Identifying Latent Structures in Panel Data*

Liangjun Su[a], Zhentao Shi[b], and Peter C. B. Phillips[c]

[a] *School of Economics, Singapore Management University*

[b] *Department of Economics, Chinese University of Hong Kong*

[c] *Yale University, University of Auckland*

University of Southampton & Singapore Management University

December 29, 2015

This paper provides a novel mechanism for identifying and estimating latent group structures in panel data using penalized techniques. We consider both linear and nonlinear models where the regression coefficients are heterogeneous across groups but homogeneous within a group and the group membership is unknown. Two approaches are considered – penalized profile likelihood (PPL) estimation for the general nonlinear models without endogenous regressors, and penalized GMM (PGMM) estimation for linear models with endogeneity. In both cases we develop a new variant of Lasso called classifier-Lasso (C-Lasso) that serves to shrink individual coefficients to the unknown group-specific coefficients. C-Lasso achieves simultaneous classification and consistent estimation in a single step and the classification exhibits the desirable property of uniform consistency. For PPL estimation C-Lasso also achieves the oracle property so that group-specific parameter estimators are asymptotically equivalent to infeasible estimators that use individual group identity information. For PGMM estimation the oracle property of C-Lasso is preserved in some special cases. Simulations demonstrate good finite-sample performance of the approach both in classification and estimation. Empirical applications to both linear and nonlinear models are presented.

fi C33, C36, C38, C51

Classification; Cluster analysis; Dynamic panel; Group Lasso; High dimensionality; Nonlinear panel; Oracle property; Panel structure model; Parameter heterogeneity; Penalized least squares; Penalized GMM; Penalized profile likelihood

---

*The authors thank the Co-editor Elie Tamer and three anonymous referees for many constructive comments on the previous version of the paper. They also thank Stéphane Bonhomme, Xiaohong Chen, Cheng Hsiao, Joon Park, and Yixiao Sun for discussions on the subject matter and comments on the paper. Su acknowledges support from the Singapore Ministry of Education for Academic Research Fund (AcRF) under the Tier-2 grant number MOE2012-T2-2-021. Phillips acknowledges NSF support under Grant Nos. SES-0956687 and SES-1285258. Address Correspondence to: Liangjun Su, School of Economics, Singapore Management University, 90 Stamford Road, Singapore 178903; E-mail: ljsu@smu.edu.sg, Phone: +65 6828 0386.

Panel data are widely used in empirical analysis in many disciplines across the social and medical sciences. Such data usually cover individual units sampled from different backgrounds and with different individual characteristics so that an abiding feature of the data is its heterogeneity, much of which is simply unobserved. Neglecting latent heterogeneity in the data can lead to many difficulties, including inconsistent estimation and misleading inference, as is well explained in the literature (e.g., Hsiao 2014, ch. 6). It is therefore widely acknowledged that an important feature of good empirical modeling is to control for heterogeneity in the data as well as for potential heterogeneity in the response mechanisms that figure within the model. Since heterogeneity is a latent feature of the data and its extent is unknown a priori, respecting the potential influence of heterogeneity on model specification is a serious challenge in empirical research. Even in the simplest linear panel data models the challenge is manifest and clearly stated: do we allow for heterogeneous slope coefficients in regression as well as heterogeneous error variances?

While it may be clearly stated, this challenge to the empirical researcher is by no means easily addressed. While allowing for cross-sectional slope heterogeneity in regression may help to avert misspecification bias, it also sacrifices the power of cross section averaging in the estimation of response patterns that may be common across individuals, or more subtly, certain groups of individuals in the panel. In the absence of prior information on such grouping and with data where every new individual to the panel may bring new idiosyncratic elements to be explained, the challenge is demanding and almost universally relevant.

Traditional panel data models frequently deal with this challenge by avoidance. Complete slope homogeneity is assumed for certain specified common parameters in the panel. Under this assumption, the regression parameters are the same across individuals and unobserved heterogeneity is modeled through individual-specific effects which typically enter the model additively. This approach is an exemplar of a convenient assumption that facilitates estimation and inference. Nevertheless, this assumption has been frequently questioned and rejected in empirical studies; see Hsiao and Tahmiscioglu (1997), Lee, Pesaran, and Smith (1997), Durlauf, Kourtellos, and Minkin (2001), Phillips and Sul (2007a), Browning and Carro (2007, 2010, 2014), and Su and Chen (2013), among others.

Despite general agreement that slope heterogeneity is endemic in empirical work with panels, few methods are available to allow for heterogeneity in the slopes when the extent of the heterogeneity is unknown. Some researchers assume complete slope heterogeneity where regression coefficients are completely different for different individuals; see the survey by Baltagi, Bresson, and Pirotte (2008) and Hsiao and Pesaran (2008). Others consider panel structure models where individuals belong to a number of homogeneous groups within a broadly heterogeneous population, and the regression parameters are the same within each group but differ across groups. Two essential questions remain: how to determine the unknown number of groups (dubbed convergence clubs in the economic growth literature); and how to identify the membership of each individual. These are longstanding questions of statistical classification in panel data. No completely satisfactory solution has yet been found, although various approaches have been adopted in empirical research. For instance, Bester and Hansen (2016) consider a panel structure model where individuals are grouped according to some external classification, geographic location, or observable explanatory variables; Ando and Bai (2014) consider a multifactor asset-pricing model with group-specific pervasive factors where the group membership is known. Here the group structure is assumed to be *completely known* to the researcher, an approach that is common in practical work because of its convenience. In spite of its convenience, this approach to panel inference is inevitably misleading when the number of groups and individual identities are incorrectly classified.

Several approaches have been proposed to determine an *unknown* group structure in modeling unobserved slope heterogeneity in panels. The first approach applies finite mixture models. For example, Sun (2005) considers a *parametric* finite mixture linear panel data model, and Kasahara and Shimotsu (2009) and Browning and Carro (2011) study identification in discrete choice panel data models for a fixed number of groups using *nonparametric* discrete mixture distributions. The second approach is based on the K-means algorithm in statistical cluster analysis. Lin and Ng (2012) and Sarafidis and Weber (2015) consider linear panel data models where the slope coefficients have latent group structure. They modify the K-means algorithm to estimate the models but do not provide any inference theory. Bonhomme and Manresa (2015, BM hereafter) consider a linear panel data model where the additive fixed effects have group structure and apply the K-means algorithm to estimate the model and study its asymptotic properties. Ando and Bai (2015) extend BM's approach to allow for group structure among the interactive fixed effects. In addition, Phillips and Sul (2007a) develop an algorithm for determining group clusters that relies on the estimation of evaporating trend functions to determine convergence clusters. Hahn and Moon (2010) argue that the group structure has sound foundations in game theory or macroeconomic models where multiplicity of Nash equilibria is expected and they consider nonlinear panel data models where the parameter of interest is common to individuals whereas the fixed effects have finite support.

The present paper proposes a new method for econometric estimation and inference in panel models when the regression parameters are heterogenous across groups, individual group membership is unknown, and classification is to be determined empirically. It is an automated data-determined procedure and does not require the specification of any modeling mechanism for the unknown group structure. The methods proposed here have several novel aspects in relation to earlier research and they contribute to both the Lasso and econometric classification literatures in various ways, which we outline in the following paragraphs.

First, our approach is motivated by a key advantage of Lasso technology in coping with parameter sparsity. This advantage is particularly useful when the set of unknown parameters is potentially large but may also embody certain *sparse* features. In a typical panel structure model, the *effective* number of unknown regression parameters $\{ \ _i \quad = 1 \quad \}$ is not of order ( ) as it would be if these parameters were all incidental, but rather of some order ($_0$) where $_0$ denotes the number of unknown groups within which the parameters are homogeneous. Hence, in many empirical applications the set of unknown parameters in a panel structure model surely exhibits the desirable sparsity feature, making the use of Lasso technology highly appealing.

Second, the procedures developed in the present paper contribute to the fused Lasso literature in which sparsity arises because some parameters take the same value. The fused Lasso was proposed by Tibshirani, Saunders, Rosset, Zhu, and Knight (2005) and was designed for problems with features that can be ordered

ec5 o esti1ns

likelihood objective function and when multiple penalty terms are used they enter the objective function *additively*. To achieve simultaneous group classification and estimation in a single step our variant of Lasso involves *additive* penalty terms, each of which takes a *multiplicative* form as a product of ${}_0$ penalty terms. To the best of our knowledge, this paper is the first to propose a mixed additive-multiplicative penalty form that can serve as an engine for simultaneous classification and estimation. The method works by using each of the ${}_0$ penalty terms in the *multiplicative* expression to shrink the individual-level regression parameter vectors to a particular *unknown* group-level parameter vector, thereby producing a joint shrinkage process to unknown quantities. This process is distinct from the prototypical Lasso method that shrinks an individual parameter to the *known* value zero and the group Lasso method that shrinks a parameter vector to a *known* vector of zeros (see Yuan and Lin, 2006). To emphasize its role as a classifier and for future reference, we describe our new Lasso method as the *classifier-Lasso* or *C-Lasso*.

Fourth, we develop a double asymptotic limit theory for the C-Lasso that demonstrates its capacity to achieve simultaneous classification and estimation in a single step. As mentioned in the Abstract, the paper develops two classes of estimators for panel structure models – penalized profile likelihood (PPL) and penalized GMM (PGMM). The former is applicable to both linear and nonlinear panel models without endogeneity and with or without dynamic structures, while the latter is applicable to linear panel models with endogeneity or dynamic structures. Both broaden the scope of applicability of our method as early literature only considers linear panels without endogeneity. In either case, we show uniform classification consistency in the sense that all individuals belonging to a certain group can be classified into the same group correctly uniformly over both individuals and group identities with probability approaching one (w.p.a.1). Conversely, all individuals that are classified into a certain group belong to the same group uniformly over both individuals and group identities w.p.a.1. Such a uniform result allows us to establish an *oracle* property of the PPL estimator that, like the BM K-means estimator, is asymptotically equivalent to the corresponding infeasible estimator of the group-specific parameter that is obtained by knowing all individual group identities. Unfortunately, our PGMM estimator generally does not have the oracle property. But the uniform classification consistency property allows us to develop a limit theory for post-C-Lasso estimators that are obtained by pooling all individuals in an estimated group to estimate the group-specific parameters and these estimators are asymptotically as efficient as the oracle ones in both the PPL and PGMM contexts.

Fifth, C-Lasso enables empirical researchers to study panel structures without *a priori* knowledge of the number of groups, without the need to specify any ancillary regression models to model individual group identities, and with no need to make any distributional assumptions. When the number ${}_0$ of groups is unknown, a BIC-type information criterion is proposed to determine the number of groups for both PPL and PGMM estimation and it is shown that this procedure selects the correct number of groups consistently.

The rest of the paper is organized as follows. We study C-Lasso PPL estimation and inference of panel structure models in Section 2. PGMM estimation and inference is addressed in Section 3. Section 4 reports Monte Carlo simulation findings. Section 5 contains two empirical applications. Section 6 concludes. Proofs of the main results in the paper are given in Appendices A and B. Additional materials may be found in the Supplemental Material.

