF-Chamberlain and LF-Mundlak estimators, respectively. For DGP1 and DGP2, these high-frequency estimators are also more efficient than the low-frequency estimators, which is consistent with the theoretical prediction. However, we do not find a similar pattern of efficiency in DGP3 and DGP4. In terms of the coverage rates, all estimators perform equally well under DGP1 and DGP2 but the HF-Chamberlain estimator outperforms other estimators in DGP3 and DGP4.

The simulation results for random weighting coefficients are collected in Table 3. Consistent with our theoretical prediction in section II, our proposed estimators continue to provide reliable estimation results under random weighting schemes. The bias and SD of the FE-NLS estimator does not change much in these new settings and its coverage rates are now marginally larger than the nominal level, which suggests that inference based on our FE-NLS estimator is slightly conservative.

Our simulation study suggests that the FE-NLS estimator provides useful inferences under various data generating processes. In other cases considered here, the naive fixed effects estimator produces unreliable inferences due to large biases. Our proposed estimators, on the other hand, have satisfying small sample performances across all designs. Among all the four viable estimators, the FE-NLS estimator (the HF-Chamberlain estimator) has relatively small bias and SD. Moreover, the actual coverage rates of the FE-NLS estimator are all very close to their theoretical level of 0.95 across all designs.

## V. Empirical applications

In recent years, climate change has been one of the most discussed issues in various fields. In economics, enormous effort has been put to address the question ‘how does the climate impact society and economics’. Most studies focus on how weather conditions have shifted in a given area and their consequential impacts on economic growth, agricultural output, industrial output, productivity, and other outcomes. Among them, Dell *et al.* (2012) find that that a 1$^\circ$C increase in average annual temperature reduces economic growth by 1.3 percentage points for poor countries. Their study calculates the annual average of the temperature as a measure of the overall weather condition of the year. This flat aggregating scheme ignores the fact that temperature in different seasons may have different impacts on economic activities.

For our purpose, we collect the Terrestrial Air Temperature and Precipitation: 1900–2017 Gridded Monthly Time Series (1900–2017) (V 5.01). Basically, the datasets provide monthly mean temperature and precipitation data at 0.5 $\times$ 0.5 degree resolution, from which we can aggregate to country-level using the Global Rural-urban Mapping Project. The datasets are the updated version of the data used in Dell *et al.* (2012) and thus minor variation may exist for the datasets. We follow exactly the same procedure as in Dell *et al.* (2012) to construct the climate panel dataset except that we keep these weather variables at quarterly frequency. By assigning different weights to weather variables of different quarters, it allows us to obtain a more accurate measurement of the yearly weather condition.

The empirical framework considered here is the same as the one adopted by Bond, Leblebicioğlu, and Schiantarelli (2010) and Dell *et al.* (2012). A simple economy is

TABLE 3

*Simulation results for random weighting coefficients*

| | FE-Naive | | | LF-Mundlak | | | HF-Munlak | | | LF-Chamberlain | | | HF-Chamberlain/ FE-NLS | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Mean weighting values | Bias | SD | CR | Bias | SD | CR | Bias | SD | CR | Bias | SD | CR | Bias | SD | CR |
| DGP5 (0.25, 0.25, 0.25, 0.25) | 0.001 | 0.075 | 0.952 | −0.007 | 0.079 | 0.937 | −0.003 | 0.078 | 0.953 | −0.004 | 0.079 | 0.941 | 0.002 | 0.079 | 0.956 |
| DGP5 (0.1, 0.2, 0.4, 0.3) | −0.022 | 0.076 | 0.935 | −0.010 | 0.078 | 0.946 | 0.002 | 0.076 | 0.955 | −0.007 | 0.079 | 0.945 | 0.003 | 0.078 | 0.957 |
| DGP5 (0.2, 0.3, 0.2, 0.3) | 0.025 | 0.075 | 0.945 | −0.003 | 0.078 | 0.944 | −0.002 | 0.078 | 0.953 | 0.001 | 0.078 | 0.944 | 0.002 | 0.079 | 0.956 |
| DGP5 (0.1, 0.4, 0.1, 0.4) | 0.075 | 0.075 | 0.826 | 0.011 | 0.074 | 0.952 | 0.002 | 0.076 | 0.954 | 0.012 | 0.075 | 0.949 | 0.005 | 0.077 | 0.954 |
| DGP6 (0.25, 0.25, 0.25, 0.25) | 0.001 | 0.055 | 0.953 | 0.073 | 0.103 | 0.909 | 0.065 | 0.113 | 0.909 | 0.003 | 0.111 | 0.906 | 0.005 | 0.120 | 0.957 |
| DGP6 (0.1, 0.2, 0.4, 0.3) | −0.287 | 0.055 | 0.000 | 0.037 | 0.107 | 0.938 | 0.066 | 0.112 | 0.906 | −0.005 | 0.130 | 0.899 | 0.007 | 0.119 | 0.958 |
| DGP6 (0.2, 0.3, 0.2, 0.3) | 0.051 | 0.055 | 0.848 | 0.080 | 0.102 | 0.896 | 0.065 | 0.113 | 0.909 | 0.020 | 0.108 | 0.922 | 0.006 | 0.119 | 0.956 |
| DGP6 (0.1, 0.4, 0.1, 0.4) | 0.150 | 0.055 | 0.227 | 0.097 | 0.100 | 0.863 | 0.067 | 0.111 | 0.902 | 0.059 | 0.104 | 0.909 | 0.012 | 0.113 | 0.951 |

modelled as follows

$$Y_{it} = e^{\beta T_{it}} A_{it} L_{it}, \tag{18}$$

$$\Delta A_{it}/A_{it} = g_i + \delta T_{it}, \tag{19}$$

where $Y_{it}$ is the aggregate output for country $i$ at year $t$, $A_{it}$ denotes the labor productivity, $L_{it}$ measures the working population and $T_{it}$ is the weather variable. $g_i$ is the country specific productivity growth rate. As described in Dell *et al.* (2012), equation (18) measures the level effect of weather on production and equation (19) captures the growth effect of weather. Taking logs of equation (18) and making difference with respect to time, we have the following distributed lag equation

$$y_{it} = (\beta + \delta)T_{it} - \beta T_{it-1} + g_i, \tag{20}$$

where $y_{it}$ is the growth rate of per-capita output. From the above equation we can see that the level effects $\beta$ is identified as the negative of the coefficient on the lag weather variable and the growth effect $\delta$ can be estimated as the summation of all parameters on the weather variables.