For any real matrix we write the transpose $'$ the Frobenius norm $\| \ \|$ and the Moore-Penrose inverse as $^+$ When is symmetric, we use ${}_{\max}(\ )$ and ${}_{\min}(\ )$ to denote the largest and smallest eigenvalues, respectively. ${}_p$ and $\mathbf{0}_{p\times 1}$ denote the $\times$ identity matrix and $\times 1$ vector of zeros, and $\mathbf{1}\{\cdot\}$ is the indicator function. The operator $\xrightarrow{P}$ denotes convergence in probability, $\xrightarrow{D}$ convergence in distribution, and plim probability limit. We use ( ) $\rightarrow \infty$ to signify that and pass jointly to infinity.

fi

This section considers panel structure models without endogeneity. It is convenient to assume first that the number of groups is known and later consider the determination of the number of unknown groups.

Given a panel data set $\{({}_{it} \quad {}_{it})\}$ for $= 1$ and $= 1$ it is proposed to use fixed effects quasi maximum likelihood to estimate the unknown parameters by solving the minimization problem

$$\min_{\{\beta_i,\mu_i\}} \frac{1}{} \sum^{N} \sum^{T}_{i=1\ t=1} ({}_{it}; \ {}_{i} \ {}_{i}) \tag{2.1}$$

Here $-$ $({}_{it}; {}_{i} \ {}_{i})$ denotes the logarithm of the pseudo-true conditional density function of ${}_{it}$ given ${}_{it}$ the history of $({}_{it} \ {}_{it})$, and $({}_{i} \ {}_{i})$, where ${}_{i}$ are scalar individual effects and ${}_{i}$ are $\times 1$ vectors of parameters of interest. Traditionally, econometric work has assumed that the ${}_{i}$ are common for all cross sectional units, leading to a homogeneous panel with individual heterogeneity modeled through ${}_{i}$ alone. At the other extreme, the ${}_{i}$ are assumed to differ across individuals and each is estimated at a slow rate without pooling across section. The present paper allows the true values of ${}_{i}$ denoted ${}^{0}_{i}$ to follow a group pattern of the general form

$$ {}^{0}_{i} = \sum^{K_0}_{k=1} {}^{0}_{k} \mathbf{1}\{ \ \in \ {}^{0}_{k}\} \tag{2.2}$$

Here ${}^{0}_{j} \neq {}^{0}_{k}$ for any $\neq$, $\cup^{K_0}_{k=1} {}^{0}_{k} = \{1 \ 2 \quad \}$ and ${}^{0}_{k} \cap {}^{0}_{j} = \varnothing$ for any $\neq$ Let ${}_{k} = \# {}^{0}_{k}$ denote the cardinality of the set ${}^{0}_{k}$ In the economic growth literature (e.g., Phillips and Sul, 2007a), ${}_{0}$ corresponds to the number of convergence clubs and countries (indexed by) within the same ${}^{\text{th}}$ club share the same (slope) parameter vector ${}^{0}_{k}$ In the market entry-exit example (e.g., Hahn and Moon, 2010), ${}_{0}$ denotes the number of pure Nash equilibria and markets (indexed by) selecting the same equilibrium over time exhibit the same parameter vector.

For now, we assume that the number of groups, ${}_{0}$, is known and fixed but that each individual's group membership is unknown. In addition, following Sun (2005), Lin and Ng (2012), and BM, we implicitly assume that individual group membership does not vary over time. Let $\boldsymbol{\alpha} \equiv ({}_{1} \quad {}_{K_0})$ and $\boldsymbol{\beta} \equiv ({}_{1} \quad {}_{N})$ We denote the true values of ${}_{i} \ {}_{k} \ {}_{i}$ $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ as ${}^{0}_{i} \ {}^{0}_{k} \ {}^{0}_{i}$ $\boldsymbol{\alpha}^{0}$ and $\boldsymbol{\beta}^{0}$ respectively. The econometric task is to infer each individual's group identity and to estimate the group-specific parameters ${}^{0}_{k}$ Some examples of models that fall within this framework and the scope of our methodology are as follows.

(Linear panel) The linear panel structure model is generated according to

$$ {}_{it} = {}^{0\prime}_{i} \ {}_{it} + {}^{0}_{i} + {}_{it} \tag{2.3}$$

where ${}_{it}$ is a $\times 1$ vector of exogenous or predetermined variables, ${}_{i}$ is an individual fixed effect, ${}_{i}$ is a $\times 1$ vector of slope parameters, and ${}_{it}$ is the idiosyncratic error term with mean zero. Gaussian quasi-maximum likelihood estimation (QMLE) of ${}_{i}$ and ${}_{i}$ is achieved by minimizing (2.1) with $({}_{it}; {}_{i} \ {}_{i}) = \frac{1}{2}({}_{it} - {}^{\prime}_{i} \ {}_{it} - {}_{i})^{2}$ and ${}_{it} = ({}_{it} \ {}^{\prime}_{it})^{\prime}$

(Linear panel with quantile restrictions) Consider the model in (2.3) with the quantile restriction: $(_{it} \leq 0| \ _{it} \ ^{0}_{i} \ ^{0}_{i}) =$; see, e.g., Kato, Galvo, and Montes-Rojas (2012). We can estimate $_i$ and $_i$ by minimizing (2.1) with $(\ _{it}; \ _i \ _i) = \ _\tau(\ _{it} - \ '_i \ _{it} - \ _i)$ where $_\tau(\ ) = \{ \ - \ (- \ )\}$ is a smoothed version of the usual check function with being a CDF-type kernel function and a bandwidth parameter.

(Binary choice panel) The dynamic binary choice panel data model is characterized by $_{it} =$ $\mathbf{1}\{ \ ^{0\prime}_{i} \ _{it} + \ ^{0}_{i} - \ _{it} \geq 0\}$ where $_{it}$ $_i$ and $_{it}$ are as defined in Example 1. In this case, $- \ (\ _{it}; \ _i \ _i) =$ $_{it} \ln (\ _{it} - \ '_i \ _{it} - \ _i) + (1 - \ _{it}) \ln[1 - (\ _{it} - \ '_i \ _{it} - \ _i)]$ where $_{it} = (\ _{it} \ '_{it})'$ and $(\cdot)$ denotes the conditional CDF (standard logistic or normal) of $_{it}$ given $_{it}$ and the history of $(\ _{it} \ _{it})$

(Tobit panel) The Tobit panel is characterized by $_{it} = \max(0 \ ^{0\prime}_{i} \ _{it} + \ ^{0}_{i} + \ _{it})$ where $_{it}$ $_i$ and $_{it}$ are defined as in the above examples. For clarity, assume that $_{it}$'s are independent and identically distributed (IID) $(0 \ ^{2}_{\varepsilon})$ given $_{it}$ and the history of $(\ _{it} \ _{it})$ In this case, $- (\ _{it}; \ _i \ _i \ ^{2}_{\varepsilon}) =$ $\mathbf{1}\{ \ _{it} = 0\} \ln ((\ _{it} - \ '_i \ _{it} - \ _i) \ ^{2}_{\varepsilon}) + \mathbf{1}\{ \ _{it} \ 0\} \ln[((\ _{it} - \ '_i \ _{it} - \ _i) \ ^{2}_{\varepsilon}) \ _{\varepsilon}]$ where $_{it} = (\ _{it} \ '_{it})'$ and denotes the standard normal PDF and CDF, respectively. The presence of the common parameter $^{2}_{\varepsilon}$ can be addressed by extending the asymptotic analysis below.

fi $\boldsymbol{\alpha}$ $\boldsymbol{\beta}$

Following Hahn and Newey (2004) and Hahn and Kuersteiner (2011), the profile log-likelihood function is

$$_{1,NT}(\boldsymbol{\beta}) = \frac{1}{\ } \sum_{i=1}^{N} \sum_{t=1}^{T} (\ _{it}; \ _i \ \hat{}_i(\ _i)) \tag{2.4}$$

where $\hat{}_i(\ _i) = \arg\min_{\mu_i} \frac{1}{T} \sum_{t=1}^{T} (\ _{it}; \ _i \ _i)$ Motivated by the literature on group Lasso (e.g., Yuan and Lin 2006), we propose to estimate $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$ by minimizing the following PPL criterion function

$$^{(K_0)}_{1NT,\lambda_1}(\boldsymbol{\beta} \ \boldsymbol{\alpha}) = \ _{1,NT}(\boldsymbol{\beta}) + \frac{1}{\ } \sum_{i=1}^{N} \Pi_{k=1}^{K_0} \| \ _i - \ _k \| \tag{2.5}$$

where $_1 = \ _{1NT}$ is a tuning parameter. Minimizing the above criterion function produces *classifier-Lasso* (C-Lasso) estimates $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\alpha}}$ of $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$ respectively. Let $\hat{}_i$ and $\hat{}_k$ denote the $^{\text{th}}$ and $^{\text{th}}$ columns of $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\alpha}}$, respectively, i.e., $\hat{\boldsymbol{\alpha}} \equiv (\hat{}_1 \ \hat{}_K)$ and $\hat{\boldsymbol{\beta}} \equiv (\hat{}_1 \ \hat{}_N)$

The penalty term in (2.5) takes a novel mixed *additive-multiplicative* form that does not appear in the literature. Traditional Lasso includes additive penalty terms to an objective function by differentiating zeros from non-zero-valued parameters to select relevant regressors. In contrast, the C-Lasso has additive terms, each of which takes a multiplicative form as the product of $_0$ separate penalties. The multiplicative component is needed because for each $^{0}_{i}$ can take any one of the $_0$ *unknown* values, $^{0}_{1}$ $^{0}_{K_0}$ We do not know *a priori* to which point $_i$ should shrink, and all $_0$ possibilities must be allowed. Each of the $_0$ penalty terms in the multiplicative expression permits $_i$ to shrink to a particular *unknown* group-level parameter vector $_k$ The summation component is needed because we need to pull information from all cross sectional units in order to identify $\{ \ ^{0}_{i}\}$ and $\{ \ ^{0}_{k}\}$ jointly. Our approach differs from the prototypical Lasso method of Tibshirani (1996) that shrinks a parameter to zero as well as the group Lasso method of Yuan and Lin (2006) that shrinks a parameter vector to a zero vector. The main purpose in the latter papers is to select relevant variables while C-Lasso is designed to determine group membership for each individual. As emphasized in the Introduction, both problems enjoy the same motivation of parameter sparsity despite their different nature. C-Lasso has the additional motivation of classification of unknown parameters into *a priori* unknown groups each with their own *unknown* parameters.

Note that the objective function in (2.5) is not convex in $\boldsymbol{\beta}$ even though it is (conditionally) convex in $_k$ when one fixes $_j$ for $\neq$ The supplement provides an iterative algorithm to obtain the estimates $\hat{\boldsymbol{\alpha}}$ and $\hat{\boldsymbol{\beta}}$

Let $_i(\ _i) \equiv \arg\min_{\mu_i} \Psi_i(\ _i\ \ _i)$ where $\Psi_i(\ _i\ \ _i) \equiv \frac{1}{T} \sum_{t=1}^{T} \mathbb{E}[\ (\ _{it};\ _i\ \ _i)]$ Note that $^0_i = \ _i(\ ^0_i)$ Let $_i(\ _{it};\ _i\ \ _i) \equiv\ (\ _{it};\ _i\ \ _i)\ \ _i$ and $_i(\ _{it};\ _i\ \ _i) \equiv\ (\ _{it};\ _i\ \ _i)\ \ _i$ Let $^{\mu_i}_i$ and $^{\mu_i\mu_i}_i$ denote the first and second derivatives of $_i$ with respect to $_i$ Define $^{\mu_i}_i$ $^{\mu_i\mu_i}_i$ $^{\beta_i}_i$ and $^{\beta_i}_i$ similarly. For notational simplicity, denote $_{it} \equiv\ _i(\ _{it};\ ^0_i\ \ ^0_i)$ and similarly for $^{\mu_i}_{it}$ $^{\mu_i\mu_i}_{it}$ $_{it}$ $^{\mu_i}_{it}$ and $^{\mu_i\mu_i}_{it}$ Define

$$_{iU} \equiv \frac{1}{T}\sum_{t=1}^{T} \mathbb{E}(\ ^{\mu_i}_{it}) \quad _{iV} \equiv \frac{1}{T}\sum_{t=1}^{T} \mathbb{E}(\ ^{\mu_i}_{it}) \quad _{iU2} \equiv \frac{1}{T}\sum_{t=1}^{T} \mathbb{E}(\ ^{\mu_i\mu_i}_{it}) \quad _{iV2} \equiv \frac{1}{T}\sum_{t=1}^{T} \mathbb{E}(\ ^{\mu_i\mu_i}_{it})$$

$$\mathbb{U}_{it} \equiv\ _{it} - \frac{_{iU}}{_{iV}}\ _{it} \quad \mathbb{U}^{\beta_i}_{it} \equiv\ ^{\beta_i}_{it} - \frac{_{iU}}{_{iV}}\ ^{\beta_i\prime}_{it} \text{ and } \mathbb{U}^{\mu_i}_{it} \equiv\ ^{\mu_i}_{it} - \frac{_{iU}}{_{iV}}\ ^{\mu_i}_{it}$$

Let $\Omega_{iT} \equiv \frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}\mathbb{E}(\mathbb{U}_{it}\mathbb{U}'_{is})$ $\mathbb{H}_{iT} \equiv \frac{1}{T}\sum_{t=1}^{T}\mathbb{E}[\mathbb{U}^{\beta_i}_{it}]$ and $\mathbb{H}_{kNT} \equiv \frac{1}{N_k}\sum_{i\in G^0_k}\mathbb{H}_{iT}$ Define the two expected Hessian matrices for cross sectional unit :

$$_{i\mu\mu}(\ _i) \equiv \frac{1}{T}\sum_{t=1}^{T}\mathbb{E}[\ ^{\mu_i}_i(\ _{it};\ _i\ \ _i(\ _i))] \text{ and } _{i\beta\beta}(\ _i) \equiv \frac{1}{T}\sum_{t=1}^{T}\mathbb{E}\left[\ ^{\beta_i}_{it}(\ _i) + \ ^{\mu_i}_{it}(\ _i)\frac{_i(\ _i)}{'_i}\right]$$

where $^{\beta_i}_{it}(\ _i) = \ ^{\beta_i}_i(\ _{it};\ _i\ \ _i(\ _i))$ and similarly for $^{\mu_i}_{it}(\ _i)$ Let $\min_i$ denote $\min_{1\leq i\leq N}$ and similarly for $\max_i$ We make the following assumptions

ASSUMPTION A1. (i) *For each* $\{\ _{it}: \ = 1\ 2\ \}$ *is stationary strong mixing with mixing coefficients* $_i(\cdot)$. $(\cdot) \equiv \max_i\ _i(\cdot)$ *satisfies* $(\ ) \leq\ _\alpha\ ^\tau$ *for some* $_\alpha\ \ 0$ *and* $\in (0\ 1)$. $\{\ _{it}: \ = 1\ 2\ \}$ *are independent across*

(ii) *For each* 0 $\min_i[\inf_{(\beta_i,\mu_i):\|(\beta_i,\mu_i)-(\beta^0_i,\mu^0_i)\|>\eta}\Psi_i(\ _i\ \ _i) - \Psi_i(\ ^0_i\ \ ^0_i)]\ \ 0$

(iii) *Let* Θ *denote the parameter space for* $_i = (\ '_i\ \ _i)'$ Θ *is a compact and convex subset of* $\mathbb{R}^{p+1}$ *such that* $^0_i = (\ ^{0\prime}_i\ \ ^0_i)'$ *lies in the interior of* Θ *for each*

(iv) *Let* $|\ | \equiv \sum_{j=1}^{p+1}\ _j$ *and* $^v(\ _{it};\ ) \equiv\ ^{|v|}(\ _{it};\ )\ (\ _{(1)}\cdots\ _{(p+1)})$ *where* $= (\ _1\cdots\ _{p+1})$ *is a vector of nonnegative integers and* $_{(j)}$ *denotes the* *th element of* . *There exists a function* $(\cdot)$ *such that* $\sup_{\theta\in\Theta}\|\ ^v(\ _{it};\ )\| \leq\ (\ _{it})\ \|\ ^v(\ _{it};\ ) - \ ^v(\ _{it};\ \bar{\ })\| \leq\ (\ _{it})\|\ - \bar{\ }\|$ *for any* $\bar{\ } \in \Theta$ *and* $|\ | \leq 3$ *and* $\max_i\mathbb{E}|\ (\ _{it})|^q\ \ _M$ *for some* $_M\ \ \infty$ *and* ≥ 6

(v) *There exists a constant* $_H\ \ 0$ *such that* $\min_i\inf_{\beta\in\mathcal{B}}\ _{i\mu\mu}(\ ) \geq\ _H$ *and* $\min_i\ _{\min}(\ _{i\beta\beta}(\ ^0_i)) \geq\ _H$

(vi) *There exists a constant* $_\alpha\ \ 0$ *such that* $\min_{1\leq k<l\leq K_0}\|\ ^0_k - \ ^0_l\| \geq\ _\alpha$

(vii) $_0$ *is fixed and* $_k\ \ \to\ _k \in (0\ 1)$ *for each* $= 1\ \ \ _0$ *as* $\to\infty$

ASSUMPTION A2. (i) $^2_1(\ln\ )^{6+2\nu} \to \infty$ *and* $_1(\ln\ )^\nu \to 0$ *for some* 0 *as* $(\ \ ) \to \infty$

(ii) $^{1/2}\ ^{-1}(\ln\ )^9 \to 0$ *and* $^2\ ^{1-q/2} \to\ \ \in [0\ \infty)$ *as* $(\ \ ) \to \infty$.

ASSUMPTION A3. (i) *For each* $= 1 \quad {}_0$ $\Omega_k \equiv \lim_{(N_k,T)\to\infty} \frac{1}{N_k} \sum_{i\in G_k^0} \Omega_{iT}$ *exists and* $\Omega_k \quad 0$

(ii) *For each* $= 1 \quad {}_0$, $\mathbb{H}_k \equiv \lim_{(N_k,T)\to\infty} \mathbb{H}_{kNT}$ *exists and* $\mathbb{H}_k \quad 0$

Assumption A1(i) imposes conditions on $\{ \ _{it}\}$ which are commonly assumed for dynamic nonlinear panel data model; see, e.g., Hahn and Kuersteiner (2011) and Lee and Phillips (2015). With more complicated notation, we can relax the stationarity assumption along the time dimension. A1(ii) imposes an identification condition for the joint identification of $(\ _i \ \ _i)$ for each A1(iii) restricts the parameter space and it is possible to allow Θ to be -dependent. A1(iv) specifies the smoothness and moment conditions on or objects associated with it. A1(v), in conjunction with A1(ii) and (iv), implies that $\min_i[\inf_{\mu_i:|\mu_i-\mu_i(\beta_i)|>\eta} \Psi_i(\ _i \ \ _i) - \Psi_i(\ _i \ \ _i(\ _i))] \quad 0$ and $\min_i[\inf_{\beta_i:\|\beta_i-\beta_i^0\|>\eta} \Psi_i(\ _i \ \ _i(\ _i)) - \Psi_i(\ _i^0 \ \ _i(\ _i^0))] \quad 0$ A1(vi) specifies that the group-specific parameters are separated from each other, similar to the separation requirement in Hahn and Moon (2010). A1(vii) implies that each group has an asymptotically non-negligible membership number of individuals as $\to \infty$ This assumption can also be relaxed at the cost of more lengthy arguments. Assumption A2(i) imposes conditions on ${}_1$ all of which hold if

$$ {}_1 \propto \quad {}^{-a} \text{ for any } \in (0 \ 1 \ 2) \tag{2.6}$$

A2(ii) is needed to ensure some higher order terms are asymptotically negligible. A3 is used to derive the asymptotic bias and variance of the C-Lasso estimator. The theory developed below under these conditions does not require correct specification of the likelihood function and the C-Lasso asymptotics apply under the general QMLE setup.

ffi

The following theorem establishes the consistency of the PPL estimates $\{\hat{\ }_i\}$ and $\{\hat{\ }_k\}$

Suppose that Assumption A1 holds and ${}_1 = \ (1)$. *Then (i)* $\hat{\ }_i - \ _i^0 = \ _P(\ ^{-1/2} + \ _1)$ *for* $= 1 \ 2$ *(ii)* $\frac{1}{N} \sum_{i=1}^{N} \left\| \hat{\ }_i - \ _i^0 \right\|^2 = \ _P(\ ^{-1})$ *and (iii)* $(\hat{\ }_{(1)} \quad \hat{\ }_{(K_0)}) - (\ _1^0 \quad \ _{K_0}^0) = \ _P(\ ^{-1/2})$ *where* $(\hat{\ }_{(1)} \quad \hat{\ }_{(K_0)})$ *is a suitable permutation of* $(\hat{\ }_1 \quad \hat{\ }_{K_0})$

Theorem 2.1(i)-(ii) establish the pointwise and mean-square convergence of $\hat{\ }_i$. Theorem 2.1(iii) indicates that the group-specific parameters $\ _1^0 \quad \ _K^0$ can be estimated consistently by $\hat{\ }_1 \quad \hat{\ }_{K_0}$ subject to permutation. As expected and consonant with other Lasso limit theory, the pointwise convergence rate of $\hat{\ }_i$ depends on the rate at which the tuning parameter ${}_1$ converges to zero. Somewhat unexpectedly, this requirement is not the case either for mean-square convergence of $\hat{\ }_i$ or convergence of $\hat{\ }_k$. For notational simplicity, hereafter we simply write $\hat{\ }_k$ for $\hat{\ }_{(k)}$ as the consistent estimator of $\ _k^0$, and define

$$\hat{\ }_k = \left\{ \ \in \{1 \ 2 \quad \} : \hat{\ }_i = \hat{\ }_k \right\} \text{ for } = 1 \quad {}_0 \tag{2.7}$$

fi

Roughly speaking, a classification method is consistent if it classifies each individual to the correct group w.p.a.1. For a rigorous statement of this property we define

$$\hat{\ }_{kNT,i} \equiv \left\{ \ \in \hat{\ }_k \mid \ \in \ _k^0 \right\} \text{ and } \hat{\ }_{kNT,i} \equiv \left\{ \ \in \ _k^0 \mid \ \in \hat{\ }_k \right\} \tag{2.8}$$

where $= 1$ and $= 1$ $_0$ Let $\hat{}_{kNT} = \cup_{i \in G_k^0} \hat{}_{kNT,i}$ and $\hat{}_{kNT} = \cup_{i \in \hat{G}_k} \hat{}_{kNT,i}$ $\hat{}_{kNT}$ and $\hat{}_{kNT}$ mimic Type I and II errors in statistical tests: $\hat{}_{kNT}$ denotes the error event of not classifying an element of ${}_k^0$ into the estimated group $\hat{}_k$; and $\hat{}_{kNT}$ denotes the error event of classifying an element that does not belong to ${}_k^0$ into the estimated group $\hat{}_k$ Both types of errors must be controlled. We use the following definition.

fi **fi** The classification is *individually consistent* if $(\hat{}_{kNT,i}) \to 0$ as $(\quad) \to \infty$ $\forall \in {}_k^0$ and $\in \{1 \quad {}_0\}$ and $(\hat{}_{kNT,i}) \to 0$ as $(\quad) \to \infty$ $\forall \in \hat{}_k$ and $\in \{1 \quad {}_0\}$ It is *uniformly consistent* if $(\cup_{k=1}^{K_0} \hat{}_{kNT}) \to 0$ and $(\cup_{k=1}^{K_0} \hat{}_{kNT}) \to 0$ as $(\quad) \to \infty$

The following theorem establishes uniform consistency for the PPL classifier.