The weather variables (temperature and precipitation) are available at a higher frequency than the GDP growth rate. We adopt the new estimation method to accommodate the possibility that weather variables of different periods are of different importance. Let $T_{ij}^{(t)}$ be the weather variable of quarter $j$ in year $t$ for country $i$ and assume $T_{it} = a_1 T_{i1}^{(t)} + a_2 T_{i2}^{(t)} + a_3 T_{i3}^{(t)} + a_4 T_{i4}^{(t)}$. In this section we compare the FE-Naive and the FE-NLS estimates of various specifications. Here we focus only on the FE-NLS estimator (or its equivalent variant HF-Chamberlain estimator) because it has better finite sample performances as demonstrated in previous simulation studies.

First we revisit the benchmark model that there are no lag effects of the temperature (i.e. imposing $\beta = 0$), which corresponds to the table 2 in Dell *et al.* (2012). Following Dell *et al.* (2012), the dummy variable 'poor' is defined as having below-median PPP-adjusted per capita GDP in the first year the country enters the dataset. The dummy variable 'hot' is defined as having above median temperature in the 1950s. The FE-Naive estimates and FE-NLS estimates are reported in Table 4. The $P$ values associated with the Wald statistic testing that all temperature weights are equal are reported in the third row from the bottom. These large p values indicate that the true aggregation weights are not statistically different from the flat weights used in the FE-Naive estimation.

The last two rows of Table 4 give the sum of the main effects of weather variables (temperature and precipitation) and their interactions with the poor dummy. From specifications (2)–(4), we can see that the increase of the temperature has a negative effect for the poor countries while the effects of the precipitation are not statistically significant. Based on the results from Table 4, we can see that the net effect of a 1°C rise in temperature is to decrease the growth rates by 1.53–1.94 percentage points for poor countries. Except for specification (4), the absolute values of the FE-Naive estimates are marginally larger than the FE-NLS estimates, which indicates that the flat weighting

TABLE 4

*FE-Naive and FE-NLS estimations without lags*

| Dependent variable is the annual growth rate | FE-Naive (1) | FE-Naive (2) | FE-Naive (3) | FE-Naive (4) | FE-Naive (5) | FE-NLS (1) | FE-NLS (2) | FE-NLS (3) | FE-NLS (4) | FE-NLS (5) |
|---|---|---|---|---|---|---|---|---|---|---|
| Temperature | −0.524* | 0.252 | 0.248 | 0.149 | 0.552* | −0.493 | 0.310 | 0.293 | 0.171 | 0.540* |
| | (0.306) | (0.307) | (0.306) | (0.294) | (0.283) | (0.333) | (0.328) | (0.332) | (0.301) | (0.303) |
| Temperature interacted with poor country dummy | | −1.999*** | −1.988*** | −2.085*** | −2.134*** | | −2.045*** | −1.982*** | −2.112*** | −2.068*** |
| | | (0.540) | (0.537) | (0.553) | (0.559) | | (0.583) | (0.583) | (0.584) | (0.595) |
| Hot country dummy | | | | 0.388 | | | | | 0.559 | |
| | | | | (0.574) | | | | | (0.635) | |
| Agricultural country dummy | | | | | -0.334 | | | | | -0.319 |
| | | | | | (0.480) | | | | | (0.473) |
| Precipitation | | | −0.038 | −0.210*** | −0.060 | | | −0.118 | −0.838*** | −0.323* |
| | | | (0.057) | (0.079) | (0.049) | | | (0.156) | (0.281) | (0.184) |
| Precipitation interacted with Poor country dummy | | | 0.051 | 0.063 | 0.074 | | | 0.379 | 0.405 | 0.426* |
| | | | (0.082) | (0.081) | (0.091) | | | (0.248) | (0.304) | (0.241) |
| Hot country dummy | | | | 0.235*** | | | | | 0.989*** | |
| | | | | (0.080) | | | | | (0.290) | |
| Agricultural country dummy | | | | | −0.013 | | | | | 0.185 |
| | | | | | (0.079) | | | | | (0.191) |
| Observations | 4924 | 4924 | 4924 | 4577 | 4924 | 4924 | 4924 | 4924 | 4924 | 4577 |
| $R^2$ | 0.141 | 0.145 | 0.145 | 0.146 | 0.157 | 0.141 | 0.145 | 0.146 | 0.148 | 0.159 |
| Test (*P*-value) of equal temp weights | | | | | | 0.972 | 0.307 | 0.344 | 0.283 | 0.250 |
| Temperature effect in poor countries | | −1.747*** | −1.740*** | −1.937** | −1.582*** | | −1.735*** | −1.689*** | −1.941** | −1.528*** |
| | | (0.487) | (0.483) | (0.540) | (0.538) | | (0.501) | (0.502) | (0.541) | (0.546) |
| Precipitation effect in poor countries | | | 0.013 | −0.147* | 0.014 | | | 0.261 | −0.433 | 0.102 |
| | | | (0.058) | (0.077) | (0.074) | | | (0.198) | (0.284) | (0.208) |

*Notes*: All specifications include country fixed effects, region×year fixed effects and poor × year fixed effects. Cluster robust SEs are reported in the parentheses. Temperature is in degrees Celsius and precipitation is in units of 100 mm. Estimates of aggregation weights in the FE-NLS estimation are not reported to save space. *** $P < 0.01$, ** $P < 0.05$, * $P < 0.1$.