Suppose that Assumptions A1-A2 hold. Then (i) $(\cup_{k=1}^{K_0} \hat{}_{kNT}) \leq \sum_{k=1}^{K_0} (\hat{}_{kNT}) \to 0$ *as* $(\quad) \to \infty$ *and (ii)* $(\cup_{k=1}^{K_0} \hat{}_{kNT}) \leq \sum_{k=1}^{K_0} (\hat{}_{kNT}) \to 0$ *as* $(\quad) \to \infty$

Theorem 2.2 implies that all individuals within a group, say ${}_k^0$ can be simultaneously correctly classified into the same group (denoted $\hat{}_k$) w.p.a.1. Conversely, all individuals that are classified into the same group, say $\hat{}_k$ simultaneously correctly belong to the same group (${}_k^0$) w.p.a.1. Let $\hat{}_0 \equiv \{1\ 2 \quad \} \backslash (\cup_{k=1}^{K_0} \hat{}_k)$ and $\hat{}_{iNT} \equiv \{ \in \hat{}_0\}$ Theorem 2.2(i) implies that $(\cup_{1 \leq i \leq N} \hat{}_{iNT}) \leq \sum_{k=1}^{K_0} (\hat{}_{kNT}) \to 0$ That is, all individuals can be classified into one of the ${}_0$ groups w.p.a.1. Nevertheless, when is not large, a small percentage of individuals could be left unclassified if we stick with the classification rule in (2.7). To ensure that all individuals are classified into one of the ${}_0$ groups in finite samples, we can modify the classifier. In particular, we classify $\in \hat{}_k$ if $\hat{}_i = \hat{}_k$ for some $= 1 \quad {}_0$ and $\in \hat{}_l$ for some $= 1 \quad {}_0$ if $||\hat{}_i - \hat{}_l|| = \min\{||\hat{}_i - \hat{}_1|| \quad ||\hat{}_i - \hat{}_{K_0}||\}$ and $\sum_{k=1}^{K_0} \mathbf{1}\{\hat{}_i = \hat{}_k\} = 0$ Since the event $\sum_{k=1}^{K_0} \mathbf{1}\{\hat{}_i = \hat{}_k\} = 0$ occurs w.p.a.1 uniformly in we can ignore it in large samples in subsequent theoretical analysis and restrict our attention to the classification rule in (2.7) to avoid confusion.

Let $\hat{}_k \equiv \sum_{i=1}^{N} \mathbf{1}\{ \in \hat{}_k\}$ The following corollary studies the consistency of $\hat{}_k$.

Suppose that Assumptions A1-A2 hold. Then $\hat{}_k - {}_k = {}_P(1)$ *for* $= 1 \quad {}_0$

The following theorem reports the oracle property of the Lasso estimator $\{\hat{}_k\}$.

Suppose Assumptions A1-A3 hold. Then $\sqrt{\quad_k}(\hat{}_k - {}_k^0) - \mathbb{H}_{kNT}^{-1}\mathbb{B}_{kNT} \xrightarrow{D} (0\ \mathbb{H}_k^{-1}\Omega_k(\mathbb{H}_k^{-1})')$ *where* $\mathbb{B}_{kNT} = \mathbb{B}_{1kNT} - \mathbb{B}_{2kNT}$ $\mathbb{B}_{1kNT} = \frac{1}{\sqrt{N_k T^3}} \sum_{i \in G_k^0} {}_{iV}^{-1} \sum_{s=1}^{T} \sum_{t=1}^{T} {}_{is}\mathbb{U}_{it}^{\mu_i}$ *and* $\mathbb{B}_{2kNT} = \frac{1}{2\sqrt{N_k T}} \sum_{i \in G_k^0} {}_{iV}^{-2}({}_{iU2} - \frac{m_{iV2}}{m_{iV}} {}_{iU})(\frac{1}{\sqrt{T}} \sum_{t=1}^{T} {}_{it})^2$ *for* $= 1 \quad {}_0$

$\mathbb{B}_{kNT}$ is written as the difference between two terms that are derived from the first and second order Taylor expansions of the PPL estimating equation, respectively. Comparing the above result with HK, we find that the quantities Ω_k $\mathbb{H}_k$ and $\mathbb{B}_k$ coincide with the corresponding terms in HK; see the remark after Lemma S1.12 for details. Then we can use the formula in HK to estimate the asymptotic bias and variance with obvious modifications. Alternatively, we can use the jackknife to correct bias; see Hahn and Newey (2004) and Dhaene and Jachmans (2015) for static and dynamic models, respectively.

If group membership is known, the *oracle* estimator of $_k$ is given by $\hat{}_{G_k^0} \equiv \arg\min_{\alpha_k} \frac{1}{N_k T} \sum_{i \in G_k^0} \sum_{t=1}^{T} (\; _{it}; \; _k \; \hat{}_i (\; _k))$ Then following our asymptotic analysis or that of HK, we can readily show that $\sqrt{\; _k}(\hat{}_{G_k^0} - \; _k^0) - \mathbb{H}_{kNT}^{-1}\mathbb{B}_{kNT} \xrightarrow{D} (0 \; \mathbb{H}_k^{-1}\Omega_k(\mathbb{H}_k^{-1})')$ under Assumptions A1 and A3. Theorem 2.4 indicates that the PPL estimator $\hat{}_k$ achieves the same limit distribution as this oracle estimator. In this sense, we say that the PPL estimators $\{\hat{}_k\}$ enjoy the asymptotic oracle property. In addition, given the estimated groups $\hat{}_k$ we can obtain the post-Lasso estimator of $_k$ by $\hat{}_{\hat{G}_k} \equiv \arg\min_{\alpha_k} \frac{1}{N_k T} \sum_{i \in \hat{G}_k} \sum_{t=1}^{T} (\; _{it}; \; _k \; \hat{}_i (\; _k))$ The following theorem reports the asymptotic distribution of $\hat{}_{\hat{G}_k}$

Suppose Assumptions A1-A3 hold. Then $\sqrt{\; _k}(\hat{}_{\hat{G}_k} - \; _k^0) - \mathbb{H}_{kNT}^{-1}\mathbb{B}_{kNT} \xrightarrow{D} (0 \; \mathbb{H}_k^{-1}\Omega_k(\mathbb{H}_k^{-1})')$ *for* $= 1 \quad _0$ *where* $\mathbb{B}_{kNT}$ *is as defined in Theorem 2.4.*

Theorems 2.4 and 2.5 indicate that $\hat{}_k$ and $\hat{}_{\hat{G}_k}$ are asymptotically equivalent. In a totally different framework, Belloni and Chernozhukov (2013) study post-Lasso estimators which apply OLS to the model selected by first-step penalized estimators and show that the post-Lasso estimators perform at least as well as Lasso in terms of rate of convergence and have the advantage of smaller bias. Correspondingly, it would be interesting to compare the higher-order asymptotic properties of $\hat{}_k$ and $\hat{}_{\hat{G}_k}$ in future work.

Note that our asymptotic results are "pointwise" in the sense that the unknown parameters are treated as fixed. The implication is that in finite samples, the distributions of our estimators can be quite different from normal, as discussed in Leeb and Pötscher (2008, 2009). This is a well-known challenge for shrinkage estimators. Despite its importance, developing a thorough theory on uniform inference in this context is beyond the scope of the present work.

In practice, the exact number of groups is typically unknown. We assume that $_0$ is bounded from above by a finite integer $_{\max}$ and study the determination of the number of groups via some information criterion (IC). By minimizing (2.5) with $_0$ replaced by we obtain the C-Lasso estimates $\{\hat{}_i(\; _1) \; \hat{}_k(\; _1)\}$ of $\{ \; _i \; _k\}$ where we make the dependence of $\hat{}_i$ and $\hat{}_k$ on $(\; _1)$ explicit. As above, we classify individual into group $\hat{}_k(\; _1)$ if and only if $\hat{}_i(\; _1) = \hat{}_k(\; _1)$, i.e., $\hat{}_k(\; _1) \equiv \{ \; \in \{1 \; 2 \quad \} : \hat{}_i(\; _1) = \hat{}_k(\; _1)\}$ for $= 1$ Let $\hat{}(\; _1) \equiv \{\hat{}_1(\; _1) \quad \hat{}_K(\; _1)\}$ The post-Lasso estimator of $_k^0$ is denoted as $\hat{}_{\hat{G}_k(K,\lambda_1)}$ We propose to select to minimize

$$_1(\quad _1) \equiv \frac{2}{\quad} \sum_{k=1}^{K} \sum_{i \in \hat{G}_k(K,\lambda_1)} \sum_{t=1}^{T} \left(\; _{it}; \hat{}_{\hat{G}_k(K,\lambda_1)} \; \hat{}_i(\hat{}_{\hat{G}_k(K,\lambda_1)})\right) + \; _{1NT} \tag{2.9}$$

where $_{1NT}$ is a tuning parameter. Let $\hat{}(\; _1) \equiv \arg\min_{1 \le K \le K_{\max}} \; _1(\quad _1)$ See Wang, Li, and Tsai (2007), Liao (2013), and Lu and Su (2016) for the use of a similar IC in various contexts.

Let $^{(K)} \equiv (\; _{K,1} \quad _{K,K})$ be any -partition of $\{1 \; 2 \quad \}$ and $\mathcal{G}_K$ a collection of all such partitions. Let $\hat{}^2_{G^{(K)}} \equiv \frac{2}{NT} \sum_{k=1}^{K} \sum_{i \in G_{K,k}} \sum_{t=1}^{T} (\; _{it}; \hat{}_{G_{K,k}} \; \hat{}_i(\hat{}_{G_{K,k}}))$ where $\hat{}_{G_{K,k}} \equiv \arg\min_{\alpha_k} \frac{1}{N_k T} \sum_{i \in G_{K,k}} \sum_{t=1}^{T} (\; _{it}; \; _k \; \hat{}_i(\; _k))$. We add the following two assumptions.