TABLE 5

*FE-Naive estimations of models with lags*

| | No lags (1) | One lag (2) | Five lags (3) | 10 lags (4) | No lags (5) | One lag (6) | Five lags (7) | 10 lags (8) |
|---|---|---|---|---|---|---|---|---|
| Temperature × poor | −1.747*** | −1.857*** | −1.907*** | −1.801*** | −1.740*** | −1.866*** | −1.967*** | −1.924*** |
| | (0.487) | (0.576) | (0.609) | (0.682) | (0.483) | (0.571) | (0.602) | (0.675) |
| L1:Temperature × poor | | 0.713 | 0.974*** | 0.967** | | 0.793* | 1.046** | 1.035** |
| | | (0.444) | (0.423) | (0.460) | | (0.445) | (0.420) | (0.466) |
| L2:Temperature × poor | | | −0.205 | −0.330 | | | −0.252 | −0.384 |
| | | | (0.461) | (0.487) | | | (0.467) | (0.504) |
| L3:Temperature × poor | | | 0.131 | 0.189 | | | 0.223 | 0.222 |
| | | | (0.364) | (0.401) | | | (0.374) | (0.406) |
| Temperature × rich | 0.252 | 0.214 | 0.210 | 0.146 | 0.248 | 0.209 | 0.216 | 0.158 |
| | (0.307) | (0.289) | (0.291) | (0.331) | (0.306) | (0.290) | (0.293) | (0.336) |
| L1:Temperature × rich | | 0.195 | 0.259 | 0.293 | | 0.193 | 0.242 | 0.282 |
| | | (0.241) | (0.206) | (0.235) | | (0.241) | (0.204) | (0.234) |
| L2:Temperature × rich | | | 0.256 | 0.286 | | | 0.260 | 0.303 |
| | | | (0.226) | (0.247) | | | (0.229) | (0.254) |
| L3:Temperature × rich | | | 0.101 | 0.085 | | | 0.098 | 0.088 |
| | | | (0.197) | (0.220) | | | (0.196) | (0.223) |
| Precipitation included | No | No | No | No | Yes | Yes | Yes | Yes |
| Observations | 4924 | 4831 | 4444 | 3931 | 4924 | 4831 | 4444 | 3931 |
| $R^2$ | 0.145 | 0.144 | 0.146 | 0.134 | 0.145 | 0.146 | 0.151 | 0.141 |
| Sum of temperature coefficient in poor countries | −1.747*** | −1.444*** | -0.723 | −0.931 | −1.740*** | −1.073** | −0.983 | −0.773 |
| | (0.487) | (0.432) | (0.630) | (0.913) | (0.483) | (0.440) | (0.671) | (0.962) |
| Sum of temperature coefficient in rich countries | 0.252 | 0.409 | 0.819 | 0.974 | 0.248 | 0.402 | 0.320 | 0.914 |
| | (0.307) | (0.383) | (0.577) | (1.024) | (0.306) | (0.382) | (0.569) | (1.013) |

*Notes*: All specifications include country fixed effects, region × year fixed effects and poor × year fixed effects. Cluster robust SEs are reported in the parentheses. In models with more than three lags, given space constraints, the table reports only the first three lags. *** $P < 0.01$, ** $P < 0.05$, * $P < 0.1$.

TABLE 6

*FE-NLS estimations of models with lags*

| | No lags (1) | One lag (2) | Five lags (3) | 10 lags (4) | No lags (5) | One lag (6) | Five lags (7) | 10 lags (8) |
|---|---|---|---|---|---|---|---|---|
| Temperature × poor | −1.735*** | −1.773*** | −1.562*** | −1.376*** | −1.689*** | −1.735*** | −1.589*** | −1.427*** |
| | (*0.501*) | (*0.580*) | (*0.481*) | (*0.518*) | (*0.502*) | (*0.577*) | (*0.487*) | (*0.529*) |
| L1:Temperature × poor | | 0.900** | 1.126*** | 1.074** | | 0.883** | 1.064*** | 0.991** |
| | | (0.432) | (0.398) | (0.427) | | (0.428) | (0.405) | (0.412) |
| L2:Temperature × poor | | | −0.341 | −0.483 | | | −0.262 | −0.350 |
| | | | (0.404) | (0.414) | | | (0.413) | (0.423) |
| L3:Temperature × poor | | | 0.816** | 1.026** | | | 0.806* | 0.900* |
| | | | (0.414) | (0.455) | | | (0.427) | (0.457) |
| Temperature × rich | 0.310 | 0.250 | 0.368 | 0.458 | 0.293 | 0.259 | 0.385 | 0.490 |
| | (0.328) | (0.335) | (0.368) | (0.380) | (0.332) | (0.344) | (0.376) | (0.390) |
| L1:Temperature × rich | | 0.265 | 0.464*** | 0.511*** | | 0.279 | 0.479*** | 0.578*** |
| | | (0.214) | (0.169) | (0.170) | | (0.216) | (0.171) | (0.173) |
| L2:Temperature × rich | | | 0.063 | 0.251 | | | 0.080 | 0.303 |
| | | | (0.196) | (0.223) | | | (0.196) | (0.223) |
| L3:Temperature × rich | | | 0.032 | 0.109 | | | 0.030 | 0.105 |
| | | | (0.228) | (0.244) | | | (0.229) | (0.245) |
| Precipitation included | No | No | No | No | Yes | Yes | Yes | Yes |
| Observations | 4,924 | 4,831 | 4,444 | 3,931 | 4,924 | 4,831 | 4,444 | 3,931 |
| $R^2$ | 0.145 | 0.145 | 0.149 | 0.136 | 0.146 | 0.149 | 0.155 | 0.146 |
| Wald test (*P*-value) of equal weights | 0.307 | 0.377 | 0.000 | 0.000 | 0.344 | 0.342 | 0.000 | 0.232 |
| Sum of temperature coefficient in poor countries | −1.735*** | −0.873*** | -0.567 | −0.576 | −1.689*** | −0.852* | −0.551 | −0.470 |
| | (0.501) | (0.441) | (0.676) | (0.976) | (0.502) | (0.445) | (0.685) | (1.011) |
| Sum of temperature coefficient in rich countries | 0.310 | 0.515 | 0.267 | 0.851 | 0.293 | 0.538 | 0.262 | 0.867 |
| | (0.328) | (0.454) | (0.608) | (0.851) | (0.332) | (0.465) | (0.606) | (0.834) |

*Notes*: all specifications include country fixed effects, region × year fixed effects and poor × year fixed effects. Cluster robust standard errors are reported in the parentheses. In models with more than three lags, given space constraints, the table reports only the first three lags. *** $P < 0.01$, ** $P < 0.05$, * $P < 0.1$.

scheme may overestimate the negative effects of high temperature on the economic growth rates in poor countries.

Next we consider more flexible models by including lags of temperature to capture the dynamics of these effects. The model with 1 lag is given by equation (20). The FE-Naive estimates and the FE-NLS estimates of the distributed lag models are reported in Tables 5 and 6, respectively. The last two rows of each column in these tables present the growth effects of temperature for poor and rich countries, calculated by summing the coefficients associated with the respective temperature variable and its lags. These results suggest that the negative effect of temperature persists only in the short run. As more lags are included, the growth effect becomes statistically insignificant. As we can see from Table 6, estimates of the first individual lags are statistically significant, which suggests the the temperature shocks mainly work through the level effects instead of the growth effects. In the first year, $1^{\circ}$C increase in temperature leads to a reduction of $1.69-1.74$ percentage points of growth rate for poor countries. In the second year, the effect of the initial temperature shock on the growth declines to $0.85-0.87$ percentage points. This is in contrast with the results from Dell *et al.* (2012), in which the main driving force is the growth effect. Based on the empirical evidences from Table 6, we conclude that the temperature shocks have a negative effect on the growth of poor countries in the short run but not in the long run. The level effect eventually reverses itself once the shock disappears.

A noticeable difference between Tables 5 and 6 is that the FE-NLS estimators of the first lag of temperature within rich countries are significant at 1% level in specifications (2), (3), (7) and (8), while these lags are not significant in the FE-Naive estimation. This finding suggests that the temperature shock might have a positive level effect for the rich countries. This difference between the FE-NLS and the FE-Naive estimators can be attributed to the unequal aggregation weights. Based on the $P$ values reported in the third row from the bottom at Table 6, we reject the null hypothesis of equal aggregation weights in specifications (2), (3) and (7). The FE-Naive and the FE-NLS estimates are not very different for poor countries, although the magnitudes of the FE-NLS estimates are slightly smaller than the FE-Naive estimates in various specifications.

## VI. Conclusion

This paper investigates the properties of some common estimators developed for balanced panel models in the context of panel MIDAS models. In particular, we consider the estimation of panel data models in which the explanatory variables are sampled at a higher frequency than the dependent variable. We first show that the usual fixed effects estimator with an equal weighting scheme is generally inconsistent except for some special cases. Motivated by the time series MIDAS model, we propose a data-driven method to estimate the true aggregation weights. The linear regression parameters together with the unknown weights can be estimated by the fixed effects nonlinear least squares method.

We further extend the Mundlak device and the Chamberlain's projection approach to the panel MIDAS model. In single-frequency panel data models, it is well known that the use of Mundlak device or the Chamberlain's projection approach all lead to the fixed effects estimator. In the panel MIDAS model, we can either project the individual fixed effects onto the aggregated explanatory variables or the high-frequency explanatory variables, which gives rise to two different versions of the Mundlak device and the Chamberlain's projection. We therefore propose four different estimators: low-frequency Mundlak estimator, high-frequency estimator, low-frequency Chamberlain estimator and the high-frequency Chamberlain estimator. We show that, when the aggregation weights are known, the low-frequency Mundlak estimator, the low-frequency and high-frequency Chamberlain estimators are numerically identical to the fixed effects nonlinear least squares estimator. When the true weights are unknown, only the high-frequency Chamberlain estimator is equivalent to the fixed effects nonlinear least squares estimator. We further demonstrate that the same equivalence results extend to the case that parametric weighting functions are used in the panel MIDAS model. The proposed estimators perform well in various simulation designs.

We revisit the paper by Dell *et al.* (2012) to examine the effects of temperature on the economic growth rates with our new method. Our results indicate that the high temperature reduces the economic growth rates of the poor countries in the short run. The empirical evidences further suggest that the mechanism of how the temperature shocks affect the growth is through the level effects instead of the growth effects.

## Appendix A

*Calculation of the scores and the Hessian matrix*. The $i$th observation contributes to the objective function as $q(W_i;\theta) = \frac{1}{2}\sum_{t=1}^{T}[\ddot{y}_{it} - \ddot{x}_{it}(\alpha)\beta]^2$. The score can be derived as

$$s_i(\theta) = \nabla_\theta q_i(\theta) = -\sum_{t=1}^{T}\nabla_\theta[\ddot{x}_{it}(\alpha)\beta]'\,[\ddot{y}_{it} - \ddot{x}_{it}(\alpha)\beta]\,. \tag{A1}$$

Note that, for $t = 1, \ldots, T$, we can rewrite $\ddot{x}_{it}(\alpha)\beta = \sum_{k=1}^{p}\ddot{x}_{k,it}(\alpha_k)\beta_k = \sum_{k=1}^{p}\sum_{j=1}^{m}\ddot{x}^{(t)}_{k,ij/m}a_{jk}\beta_k$. Recall that we internalize the constraint that $\sum_{j=1}^{m}a_{jk} = 1$ by setting $a_{mk} = 1 - \sum_{j=1}^{m-1}a_{jk}$ for $k = 1, \ldots, p$. The total number of unrestricted weighting parameter is $(m-1)p$, so $\dim(\theta) = \dim(\alpha) + \dim(\beta) = (m-1)p + p = mp$. Then, for $t = 1, \ldots, T$,

$$\begin{aligned}\frac{\partial\,[\ddot{x}_{it}(\alpha)\beta]}{\partial a_{jk}} &= \frac{\partial\left[\sum_{k=1}^{p}\sum_{j=1}^{m}\ddot{x}^{(t)}_{k,ij/m}a_{jk}\beta_k\right]}{\partial a_{jk}}\\ &= \frac{\partial\left[\sum_{k=1}^{p}\left(\sum_{j=1}^{m-1}\ddot{x}^{(t)}_{k,ij/m}a_{jk}\beta_k + \ddot{x}^{(t)}_{k,im/m}a_{mk}\beta_k\right)\right]}{\partial a_{jk}}\\ &= \left(\ddot{x}^{(t)}_{k,ij/m} - \ddot{x}^{(t)}_{k,im/m}\right)\beta_k,\ j = 1, \ldots, m-1,\ k = 1, \ldots, p. \end{aligned} \tag{A2}$$

$$\frac{\partial\,[\ddot{x}_{it}(\alpha)\beta]}{\partial\beta_k} = \frac{\partial\left[\sum_{k=1}^{p}\sum_{j=1}^{m}\ddot{x}^{(t)}_{k,ij/m}a_{jk}\beta_k\right]}{\partial\beta_k} = \sum_{j=1}^{m}\ddot{x}^{(t)}_{k,ij/m}\alpha_{jk} = \ddot{x}_{k,it}(\alpha_k),\ k = 1, \ldots, p. \tag{A3}$$