ASSUMPTION A4. *As* $(\quad) \to \infty$ $\min_{1 \le K < K_0} \inf_{G^{(K)} \in \mathcal{G}_K} \hat{}^2_{G^{(K)}} \xrightarrow{P} \underline{\;}^2 \quad _0^2$ *where* $_0^2 \equiv lim_{(N,T)\to\infty} \frac{2}{NT} \sum_{k=1}^{K_0} \sum_{i \in G_k^0} \sum_{t=1}^{T} \mathbb{E}[(\; _{it}; \; _k^0 \; _i^0)]$

ASSUMPTION A5. *As* $(\quad) \to \infty$ $\quad_{1NT} \to 0$ *and* $\quad_{1NT} \quad \to \infty.$

Assumption A4 is intuitively clear and applies under primitive conditions in a variety of models, such as panel autoregressions. It requires that all under-fitted models yield asymptotic mean square errors that are larger than $\ _0^2$, which is delivered by the true model. A5 reflects the usual conditions for the consistency of model selection: $\ _{1NT}$ cannot shrink to zero either too fast or too slowly.

The following theorem justifies the use of (2.9) as a selector criterion for

Suppose Assumptions A1-A5 hold. Then $(\hat{\ }(\ _1) = \ _0) \to 1$ *as* $(\quad) \to \infty$

As Theorem 2.6 indicates, as long as $\ _1$ satisfies Assumption A2(i), we can ensure that the correct number of groups is chosen w.p.a.1. In practice, we can fine-tune this parameter over a finite set, e.g., $\Lambda_1 \equiv \{\ _1 = \ _j \ ^{-1/3} \ _j = \ _0 \ ^j$ for $\ = 1 \quad \}$ for some $\ _0 \quad 0$ and $\quad 1$ That is, we pick up $\ _1 \in \Lambda_1$ such that $\ _1(\hat{\ }(\ _1) \quad _1)$ is minimized. We can show that with such a choice of $\ _1$ Theorem 2.6 continues to hold. Alternatively, we can consider a data-driven cross-validation procedure.

For the linear model in (2.3) with $\mathbb{E}(\ _{it}| \ _{it} \ _i^0) = 0$, we have $(\ _{it}; \ _i \ _i) = \frac{1}{2}(\ _{it} - \ _i' \ _{it} - \ _i)^2$ $\hat{\ }_i(\ _i) = \bar{\ }_{i\cdot} - \ _i' \bar{\ }_{i\cdot}$ and $\ _{1,NT}(\boldsymbol{\beta}) = \frac{1}{NT}\sum_{i=1}^N \sum_{t=1}^T (\tilde{\ }_{it} - \ _i' \tilde{\ }_{it})^2$ where $\bar{\ }_{i\cdot} = \frac{1}{T}\sum_{t=1}^T \ _{it}$ $\tilde{\ }_{it} = \ _{it} - \bar{\ }_{i\cdot}$ and $\bar{\ }_{i\cdot}$ and $\tilde{\ }_{it}$ are analogously defined. So the PPL problem becomes the penalized least squares (PLS) problem considered in Su, Shi, and Phillips (2014, SSP hereafter). In addition, we can verify that $\ _i(\ _i) = \mathbb{E}(\bar{\ }_{i\cdot}) - \ _i'\mathbb{E}(\bar{\ }_{i\cdot})$ $\ _i(\ _{it}; \ _i \ _i) = -(\ _{it} - \ _i' \ _{it} - \ _i) \ _{it}$ $\ _i(\ _{it}; \ _i \ _i) = -(\ _{it} - \ _i' \ _{it} - \ _i)$ $\ _i^{\mu_i}(\ _{it}; \ _i \ _i) = \ _{it} = \ _i^{\beta_i}(\ _{it}; \ _i \ _i)$ $\ _i^{\mu_i}(\ _{it}; \ _i \ _i) = 1$ $\ _i^{\beta_i}(\ _{it}; \ _i \ _i) = \ _{it} \ _{it}'$ $\mathbb{U}_{it} = -\ _{it}[\ _{it} - \mathbb{E}(\bar{\ }_{i\cdot})]$ $\mathbb{U}_{it}^{\beta_i} = [\ _{it} - \mathbb{E}(\bar{\ }_{i\cdot})] \ _{it}'$ $\mathbb{U}_{it}^{\mu_i} = \ _{it} - \mathbb{E}(\bar{\ }_{i\cdot})$ $\ _{i\mu\mu}(\ _i) = 1$ $\ _{i\beta\beta}(\ _i) = \frac{1}{T}\sum_{t=1}^T \mathbb{E}\{[(\ _{it} - \mathbb{E}(\bar{\ }_{i\cdot})][\ _{it} - \mathbb{E}(\bar{\ }_{i\cdot})]'\}$

$$\Omega_{iT} = \frac{1}{\ }\sum_{t=1}^T \sum_{s=1}^T \mathbb{E}\{\ _{it} \ _{is}[\ _{it} - \mathbb{E}(\bar{\ }_{i\cdot})][\ _{is} - \mathbb{E}(\bar{\ }_{i\cdot})]'\} \text{ and}$$

$$\mathbb{H}_{iT} = \frac{1}{\ }\sum_{t=1}^T \mathbb{E}\{[\ _{it} - \mathbb{E}(\bar{\ }_{i\cdot})][\ _{it} - \mathbb{E}(\bar{\ }_{i\cdot})]'\}$$

With the above calculations, we can readily verify that Assumptions A1(ii), (iv)-(v) and A3 hold under weak conditions. In addition, we can show that

$$\mathbb{B}_{1kNT} = \frac{-1}{\sqrt{\ _k \ ^3}} \sum_{i \in G_k^0} \sum_{t=1}^T \sum_{s=1}^T \ _{it}[\ _{is} - \mathbb{E}(\bar{\ }_{i\cdot})] = \mathbb{B}_{1k} + \ _P(1) \text{ and } \mathbb{B}_{2kNT}$$

analogously defined as $\hat{\ }_{\hat{G}_k}$ In practice, $\hat{\ }^2_{\hat{G}(K,\lambda_1)}$ is frequently replaced by its natural logarithm as in standard BIC to obtain

$$_1(\quad _1) = \ln \hat{\ }^2_{\hat{G}(K,\lambda_1)} + \quad _{1NT} \tag{2.10}$$

which will be used in our simulations and applications. But because the fixed effects are eliminated in the within-group transformed model, the $\sqrt{\ }$-convergence rates of their estimates won't play a role to ensure the selection consistency of $\ _1$ SSP show that the requirement on $\ _{1NT}$ can be relaxed with Assumption A5 replaced by:

ASSUMPTION A5*. *As* $(\quad) \to \infty \quad _{1NT} \to 0$ *and* $\ _{1NT} \ ^2_{NT} \to \infty$ *where* $\ _{NT} = \ ^{1/2} \ ^{1/2}$ *if* $\ _{it}$ *is strictly exogenous and* $\min(\ ^{1/2} \ ^{1/2} \ )$ *otherwise.*

In some applications, certain parameters of interest may be common across all individuals whereas others are group-specific. For instance, Pesaran, Shin, and Smith (1999) constrain the long-run coefficients to be identical across individuals while assuming the short-run coefficients to be heterogenous, or in our case, group-specific. Example 4 above is another instance. To keep up with the early notation, we write the negative log-likelihood function as $(\ _{it}; \ _i \quad _i)$ where is the common parameter and the $\ _i$ have a group structure as before. The negative profile log-likelihood function now becomes $\ _{1,NT}(\boldsymbol{\beta} \quad) = \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} (\ _{it}; \ _i \quad \hat{\ }_i(\ _i \ ))$ where $\hat{\ }_i(\ _i \ ) = \arg\min_{\mu_i} \frac{1}{T} \sum_{t=1}^{T} (\ _{it}; \ _i \quad _i)$ Then we can estimate $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$ by minimizing the following PPL criterion function

$$^{(K_0)}_{1NT,\lambda_1}(\boldsymbol{\beta} \ \boldsymbol{\alpha} \quad) = \ _{1,NT}(\boldsymbol{\beta} \quad) + \frac{1}{\ } \sum_{i=1}^{N} \Pi_{k=1}^{K_0} \| \ _i - \ _k \| \tag{2.11}$$

Our previous analysis can be followed to establish uniform consistency for the classifier and the oracle property for the resulting estimators of the group-specific parameters $\ _k$ and the common parameter .

When we have time effects $\{ \ _t \}$ we generally cannot eliminate them through transformation even in a linear panel structure model because of the slope heterogeneity. In this case, we need to estimate $= (\ _1 \quad \ _T)'$ jointly with $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$ in (2.11). A formal asymptotic analysis of this case is left for future work.

This section considers penalized GMM estimation of linear panel structure models when some regressors are lagged dependent variables or endogenous.

$\boldsymbol{\alpha}$ $\boldsymbol{\beta}$

To stay focused, we restrict attention to the linear panel structure model in (2.3).[1] We consider the first differenced system

$$\Delta \ _{it} = \ ^{0\prime}_i \Delta \ _{it} + \Delta \ _{it} \tag{3.1}$$

[1] Extension to general nonlinear panel data models with endogeneity and nonadditive fixed effects (e.g., Fernández-Val and Lee 2013) is possible but rigorous analysis raises additional statistical challenges and is left for future research.

where, e.g., $\Delta \ _{it} = \ _{it} - \ _{i,t-1}$ for $\ = 1$ and $\ = 1$ and we assume that $\ _{i0}$ and $\ _{i0}$ are observed. Let $\ _{it}$ be a $\ \times 1$ vector of instruments for $\Delta \ _{it}$ with $\ \geq$ Define $\Delta \ _i = (\Delta \ _{i1} \quad \Delta \ _{iT})'$ with similar definitions for $\Delta \ _i$ and $\Delta \ _i$ We propose to estimate $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$ by minimizing the following penalized GMM (PGMM) criterion function[2]

$$ {}^{(K_0)}_{2NT,\lambda_2}(\boldsymbol{\beta} \ \boldsymbol{\alpha}) = \ _{2,NT}(\boldsymbol{\beta}) + \frac{2}{} \sum_{i=1}^{N} \Pi_{k=1}^{K_0} \| \ _i - \ _k \| \tag{3.2} $$

where $\ _{2,NT}(\boldsymbol{\beta}) = \frac{1}{N} \sum_{i=1}^{N} \frac{1}{T} \sum_{t=1}^{T} \ _{it} (\Delta \ _{it} - \ '_i \Delta \ _{it})' \ _{iNT} \frac{1}{T} \sum_{t=1}^{T} \ _{it} (\Delta \ _{it} - \ '_i \Delta \ _{it}) \ _{iNT}$ is a $\times$ symmetric matrix that is asymptotically nonsingular and $\ _2 = \ _{2NT}$ is a tuning parameter. Minimizing (3.2) produces the PGMM estimates $\tilde{\boldsymbol{\alpha}}$ and $\tilde{\boldsymbol{\beta}}$ where $\tilde{\boldsymbol{\alpha}} \equiv (\tilde{\ }_1 \quad \tilde{\ }_{K_0})$ and $\tilde{\boldsymbol{\beta}} \equiv (\tilde{\ }_1 \quad \tilde{\ }_N)$

Let $\tilde{\ }_{i,z\Delta x} \equiv \frac{1}{T} \sum_{t=1}^{T} \ _{it}(\Delta \ _{it})'$ and $\bar{\ }_{i,z\Delta x} \equiv \mathbb{E}[\tilde{\ }_{i,z\Delta x}]$ Let $\ _{it} \equiv (\Delta \ _{it} \ (\Delta \ _{it})' \ '_{it})'$ $(\ _{it} \quad) \equiv \ _{it}(\Delta \ _{it} - \ '\Delta \ _{it})$ and $\bar{\ }_{i,T}(\ ) \equiv \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \{ \ (\ _{it} \quad) - \mathbb{E}[\ (\ _{it} \quad)]\}$ Let $\mathcal{B}_i$ denote the parameter space for $\ _i$ We make the following assumptions.

ASSUMPTION B1. (i) $\mathbb{E}[\ (\ _{it} \ ^0_i)] = 0$ *for each* $\ = 1$ *and* $\ = 1$

(ii) $\sup_{\beta \in \mathcal{B}_i} \| \bar{\ }_{i,T}(\ ) \| = \ _P(1)$ $\frac{1}{N} \sum_{i=1}^{N} \| \bar{\ }_{i,T}(\ _i) \|^2 = \ _P(1)$ *where* $\ _i \in \mathcal{B}_i$ *and* $(\max_i \| \bar{\ }_{i,T}(\ _i) \| \geq (\ln \ )^{3+\nu}) = (\ ^{-1})$ *for any* $\ 0$ *and* $\ 0$

(iii) $(\max_i \| \tilde{\ }_{i,z\Delta x} - \bar{\ }_{i,z\Delta x} \| \geq \ ) = (\ ^{-1})$ *for any* $\ 0$ *and* $\liminf_{(N,T)\to\infty} \min_i \ _{\min}(\bar{\ }'_{i,z\Delta x} \ \bar{\ }_{i,z\Delta x}) = \ ^2_{\bar{Q}} \quad 0$

(iv) *There exist nonrandom matrices* $\ _i$ *such that* $(\max_i \| \ _{iNT} - \ _i \| \geq \ ) = (\ ^{-1})$ *for any* $\ 0$ *and* $\liminf_{N\to\infty} \min_i \ _{\min}(\ _i) = \ _W \quad 0$

(v) *There exists a constant* $\ _\alpha \quad 0$ *such that* $\min_{1 \leq k < l \leq K_0} \| \ ^0_k - \ ^0_l \| \geq \ _\alpha$

(vi) $\ _0$ *is fixed and* $\ _k \quad \to \ _k \in (0 \ 1)$ *for each* $\ = 1 \quad \ _0$ *as* $\ \to \infty$

ASSUMPTION B2. (i) $\ ^2_2 (\ln \ )^{6+2\nu} \to \infty$ *and* $\ _2 (\ln \ )^\nu \to 0$ *for some* $\ 0$ *as* $(\quad) \to \infty$

(ii) *For any given* $\ 0$ $\max_i (\| \ ^{-1} \sum_{t=1}^{T} \ _{it} \Delta \ _{it} \| \geq \ _2) \to 0$ *as* $(\quad) \to \infty$

ASSUMPTION B3. (i) *For each* $\ = 1 \quad \ _0$, $\bar{\ }_k \equiv \frac{1}{N_k} \sum_{i \in G^0_k} \bar{\ }'_{i,z\Delta x} \ _i \ \bar{\ }_{i,z\Delta x} \to \ _k \quad 0$ *as* $(\quad) \to \infty$

(ii) *For each* $\ = 1 \quad \ _0$, $\frac{1}{\sqrt{N_k T}} \sum_{i \in G^0_k} \tilde{\ }'_{i,z\Delta x} \ _{iNT} \sum_{t=1}^{T} \ _{it} \Delta \ _{it} - \ _{kNT} \xrightarrow{D} (0 \ \ _k)$ *as* $(\quad) \to \infty$

Assumption B1(i) specifies moment conditions to identify $\ ^0_i$ B1(ii) is a high level condition. Its first part can be verified by applying Donsker's theorem. For example, if there exists $\mathcal{F}_{it}$ a -field, such that $\{\ _{it} \ \mathcal{F}_{it}\}$ is a stationary ergodic adapted mixingale with size -1 (e.g., White 2001, pp. 124-125), and $\mathrm{Var}(\ '\bar{\ }_{i,T}(\ _i)) \to \ '\Sigma_i \quad \in (0 \ \infty)$ as $\ \to \infty$ for some $\Sigma_i \quad 0$ and any nonrandom $\ \in \mathbb{R}^d$ with $\| \ \| = 1$ then $\bar{\ }_{i,T}(\ _i) \xrightarrow{D} (0 \ \Sigma_i)$ and the first part of B1(ii) follows. The second and third parts of B1(ii) can be verified by the Markov inequality and the application of Lemma S1.2(iii) in the supplement under strong

[2]We were unable to establish asymptotic theory for the case where the criterion $Q_{2,NT}$ is replaced by the fully pooled criterion $Q_{2,NT}$ $\frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} z_{it} \ \Delta y_{it} - \beta^0_i \Delta x_{it} \ ^0 W_{NT} \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} z_{it} \ \Delta y_{it} - \beta^0_i \Delta x_{it}$, where W_{NT} is asymptotically nonsingular. We also found that Arellano and Bond (1991) GMM estimation is not applicable to handle unobserved slope heterogeneity. Noticing this, Fernández-Val and Lee (2013) used a criterion similar to $Q_{2,NT}$ in the nonlinear panel setup. As we shall see, the use of $Q_{2,NT}$ means that the PGMM estimator generally does not have the oracle property.

mixing conditions. B1(iii) provides a rank condition to identify ^0_i B1(iv) is automatically satisfied for $_{iNT} = \ _d$ the $\times$ identity matrix. B1(v)-(vi) and B2(i) parallel A1(vi)-(vii) and A2(i). B2(ii) holds true by Lemma S1.2 in the supplement if $\{(\ _{it} \ \Delta \ _{it}) \quad \geq 1\}$ is strong mixing with geometric decay rate and $_{it}\Delta \ _{it}$ has six plus moments.

B3(i)-(ii) can be verified under various primitive conditions. For example, if (a) $\mathbb{E}\| \ _{it}(\Delta \ _{it})'\|^{2+\sigma} \quad 0$ for some $\quad 0$ (b) $\{(\Delta \ _{it} \ _{it} \ \Delta \ _{it}) \quad \geq 1\}$ is strong mixing for each with mixing coefficients $_i(\ )$ that satisfy $\frac{1}{N_k} \sum_{i\in G_k^0} \sum_{\tau=1}^{\infty} \ _i(\ )^{(2+\sigma)/\sigma} \quad \infty$ (c) $\{(\Delta \ _{it} \ _{it})\}$ is stationary along the time dimension and IID along the individual dimension for all $\in \ ^0_k$, and (d) $_i = \quad \forall \ \in \ ^0_k$ then B3(i) is satisfied with $_k = \{\mathbb{E}[\ _{it}(\Delta \ _{it})']\}' \quad \mathbb{E}[\ _{it}(\Delta \ _{it})'] \ \forall \ \in \ ^0_k$ To verify B3(ii), for simplicity we assume that $_{iNT} = \ _d$ and make the following decomposition

$$
\begin{aligned}
& \frac{1}{\sqrt{\ _k}} \sum_{i\in G_k^0} \tilde{\ }'_{i,z\Delta x} \sum_{t=1}^{T} \ _{it}\Delta \ _{it} \\
= \ & \frac{1}{\ _k^{1/2} \ ^{3/2}} \sum_{i\in G_k^0} \sum_{s=1}^{T} \sum_{t=1}^{T} \mathbb{E}(\Delta \ _{is} \ '_{is} \ _{it}\Delta \ _{it}) + \frac{1}{\ _k^{1/2} \ ^{3/2}} \sum_{i\in G_k^0} \sum_{s=1}^{T} \sum_{t=1}^{T} \mathbb{E}(\Delta \ _{is} \ '_{is}) \ _{it}\Delta \ _{it} \\
& + \frac{1}{\ _k^{1/2} \ ^{3/2}} \sum_{i\in G_k^0} \sum_{s=1}^{T} \sum_{t=1}^{T} \{[\Delta \ _{is} \ '_{is} - \mathbb{E}(\Delta \ _{is} \ '_{is})] \ _{it}\Delta \ _{it} - \mathbb{E}(\Delta \ _{is} \ '_{is} \ _{it}\Delta \ _{it})\} \\
\equiv \ & \ _{kNT} + \ _{kNT} + \ _{kNT} \text{ say,}
\end{aligned}
\tag{3.3}
$$

where $_{kNT}$ and $_{kNT}$ contribute to the asymptotic bias and variance, respectively, and $_{kNT}$ is a term that is asymptotically negligible under suitable conditions. Then B3(ii) is satisfied with $_{iNT} = \ _d$ if $_{kNT} = \frac{1}{N_k^{1/2}T^{1/2}} \sum_{i\in G_k^0} \sum_{t=1}^{T} \bar{\ }'_{i,z\Delta x} \ _{it}\Delta \ _{it} \xrightarrow{D} (0 \quad _k)$ and $_{kNT} = \ _P(1)$ both of which can be verified by strengthening the conditions given in (a)-(c) above. Note that $\bar{\ }^{-1}_k \ _{kNT}$ signifies the asymptotic bias of $\tilde{\ }_k$ which may not vanish asymptotically but can be corrected; see Section S2.2 in the supplement.[3]

We first establish the preliminary consistency rate of $(\tilde{\boldsymbol{\beta}}, \tilde{\boldsymbol{\alpha}})$.

Suppose Assumption B1 holds and $_2 = \quad (1)$ *Then (i)* $\tilde{\ }_i - \ ^0_i = \ _P(\ ^{-1/2} + \ _2)$ *for* $= 1$ *(ii)* $\frac{1}{N} \sum_{i=1}^{N} \| \tilde{\ }_i - \ ^0_i \|^2 = \ _P(\ ^{-1})$ *and (iii)* $(\tilde{\ }_{(1)} \quad \tilde{\ }_{(K_0)}) - (\ ^0_1 \quad \ ^0_{K_0}) = \ _P(\ ^{-1/2})$ *where* $(\tilde{\ }_{(1)} \quad \tilde{\ }_{(K_0)})$ *is a suitable permutation of* $(\tilde{\ }_1 \quad \tilde{\ }_{K_0})$

Remark 1 applies here with obvious modifications. As before, hereafter we simply write $\tilde{\ }_k$ for $\tilde{\ }_{(k)}$ as the consistent estimator of ^0_k and define $\tilde{\ }_k \equiv \{ \ \in \{1 \ 2 \quad \} : \tilde{\ }_i = \tilde{\ }_k\}$ for $= 1 \quad _0$

[3] If Conditions (a)-(b) are satisfied and $E\|z_{it}\Delta\varepsilon_{it}\|^{2+\sigma} \ > \quad$, by the Davydov inequality, we have $\|B_{kNT}\| \leq \frac{1}{T\sqrt{N_k T}} \sum_{i\in G_k^0} \sum_{t=1}^{T} \sum_{s=1}^{T} \|E[\Delta x_{is} z'_{is} z_{it}\Delta\varepsilon_{it}]\| \quad O(\ N/T \ ^{1/2})$, which is o if $T \gg N$ and usually asymptotically non-negligible otherwise.