It follows that the transpose of the score can be explicitly written as

$$
\begin{aligned}
s_i(\theta)' &= -\sum_{t=1}^{T} \nabla_\theta \left[ \ddot{x}_{it}(\alpha)\beta \right] \ddot{u}_{it} \\
&= -\sum_{t=1}^{T} \left[ \frac{\partial[\ddot{x}_{it}(\alpha)\beta]}{\partial a_{11}}, \quad \ldots, \quad \frac{\partial[\ddot{x}_{it}(\alpha)\beta]}{\partial a_{m-1,p}}, \quad \frac{\partial[\ddot{x}_{it}(\alpha)\beta]}{\partial \beta_1}, \quad \ldots, \quad \frac{\partial[\ddot{x}_{it}(\alpha)\beta]}{\partial \beta_p} \right] \ddot{u}_{it} \\
&= -\sum_{t=1}^{T} \left[ \left( \ddot{x}^{(t)}_{1,i1/m} - \ddot{x}^{(t)}_{1,im/m} \right) \beta_1, \quad \ldots, \quad \left( \ddot{x}^{(t)}_{p,i(m-1)/m} - \ddot{x}^{(t)}_{p,im/m} \right) \beta_p, \right. \\
&\qquad \left. \times\, \ddot{x}_{1,it}(\alpha_1), \quad \ldots, \quad \ddot{x}_{p,it}(\alpha_p) \right] \ddot{u}_{it}. \qquad \text{(A4)}
\end{aligned}
$$

Next we calculate the Hessian matrix associated with the objective function as

$$
H_i(\theta) = \nabla_\theta s_i(\theta) = -\sum_{t=1}^{T} \nabla^2_\theta \left[ \ddot{x}_{it}(\alpha)\beta \right] \left[ \ddot{y}_{it} - \ddot{x}_{it}(\alpha)\beta \right] + \sum_{t=1}^{T} \nabla_\theta [\ddot{x}_{it}(\alpha)\beta]' \nabla_\theta \left[ \ddot{x}_{it}(\alpha)\beta \right]. \qquad \text{(A5)}
$$

The first term in the above equation has expectation 0 when evaluated at the true parameter value. Hence, the expectation of the Hessian matrix can be expressed as

$$
B^* \equiv E\left[ H_i(\theta^*) \right] = \sum_{t=1}^{T} B^*_t = \sum_{t=1}^{T} E\left\{ \nabla_\theta \left[ \ddot{x}_{it}(\alpha^*)\beta^* \right]' \nabla_\theta \left[ \ddot{x}_{it}(\alpha^*)\beta^* \right] \right\}. \qquad \text{(A6)}
$$

Based on the score in equation (A4), we can see that a typical element of $B^*_t$ is $E\left[ \ldots x^{(t)}_{k,ij/m} \ldots x^{(t)}_{l,ij/m} \right] \beta_k \beta_l$ or $E\left[ \ldots x^{(t)}_{k,ij/m} \ddot{x}_{l,it}(\alpha_l) \right] \beta_k$ or $E\left[ \ddot{x}_{k,it}(\alpha_k) \ddot{x}_{l,it}(\alpha_l) \right]$, where $\ldots x^{(t)}_{k,ij/m} \equiv \ddot{x}^{(t)}_{k,ij/m} - \ddot{x}^{(t)}_{k,im/m}$.

## Appendix B: Parametric weighting functions

As pointed out by Ghysels *et al.* (2007), even with a moderate number of $m$ and $p$, the number of unrestricted parameters $\alpha$ can be very large. A distinguishing feature of MIDAS model is to use a suitable parametrization with a low-dimension parameter vector $\xi$ to avoid the problem of parameter proliferation. While there are a variety of parametrization designs, one popular choice is the exponential Almon lag polynomial. Take the $k$th covariate as an example, the weight assigned to the $j$th high frequency term is given by

$$
a_{jk} = \frac{\exp(\xi_{k1} j + \xi_{k2} j^2 + \cdots + \xi_{kL} j^L)}{\sum_{s=1}^{m} \exp(\xi_{k1} s + \xi_{k2} s^2 + \cdots + \xi_{kL} s^L)}, \quad j = 1, \ldots, m. \qquad \text{(B1)}
$$

Ghysels *et al.* (2006) use the above function with two parameters ($L = 2$) to show that the Almon lag polynomial can take various forms, including equal weights, slow decaying weights, rapidly declining weights, hump shape weights and so on. A different weighting scheme is referred to as the beta lag function, which is parametrized with two parameters.

See Ghysels *et al.* (2007) for more details. With proper parametrization of the weights, the number of parameters to be estimated can be drastically reduced. Since the weights $\alpha$ are now determined by a low-dimension vector $\xi = (\xi_1', ..., \xi_p')'$, the FE-NLS estimator is now obtained by the following procedure

$$(\hat{\xi}_{\text{FE-NLS}}, \hat{\beta}_{\text{FE-NLS}}) = \underset{\xi,\beta}{\text{argmin}} \sum_{i=1}^{N} (y_i - X_i(\xi)\beta)' Q_T (y_i - X_i(\xi)\beta). \tag{B2}$$

Compared with equation (5), the above objective function has a smaller number of parameters to be estimated, especially when $m$ is relatively large. Provided with suitable regularity conditions, the FE-NLS estimator can be shown to be consistent and asymptotically normally distributed. Obviously the scores and the Hessian matrix associated with equation (B2) are different as we are now using flexible parametric functions to model the aggregation weights.

## Appendix C: Proof of Proposition 1

For the simplicity of notation we demean all variables by their sample averages so that the intercept can be excluded from all regressions. We first investigate the simple case that the aggregating weights are known.

(a) the FE-NLS estimator: $\hat{\beta}_{\text{FE-NLS}} = \left(\sum_{i=1}^{N} \sum_{t=1}^{T} \ddot{x}_{it}(\alpha^*)' \ddot{x}_{it}(\alpha^*)\right)^{-1} \left(\sum_{i=1}^{N} \sum_{t=1}^{T} \ddot{x}_{it}(\alpha^*)' \ddot{y}_{it}\right)$.

(b) the low-frequency Mundlak regression estimator: using the partition out theorem, we can first regress $x_{it}(\alpha^*)$ on $\bar{x}_i(\alpha^*)$ to obtain the residuals $\ddot{x}_{it}(\alpha^*)$ and then regress $y_{it}$ on $\ddot{x}_{it}(\alpha^*)$ to obtain the low-frequency Mundlak estimator $\hat{\beta}_{\text{LF-Mundlak}} = \left(\sum_{i=1}^{N} \sum_{t=1}^{T} \ddot{x}_{it}(\alpha^*)' \ddot{x}_{it}(\alpha^*)\right)^{-1} \left(\sum_{i=1}^{N} \sum_{t=1}^{T} \ddot{x}_{it}(\alpha^*)' y_{it}\right) = \hat{\beta}_{\text{FE-NLS}}$.