3.3.2 Classification Consistency

Let $\hat{E}_{kNT,i} = \{i \notin \hat{G}_k \mid i \in G_k^0\}$ and $\hat{F}_{kNT,i} = \{i \notin G_k^0 \mid i \in \hat{G}_k\}$ for $i = 1, \ldots, N$ and $k = 1, \ldots, K_0$. Let $\hat{E}_{kNT} = \cup_{i \in G_k^0} \hat{E}_{kNT,i}$ and $\hat{F}_{kNT} = \cup_{i \in \hat{G}_k} \hat{F}_{kNT,i}$. We establish uniform classification consistency in the next theorem.

Theorem 3.2 Suppose that Assumptions B1-B2 hold. Then (i) $P(\cup_{k=1}^{K_0} \hat{E}_{kNT}) \le \sum_{k=1}^{K_0} P(\hat{E}_{kNT}) \to 0$ as $(N, T) \to \infty$, and (ii) $P(\cup_{k=1}^{K_0} \hat{F}_{kNT}) \le \sum_{k=1}^{K_0} P(\hat{F}_{kNT}) \to 0$ as $(N, T) \to \infty$.

REMARK 8. Remark 2 also holds for the above theorem with obvious modifications. Let $\hat{G}_0 = \{1, 2, \ldots, N\} \setminus (\cup_{k=1}^{K_0} \hat{G}_k)$ and $\hat{E}_{iNT} = \{i \in \hat{G}_0\}$. Theorem 3.2(i) implies that $P(\cup_{1 \le i \le N} \hat{E}_{iNT}) \le \sum_{k=1}^{K_0} P(\hat{E}_{kNT}) \to 0$, meaning that all individuals are classified into one of the K_0 groups w.p.a.1.

Let $\hat{N}_k = \sum_{i=1}^{N} 1\{i \in \hat{G}_k\}$. The following corollary parallels Corollary 2.3.

Corollary 3.3 Suppose that Assumptions B1-B2 hold. Then $\hat{N}_k - N_k = o_P(1)$.

3.3.3 Improved Convergence and Asymptotic Properties of Post-Lasso

The following theorem establishes the asymptotic distribution of the C-Lasso estimators $\{\hat{\alpha}_k\}$.

Theorem 3.4 Suppose Assumptions B1-B3 hold. Then $\sqrt{N_k T}(\hat{\alpha}_k - \alpha_k^0) \ldots \to N(0, \ldots)$ [illegible] for $k = 1, \ldots, K_0$.

REMARK 9. In contrast to the PPL case, the PGMM estimators $\{\hat{\alpha}_k\}$ may fail to [illegible] the oracle property. If the group identities were known in advance, one could obtain the GMM es[illegible] of α_k^0

Suppose Assumptions B1-B4 hold. Then $\sqrt{\ \ k}(\tilde{\ }_{\tilde{G}_k} - \ ^0_k) \xrightarrow{D} (0\ \Omega_k)$ *where* $\Omega_k = \ {}^{(k)\prime}_{z\Delta x}\ {}^{(k)}\ {}^{(k)}_{z\Delta x}\ {}^{-1}\ {}^{(k)\prime}_{z\Delta x}\ {}^{(k)}\ {}_k\ {}^{(k)}\ {}^{(k)}_{z\Delta x}\ \ {}^{(k)\prime}_{z\Delta x}\ {}^{(k)}\ {}^{(k)}_{z\Delta x}\ {}^{-1}$ *and* $= 1 \quad {}_0$

To prove the above theorem, we first apply Theorem 3.2 and show that $\sqrt{\ \ k}(\tilde{\ }_{\tilde{G}_k} - \ ^0_k) = \sqrt{\ \ k}(\tilde{\ }_{G^0_k} - \ ^0_k) + \ _P(1)$ That is, the post-Lasso GMM estimator $\tilde{\ }_{\tilde{G}_k}$ is asymptotically equivalent to the oracle estimator $\tilde{\ }_{G^0_k}$. To obtain the most efficient estimator among the class of GMM estimators based on the moment conditions specified in Assumption B1(i), one can set ${}^{(k)}_{NT}$ to be a consistent estimator of ${}_k^{-1}$ Alternatively, we can consider Arellano and Bond (1991) GMM estimation based on the estimated groups. The procedure is standard and details are omitted.

If ${}_{iNT} = {}^{(k)}_{NT}\ \bar{\ }_{i,z\Delta x} = {}^{(k)}_{z\Delta x}$ for each $\in {}^0_k$ in Assumption B3(i) (which is unrealistic before knowing the group identity), and ${}_{kNT} = 0$ in Assumption B3(ii), then ${}_k = {}^{(k)\prime}_{z\Delta x}\ {}^{(k)}\ {}^{(k)}_{z\Delta x}$, ${}_k = {}^{(k)\prime}_{z\Delta x}\ {}^{(k)}\Omega_k\ {}^{(k)}\ {}^{(k)}_{z\Delta x}$ and $\sqrt{\ \ k}(\tilde{\ }_k - \ ^0_k) \xrightarrow{D} (0\ \Omega_k)$ Thus in this special.55(p)-2s 75us

Table 1: Frequency of selecting $= 1$ 5 groups when $_0 = 3$

N	T	DGP 1					DGP 2					DGP 3				
		1	2		4	5	1	2		4	5	1	2		4	5
100	15	0	0		0.004	0.002	0	0.232		0.004	0.002					
100	25	0	0		0	0	0	0.016		0	0	0	0.096		0.242	0.016
100	50	0	0		0	0	0	0		0	0	0	0		0.014	0
200	15	0	0		0.106	0.004	0	0.022		0.008	0					
200	25	0	0		0	0	0	0		0	0	0	0.106		0.226	0
200	50	0	0		0	0	0	0		0	0	0	0	1	0	0

Next, given the true number of groups, we focus on the classification of individual units and the point estimation of post-Lasso.[4] Due to space limitation, all tabulated results are produced under $\lambda_j = 0\ 5$ $= 1\ 2$, for the linear models, and $\lambda_1 = 0\ 05$ for the Probit model. The outcomes are found robust over the specified range of constants. Column 4 of Tables 2 shows the percentage of correct classification of the units, calculated as $\frac{1}{N} \sum_{k=1}^{K_0} \sum_{i \in \hat{G}_k} \mathbf{1}\{ {}^0_i = {}^0_k \}$, averaged over the Monte Carlo replications. Columns 5–7 summarize the post-Lasso estimator's roo

Table 2: Classification and Point Estimation of $\boldsymbol{\alpha}_1$

			% of correct	Post-Lasso			Oracle		
	N	T	classification	RMSE	Bias	Coverage	RMSE	Bias	Coverage
DGP 1	100	15	0.8935	0.0594	0.0105	0.8758	0.0463	0.0012	0.9336
	100	25	0.9674	0.0384	0.0018	0.9344	0.0353	0.0001	0.9362
	100	50	0.9964	0.0249	0.0000	0.9528	0.0245	-0.0002	0.9348
	200	15	0.8987	0.0432	0.0077	0.8650	0.0324	-0.0013	0.9410
	200	25	0.9661	0.0272	0.0015	0.9228	0.0250	-0.0006	0.9394
	200	50	0.9966	0.0174	-0.0001	0.9496	0.0171	-0.0002	0.9424
DGP 2	100	15	0.8063	0.0711	-0.0123	0.9562	0.0502	-0.0037	0.9090
	100	25	0.8974	0.0461	-0.0060	0.9760	0.0351	0.0011	0.9336
	100	50	0.9689	0.0278	-0.0011	0.9860	0.0242	-0.0010	0.9320
	200	15	0.8151	0.0557	-0.0159	0.9436	0.0352	-0.0017	0.9308
	200	25	0.9037	0.0328	-0.0047	0.9664	0.0252	-0.0006	0.9442
	200	50	0.9711	0.0193	-0.0014	0.9842	0.0164	0.0000	0.9304
DGP 3	100	25	0.7941	0.1701	0.0805	0.7856	0.1077	0.0114	0.9376
	100	50	0.9456	0.0859	0.0231	0.8970	0.0752	0.0090	0.9504
	200	25	0.8277	0.1325	0.0777	0.7214	0.0821	0.0116	0.9104
	200	50	0.9527	0.0635	0.0223	0.8818	0.0573	0.0121	0.9280

via a dynamic Probit model. Due to space limitation, we only report the estimated coefficients in the main text. Summary statistics, group membership, and additional details of implementation can be found in the Supplementary Material.

fi

Understanding the disparate savings behavior across countries is a longstanding research interest in development economics. Theoretical advances and empirical studies have accumulated over many years; see Feldstein (1980), Deaton (1990), Edwards (1996) Bosworth, Collins, and Reinhart (1999), Rodrik (2000), and Li, Zhang, and Zhang (2007), among many others. Empirical research in this area typically employs standard panel data methods to handle heterogeneity or relies on prior information to categorize countries into groups. Classification criteria vary from geographic locations to the notion of developed countries versus developing countries (Loayza, Schmidt-Hebbel and Servén, 2000). This section applies the methodology developed in the present paper to revisit this empirical problem.

Following Edwards (1996), we consider the simple regression model

$$ {}_{it} = {}_{1i}\ {}_{i,t-1} + {}_{2i}\ {}_{it} + {}_{3i}\ {}_{it} + {}_{4i}\ {}_{it} + {}_{i} + {}_{it} \tag{5.1} $$

where $_{it}$ is the ratio of savings to GDP, $_{it}$ is the CPI-based inflation rate, $_{it}$ is the real interest rate, $_{it}$ is the per capita GDP growth rate, $_{i}$ is a fixed effect, and $_{it}$ is an idiosyncratic error term. Inflation characterizes the degree of the macroeconomic stability and the real interest rate reflects the price of money. The relationship between the savings rate and GDP growth rate is well documented, with the latter being found to Granger-cause the former (Carroll and Weil, 1994). The first-order lagged savings rate is added to the specification to capture persistence of the savings rate.

Data are obtained from the widely used World Development Indicators, a comprehensive dataset compiled by the World Bank. For many countries the time series of real interest rates are often short in comparison with the other variables. Using the time span 1995–2010, we were able to construct a balanced panel of 56 countries. Substantial heterogeneity across countries was observed in all these major macroeconomic indicators. Evidence of within group homogeneity is therefore particularly important in supporting panel data pooling techniques.

This dynamic panel model can be estimated by either PLS or PGMM. We first try PLS, which has higher correct classification ratio in our simulation when $= 15$. Following the simulation, $_{1NT}$ is set as $\frac{2}{3}(\quad)^{-1/2}$, and the IC picks two groups and the tuning parameter constant $_{\lambda_1} = 1\ 55$ over all combinations of $= 1 \quad 5$ and $_{\lambda_1}$ in a geometrically increasing sequence of 10 points in (0 2 2). Based on this choice of tuning parameter, the data determine the group identities. Interestingly, some geographic features remain salient in the classification. For example, we observe a strong collection of Asian countries in Group 1. In particular, except for South Korea and the city state Singapore, Group 1 includes all Eastern Asian and Southeastern Asian countries in our sample, namely, China, Japan, Indonesia, Malaysia, Philippines, and Thailand.