(c) the low-frequency Chamberlain regression estimator: applying the Frisch–Waugh theorem, the residuals obtained from regressing $x_{it}(\alpha^*)$ on $x_i(\alpha^*)$ is also $\ddot{x}_{it}(\alpha^*)$ so the low-frequency Chamberlain estimator is also identical to the FE-NLS estimator.

(d) the high-frequency Chamberlain regression estimator: Since $x_{it}(\alpha^*) = x_i^{(t)} A(\alpha^*)$ and $x_i^{(t)} = x_i S_t$, where $S_t$ is a selection matrix that selects $x_i^{(t)}$ from $x_i$, it follows that $x_{it}(\alpha^*) = x_i S_t A(\alpha^*)$ is also a linear combination of $x_i$. If we regress $x_{it}(\alpha^*)$ on $x_i$, the residuals are given by

$$\begin{aligned} r_{it}(\alpha^*) &= x_{it}(\alpha^*) - x_i \left[\sum_{i=1}^{N}\sum_{t=1}^{T} x_i' x_i\right]^{-1} \left[\sum_{i=1}^{N}\sum_{t=1}^{T} x_i' x_{it}(\alpha^*)\right] \\ &= x_{it} - x_i \left[T \sum_{i=1}^{N} x_i' x_i\right]^{-1} \left[\sum_{i=1}^{N}\sum_{t=1}^{T} x_i' x_i S_t A(\alpha^*)\right] \\ &= x_{it} - T^{-1} \sum_{t=1}^{T} x_i S_t A(\alpha^*) = x_{it}(\alpha^*) - \bar{x}_i(\alpha^*) = \ddot{x}_{it}(\alpha^*). \end{aligned}$$

It follows that the HF-Chamberlain estimator is also identical to the FE-NLS estimator.

A more realistic case is that the true aggregating weights are unknown and they are estimated along with the slope parameters.
(a1) The objective function for the FE-NLS estimator is

$$\min_{\alpha,\beta} \sum_{i=1}^{N} \sum_{t=1}^{T} (\ddot{y}_{it} - \ddot{x}_{it}(\alpha)\beta)^2. \tag{C1}$$

Define $\ddot{x}^{(t)}_{k,ij/m} = x^{(t)}_{k,ij/m} - T^{-1} \sum_{s=1}^{T} x^{(s)}_{k,ij/m}$. The associated first-order conditions, with respect to $\beta$ and $\alpha_{jk}$, are given by the following

$$\sum_{i=1}^{N} \sum_{t=1}^{T} \ddot{x}_{it}(\hat{\alpha})'(\ddot{y}_{it} - \ddot{x}_{it}(\hat{\alpha})\hat{\beta}) = 0. \tag{C2}$$

$$\sum_{i=1}^{N} \sum_{t=1}^{T} \ddot{x}^{(t)}_{k,ij/m} \hat{\beta}_k (\ddot{y}_{it} - \ddot{x}_{it}(\hat{\alpha})\hat{\beta}) = 0. \tag{C3}$$

(d1) The objective function for the high-frequency Chamberlain regression estimator is

$$\min_{\alpha,\beta,\lambda^H} \sum_{i=1}^{N} \sum_{t=1}^{T} (y_{it} - x_{it}(\alpha)\beta - x_i\lambda^H)^2. \tag{C4}$$

In this case the first-order conditions, with respect to $\beta$, $\lambda^H$ and $\alpha_{jk}$, are given by the following

$$\sum_{i=1}^{N} \sum_{t=1}^{T} x_{it}(\hat{\alpha})'\left(y_{it} - x_{it}(\hat{\alpha})\hat{\beta} - x_i\hat{\lambda}^H\right) = 0. \tag{C5}$$

$$\sum_{i=1}^{N} x_i'\left(\bar{y}_i - \bar{x}_i(\hat{\alpha})\hat{\beta} - x_i\hat{\lambda}^H\right) = 0. \tag{C6}$$

$$\sum_{i=1}^{N} \sum_{t=1}^{T} x^{(t)}_{k,ij/m} \hat{\beta}_k \left(y_{it} - x_{it}(\hat{\alpha})\hat{\beta} - x_i\hat{\lambda}^H\right) = 0. \tag{C7}$$

Since $\bar{x}_i(\alpha) = x_i A$, it is straightforward to show that $\hat{\lambda}^H = \check{\beta} - A\hat{\beta}$, where $\check{\beta} = \left[\sum_{i=1}^{N} x_i' x_i\right]^{-1} \left[\sum_{i=1}^{N} x_i' \bar{y}_i\right]$. Making use of equation (C5), it then follows that the $\hat{\beta} = \left[\sum_{i=1}^{N} \sum_{t=1}^{T} \ddot{x}_{it}(\hat{\alpha})' \ddot{x}_{it}(\hat{\alpha})\right]^{-1} \left[\sum_{i=1}^{N} \sum_{t=1}^{T} \ddot{x}_{it}(\hat{\alpha})' \ddot{y}_{it}\right]$. Equation (C7) can be rewritten as

$$\sum_{i=1}^{N} \sum_{t=1}^{T} x^{(t)}_{k,ij/m} \hat{\beta}_k (y_{it} - x_i\check{\beta} - \ddot{x}_{it}(\hat{\alpha})\hat{\beta}) = 0$$

$$\sum_{i=1}^{N}\sum_{t=1}^{T}x_{k,ij/m}^{(t)}\hat{\beta}_k(\bar{y}_i - x_i\check{\beta} + \ddot{y}_{it} - \ddot{x}_{it}(\hat{\alpha})\hat{\beta}) = 0$$

$$\sum_{i=1}^{N}\sum_{t=1}^{T}\ddot{x}_{k,ij/m}^{(t)}\hat{\beta}_k(\ddot{y}_{it} - \ddot{x}_{it}(\hat{\alpha})\hat{\beta}) = 0, \tag{C8}$$

where (C8) follows from the first-order condition (C6) that $\sum_{i=1}^{N}\sum_{t=1}^{T}x_{k,ij/m}^{(t)}\hat{\beta}_k(\bar{y}_i - x_i\check{\beta}) = 0$ (recall that $x_i$ is a vector of all high-frequency variables $x_{ij/m}^{(t)}$). Thus we have shown that the first-order conditions associated with the high-frequency Chamberlain estimator is exactly the same as the F.O.C of the FE-NLS estimator.

The low-frequency and high-frequency Mundlak estimators, as well as the low-frequency Chamberlain estimator can all be viewed as some restricted estimators from the high-frequency Chamberlain regression. For a concrete example, consider a panel MIDAS model with a single covariate for $T = 2$ and $m = 3$. Then the high-frequency Chamberlain regression equation can be written as

$$y_{it} = \alpha_1 x_{i1}^{(t)}\beta + \alpha_2 x_{i2}^{(t)}\beta + \alpha_3 x_{i3}^{(t)}\beta + x_{i1}^{(1)}\lambda_1^H + x_{i2}^{(1)}\lambda_2^H + x_{i3}^{(1)}\lambda_3^H + x_{i1}^{(2)}\lambda_4^H + x_{i2}^{(2)}\lambda_5^H + x_{i3}^{(2)}\lambda_6^H + a_{it}, \quad t = 1, 2. \tag{C9}$$

we can simply run a regression of $y_{it}$ on $(x_i^{(t)}, x_i)$ to consistently estimate all the parameters with the restriction that the sum of $a$ is one. Define $\bar{x}_{ij} = T^{-1}\sum_{t=1}x_{ij}^{(t)}$. The low-frequency Mundlak estimator can be viewed as a restricted estimator from (C9).

$$y_{it} = a_1 x_{i1}^{(t)}\beta + a_2 x_{i2}^{(t)}\beta + a_3 x_{i3}^{(t)}\beta + a_1\bar{x}_{i1}\gamma^L + a_2\bar{x}_{i2}\gamma^L + a_3\bar{x}_{i3}\gamma^L + e_{it}, \quad t = 1, 2. \tag{C10}$$

An important restrictions imposed on (C10) compared to the unrestricted equation (C9) is that the ratios of coefficients on $(x_{i1}^{(t)}, x_{i2}^{(t)}, x_{i3}^{(t)})$, $(x_{i1}^{(1)}, x_{i2}^{(1)}, x_{i3}^{(1)})$ and $(x_{i1}^{(2)}, x_{i2}^{(2)}, x_{i3}^{(2)})$ from equation (C9) are all equal to $a_1 : a_2 : a_3$. When the unrestricted parameter set $(\lambda_1^H, \lambda_2^H, \lambda_3^H)$, or similarly $(\lambda_4^H, \lambda_5^H, \lambda_6^H)$, does not satisfy this ratio restriction, the low-frequency Mundlak estimator is not the same as the high-frequency estimator, thus is different from the FE-NLS estimator. Following the same line, we can conclude that the high-frequency Mundlak estimator and the low-frequency Chamberlain estimator are also different from the FE-NLS estimator. **Q.E.D**

*Proof of Proposition 3.2.* When $\xi$ is known to us, the first part of the proposition is exactly the same as the part that $\alpha$ is known in Proposition 1. The proof is thus omitted. We focus on the proof of the second part only. When the parameter vector $\xi$ associated with the parametric aggregating functions are unknown and they are estimated along with the slope parameters. □

(a2) The objective function for the FE-NLS estimator is

$$\min_{\xi,\beta}\sum_{i=1}^{N}\sum_{t=1}^{T}(\ddot{y}_{it} - \ddot{x}_{it}(\xi)\beta)^2. \tag{C11}$$

The associated first-order conditions, with respect to $\beta$ and $\xi_k$, are given by the following

$$\sum_{i=1}^{N}\sum_{t=1}^{T}\ddot{x}_{it}(\hat{\xi})'(\ddot{y}_{it}-\ddot{x}_{it}(\hat{\xi})\hat{\beta})=0. \tag{C12}$$

$$\sum_{i=1}^{N}\sum_{t=1}^{T}\left[\sum_{j=1}^{m}\frac{\partial f_{jk}}{\partial \xi_k}\ddot{x}^{(t)}_{k,ij/m}\hat{\beta}_k\right](\ddot{y}_{it}-\ddot{x}_{it}(\hat{\alpha})\hat{\beta})=0. \tag{C13}$$

(d2) The objective function for the high-frequency Chamberlain regression estimator is

$$\min_{\xi,\beta,\lambda^H}\sum_{i=1}^{N}\sum_{t=1}^{T}(y_{it}-x_{it}(\xi)\beta-x_i\lambda^H)^2. \tag{C14}$$

In this case the first-order conditions, with respect to $\beta$, $\lambda^H$ and $\xi_k$, are given by the following

$$\sum_{i=1}^{N}\sum_{t=1}^{T}x_{it}(\hat{\alpha})'\left(y_{it}-x_{it}(\hat{\alpha})\hat{\beta}-x_i\hat{\lambda}^H\right)=0. \tag{C15}$$

$$\sum_{i=1}^{N}x_i'\left(\bar{y}_i-\bar{x}_i(\hat{\alpha})\hat{\beta}-x_i\hat{\lambda}^H\right)=0. \tag{C16}$$

$$\sum_{i=1}^{N}\sum_{t=1}^{T}\left[\sum_{j=1}^{m}\frac{\partial f_{jk}}{\partial \xi_k}x^{(t)}_{k,ij/m}\hat{\beta}_k\right]\left(y_{it}-x_{it}(\hat{\alpha})\hat{\beta}-x_i\hat{\lambda}^H\right)=0. \tag{C17}$$

Since $\bar{x}_i(\xi)=x_i\tilde{A}$, it is straightforward to show that $\hat{\lambda}^H=\check{\beta}-\tilde{A}\hat{\beta}$, where $\check{\beta}=\left[\sum_{i=1}^{N}x_i'x_i\right]^{-1}\left[\sum_{i=1}^{N}x_i'\bar{y}_i\right]$. Making use of equation (C15), it can be shown that $\hat{\beta}=\left[\sum_{i=1}^{N}\sum_{t=1}^{T}\ddot{x}_{it}(\hat{\xi})'\ddot{x}_{it}(\hat{\xi})\right]^{-1}\left[\sum_{i=1}^{N}\sum_{t=1}^{T}\ddot{x}_{it}(\hat{\xi})'\ddot{y}_{it}\right]$. Equation (C17) can be rewritten as

$$\sum_{i=1}^{N}\sum_{t=1}^{T}\left[\sum_{j=1}^{m}\frac{\partial f_{jk}}{\partial \xi_k}x^{(t)}_{k,ij/m}\hat{\beta}_k\right](y_{it}-x_i\check{\beta}-\ddot{x}_{it}(\hat{\xi})\hat{\beta})=0$$

$$\sum_{i=1}^{N}\sum_{t=1}^{T}\left[\sum_{j=1}^{m}\frac{\partial f_{jk}}{\partial \xi_k}x^{(t)}_{k,ij/m}\hat{\beta}_k\right](\bar{y}_i-x_i\check{\beta}+\ddot{y}_{it}-\ddot{x}_{it}(\hat{\xi})\hat{\beta})=0$$

$$\sum_{i=1}^{N}\sum_{t=1}^{T}(\ddot{y}_{it}-\ddot{x}_{it}(\hat{\xi})\hat{\beta})=0, \tag{C18}$$

where (C18) follows from the first-order condition (C16) that $\sum_{i=1}^{N}\sum_{t=1}^{T} x_{k,ij/m}^{(t)}\hat{\beta}_k(\overline{y}_i - x_i\check{\beta}) = 0$ (recall that $x_i$ is a vector of all high-frequency variables $x_{ij/m}^{(t)}$). Thus we have shown that the F.O.C associated with the high-frequency Chamberlain estimator is exactly the same as the F.O.C of the FE-NLS estimator.

The reason that low-frequency and high-frequency Mundlak estimators, as well as the low-frequency Chamberlain estimator, are no longer identical to the FE-NLS estimator is the same as the one given in the proof of proposition 1.

*Final Manuscript Received: February 2022*

## References


Abrevaya, J. (2013). 'The projection approach for unbalanced panel data', *The Econometrics Journal*, Vol. 16, pp. 161–178.

Abrevaya, J. (2019). 'Missing dependent variables in fixed-effects models', *Journal of Econometrics*, Vol. 211, pp. 151–165.

Anderson, T. and Hsiao, C. (1982). 'Formulation and estimation of dynamic models using panel data', *Journal of Econometrics*, Vol. 18, pp. 47–82.

Andreou, E., Ghysels, E. and Kourtellos, A. (2010). 'Regression models with mixed sampling frequencies', *Journal of Econometrics*, Vol. 158, pp. 246–261.

Arellano, M. A. and Bond, S. R. (1991). 'Some tests of specification for panel data: Monte carlo evidence and an application to employment equations', *The Review of Economic Studies*, Vol. 58, pp. 277–297.

Bond, S., Leblebicioğlu, A. and Schiantarelli, F. (2010). 'Capital accumulation and growth: a new look at the empirical evidence', *Journal of Applied Econometrics*, Vol. 25, pp. 1073–1099.

Chamberlain, G. (1982). 'Multivariate regression models for panel data', *Journal of Econometrics*, Vol. 18, pp. 5–46.

Davies, R. B. (1977). 'Hypothesis testing when a nuisance parameter is present only under the alternative', *Biometrika*, Vol. 64, pp. 247–254.

Davies, R. B. (1987). 'Hypothesis testing when a nuisance parameter is present only under the alternatives', *Biometrika*, Vol. 74, pp. 33–43.

Dell, M., Jones, B. F. and Olken, B. A. (2012). 'Temperature shocks and economic growth: Evidence from the last half century', *American Economic Journal: Macroeconomics*, Vol. 4, pp. 66–95.

Foroni, C. and Marcellino, M. (2013). *A Survey of Econometric Methods for Mixed-Frequency Data*. Working Paper No. 2013/06, Norges Bank.

Ghysels, E. and Wright, J. H. (2009). 'Forecasting professional forecasters', *Journal of Business & Economic Statistics*, Vol. 27, pp. 504–516.

Ghysels, E., Santa-Clara, P. and Valkanov, R. (2006). 'Predicting volatility: getting the most out of return data sampled at different frequencies', *Journal of Econometrics*, Vol. 131, pp. 59–95.

Ghysels, E., Sinko, A. and Valkanov, R. (2007). 'Midas regressions: further results and new directions', *Econometric Reviews*, Vol. 26, pp. 53–90.

Granger, C. (1987). 'Implications of aggregation in the multiple regression model', *Econometric Theory*, Vol. 3, pp. 208–222.

Hsiao, C. (1979). 'Linear regression using both temporally aggregated and temporally disaggregated data', *Journal of Econometrics*, Vol. 10, pp. 243–252.

Khalaf, L., Kichian, M., Saunders, C. J. and Voia, M. (2021). 'Dynamic panels with midas covariates: nonlinearity, estimation and fit', *Journal of Econometrics*, Vol. 220, pp. 589–605.

Mundlak, Y. (1978). 'On the pooling of time series and cross section data', *Econometrica*, Vol. 46, pp. 69–85.

Muris, C. (2020). 'Efficient GMM estimation with incomplete data', *The Review of Economics and Statistics*, Vol. 102, pp. 518–530.

Newey, W. K. and McFadden, D. (1994). ‘Large sample estimation and hypothesis testing’, in: Engle, R. and McFadden, D. (eds), *Handbook of Econometrics*, Vol. 4, Chapter 36. North-Holland: Elsevier, pp. 2111–2245.

Phillips, P. C. B. (1972). ‘The structural estimation of a stochastic differential equation system’, *Econometrica*, Vol. 40, pp. 1021–1041.

Phillips, P. C. B. (1974). ‘The estimation of some continuous time models’, *Econometrica*, Vol. 42, pp. 803–823.

Sims, C. A. (1971). ‘Discrete approximations to continuous time distributed lags in econometrics’, *Econometrica*, Vol. 39, pp. 545–563.

Wansbeek, T. and Kapteyn, A. (1989). ‘Estimation of the error-components model with incomplete panels’, *Journal of Econometrics*, Vol. 41, pp. 341–361.

Wooldridge, J. M. (2010). *Econometric Analysis of Cross Section and Panel Data* Vol. 1, 2nd ed., The MIT Press, Cambridge.

Wooldridge, J. M. (2019). ‘Correlated random effects models with unbalanced panels’, *Journal of Econometrics*, Vol. 211, pp. 137–150.