Table 3: PLS and PGMM estimation results

Variables		PLS			PGMM	
	Pooled FE	Group1	Group2	Pooled GMM	Group1	Group2
Lagged savings	. ***	. ***	. ***	.	.	. **
	.	.	.	.	.	.
Inflation	− .	− . ***	. ***	.	− . **	. ***
	.	.	.	.	.	.
Interest rate	− .	− . ***	. ***	− .	− . **	. *
	.	.	.	.	.	.
GDP growth	. ***	. ***	. **	. ***	. ***	. **
	.	.	.	.	.	.

Note: *** 1% significant, ** 5% significant, * 10% significant

Columns 3–4 in Table 3 report the results for the PLS-based post-Lasso estimation, in comparison with those for the pooled FE estimation in Column 2. The estimates are bias-corrected by the half-panel jackknife (Dhaene and Jochmans, 2015), and the standard errors (in parentheses) are clustered at the country level. Compared with Edwards (1996), the FE results re-confirm the significance of lagged savings and GDP growth rate as well as the insignificance of inflation and interest rates in the determination of savings rate. This result also lends support to the *conventional wisdom* estimated group identities reveal 84% overlap with the PLS classified membership, and the coefficients in columns 6–7 of Table 3 are comparable to those from PLS.

ﬂ

According to a conservative estimate, direct casualties from civil conflicts were at million in the second half of the tentieth century, a figure five times as large as the inter-state tollFearon and Laitin, 2003). Civil war damage to national devlopment has at tracted interest among economists and political scientists, looking at both causes and consequences, andng to an explosion research (Miguel, Satyanath, and Sergenti, 2004; Besley and Persson, 2010; Nunn and Qian, 2014). A comprehensive overview was given in Blattman and Miguel (2010).

This section revisits the connection between civil wars and poverty, a topic of enduring research interest.[5] Cross-country empirical work mostly follows Fearon and Laitin (2003) and Collier and Hoeffler (2004) in regressing war onset or incidence against posited causes civil con flict. Country-specific heterogeneity is handled either by control variables or fixed effects. In view of the m5asurement error in many macro variables and the difficulty in exhausting all relevant factors, Djankov and Reynal-Quer effect approach in linear regressions. Group-specific heterogeneity was also i groups using observed information relating to former families (Djankov and R or continental regions (Esteban, Mayoral, and Ray, 2012). Without such infor unobservable country- and group-specific heterogeneity simultaneously in a nor

We use the replication data in Fearon and Laitin (2003). Since most of the time-invariant country characteristics, we collect *civil war incidence*, *GDPer ca* generate a balanced panel of 38 countries and 39 years spanning from 1960 to [6] We specify a panel AR(1) Probit model as in DGP 3 to capture the high persistence of civil war incidence, and we transform *GDPer capita* and *log* into growth rates to avoid non-stationarity.

The IC with $_{1NT} = \frac{1}{4}(\ln\ln\ )$ selects two groups and the tuning parameter constant $_{\lambda_1} = 0\ 046$ from all the combinations of $= 1 \quad 5$ and $_{\lambda_1}$ from 10 points in the geometrically increasing sequence $(0\ 01 \quad 0\ 1)$. C-Lasso classifies 23 countries into a "high-occurrence" group (with mean civil war incidence 0.4302), and the other 15 countries into a "low-occurrence" group (with mean incidence 0.2263). In terms of geographic features, Iran and Jordan are separated from all the other 12 Asian countries, most of which are plagued by civil wars; the four included Europ5an countries (Cyprus, Russia, UK, Yugoslavia) all fall into the low-occurrence group.

Table 4 displays the estimated PPL coefficients along with those for standard Probit and FE Probit regressions. Again, the estimates are bias-corrected by the half-panel jackk are clustered at the countryel. Obviously, civil war incidence is highlyersist with GDP per capita growth remains robust in Probit and FE Probit regr are distinguished in the two grouBey Aga0-336.10ID-.00gati.4(d)-376.25.9(tr6 group, but no such relationship is found in the high-occurrence group.

Table 4: Probit, FE Probit and PPL estimation results

Variables	Probit		FE Probit		Post-Lasso PPL			
					high-occurrence		low-occurrence	
	coef.	s.e.	coef.	s.e.	coef.	s.e.	coef.	s.e.
Lagged civil war	. ***	0.1156	. ***	0.1140	. ***	0.1363	. ***	0.2707
GDP per capita growth	− . ***	0.1155	− . ***	0.1389	.	0.1193	− . ***	0.2220
population growth	− .	0.1107	.	0.1284	− .	0.1429	.	0.1736

Note: *** 1% significant, ** 5% significant, * 10% significant

We propose a novel and systematic approach to identify and estimate latent group structures in panel data, developing panel penalized profile likelihood (PPL) and panel GMM (PGMM) methods for classification and estimation, and providing asymptotic properties for use in inference. The PPL method enjoys the oracle property but PGMM typically does not. Post-Lasso estimates are also studied and a BIC-type information criterion is proposed to determine the number of groups. These techniques combine to provide a general approach to classifying and estimating panel models with unknown homogeneous groups, heterogeneity across groups, and an unknown number of groups. Simulations show that the approach has good finite sample performance and can be readily implemented in practical work. Two applications reveal the advantages of data-determined identification of latent group structures in empirical panel modeling.

The present work raises interesting issues for further research. First, it may be appealing to consider a more general framework that allows the number ($_0$) of groups to grow with the sample size. Close examination of the theory provided in this paper suggests that it is possible to permit $_0$ to increases with but at a very slow rate. Second, both the linear and nonlinear models may be extended to include time effects or interactive fixed effects (IFE). In linear models with IFE but without endogeneity, we remark that the present approach can be used in conjunction with principal component analysis to address cross sectional dependence modeled through IFE. Extension to nonlinear models or to models with endogeneity will raise new statistical and computational challenges. Third, our method can be extended to nonstationary panels where panel unit and cointegrating relationships may possess latent group structures. Some of these topics will be explored in future work.

APPENDIX

. (i) Let $_{1NT,i}(\ _i) = \frac{1}{T}\ \ _{t=1}^{T}\ (\ _{it};\ _i\ \hat{\ }_i(\ _i))$ and $^{(K_0)}_{1iNT,\lambda_1}(\ _i\ \boldsymbol{\alpha}) = \ _{1NT,i}(\ _i)$ $+\ _1\Pi_{k=1}^{K_0}\|\ _i - \ _k\|$ Let $_i = \ _i - \ _i^0$ and $\hat{\ }_i = \hat{\ }_i - \ _i^0$ Since $\hat{\ }_i(\ _i) = \arg\min_{\mu_i} \frac{1}{T}\ \ _{t=1}^{T}\ (\ _{it};\ _i\ \ _i)$ we have $\frac{1}{T}\ \ _{t=1}^{T}\ _i(\ _{it};\ _i\ \hat{\ }_i(\ _i)) = 0\ \forall\ \ _i$ Then by second order Taylor expansion and the envelope theorem, we have

$$
\begin{aligned}
_{1NT,i}(\hat{\ }_i) - \ _{1NT,i}(\ _i^0) &= \frac{1}{T}\ \sum_{t=1}^{T}\ (\ _{it};\hat{\ }_i\ \hat{\ }_i(\hat{\ }_i)) - \frac{1}{T}\ \sum_{t=1}^{T}\ (\ _{it};\ _i^0\ \hat{\ }_i(\ _i^0)) \\
&= \hat{\ }_i'\ \hat{\ }_i + \frac{1}{2}\hat{\ }_i'\ \hat{\ }_{i\beta\beta}(\breve{\ }_i)\hat{\ }_i \qquad \text{(A.1)}
\end{aligned}
$$

where $\breve{\ }_i$

of the property of $\hat{\boldsymbol{\alpha}}$ By (A.1) and the Cauchy-Schwarz inequality

$$\Box \quad {}^{(K_0)}_{1NT,\lambda_1} \left(\boldsymbol{\beta}^0 + \quad {}^{-1/2} \mathsf{V} \ \hat{\boldsymbol{\alpha}} \right) - \quad {}^{(K_0)}_{1NT,\lambda_1} \left(\boldsymbol{\beta}^0 \ \boldsymbol{\alpha}^0 \right) \Box$$

$$= \frac{1}{2} \Box^{N}_{i=1} \ {}'_{i} \ \hat{}\ {}_{i\beta\beta} \left(\ \breve{}_{i} \right) \quad {}_{i} + \frac{\sqrt{\quad}}{} \Box^{N}_{i=1} \ {}'_{i} \ \hat{}\ {}_{i} + \frac{1}{} \Box^{N}_{i=1} \Pi^{K_0}_{k=1} \left\| \ \hat{}\ {}_{i} - \ \hat{}\ {}_{k} \right\|$$

envelope theorem, the first order condition (with respect to $_i$) for the minimization problem in (2.5) yields that

$$
\begin{aligned}
\mathbf{0}_{p\times 1} &= \frac{1}{\sqrt{\ }}\sum_{t=1}^{T} {}_i(\ {}_{it};\hat{\ }_i\ \hat{\ }_i(\hat{\ }_i)) + \sqrt{\ }\ {}_1\sum_{j=1}^{K_0}\hat{\ }_{ij}\Pi_{l=1,l\neq j}^{K_0}\left\|\hat{\ }_i-\hat{\ }_l\right\| \\
&= \frac{1}{\sqrt{\ }}\sum_{t=1}^{T} {}_{it} + \left(\frac{{}_1\hat{\ }_{ki}}{\left\|\hat{\ }_i-\hat{\ }_k\right\|}\ {}_p + \bar{\ }_{i\beta\beta}\right)\sqrt{\ }\left(\hat{\ }_i-\hat{\ }_k\right) + \frac{1}{\sqrt{\ }}\sum_{t=1}^{T}\left[{}_i(\ {}_{it};\ {}^0_i\ \hat{\ }_i(\ {}^0_i)) - {}_{it}\right] \\
&\quad + \bar{\ }_{i\beta\beta}\sqrt{\ }(\hat{\ }_k - {}^0_k) + \sqrt{\ }\ {}_1\sum_{j=1,j\neq k}^{K_0}\hat{\ }_{ij}\Pi_{l=1,l\neq j}^{K_0}\left\|\hat{\ }_i-\hat{\ }_l\right\| \\
&\equiv \hat{\ }_{i1}+\hat{\ }_{i2}+\hat{\ }_{i3}+\hat{\ }_{i4}+\hat{\ }_{i5} \qquad \text{(A.9)}
\end{aligned}
$$

where $\hat{\ }_{ij}=\frac{\hat{\beta}_i-\hat{\alpha}_j}{\|\hat{\beta}_i-\hat{\alpha}_j\|}$ if $\|\hat{\ }_i-\hat{\ }_j\|\neq 0$ and $\|\hat{\ }_{ij}\|\leq 1$ otherwise, the second equality follows from the first order Taylor expansion and rearrangement of terms, $\bar{\ }_{i\beta\beta}\equiv\hat{\ }_{i\beta\beta}(\bar{\ }_i)$ $\hat{\ }_{i\beta\beta}(\cdot)$ is defined in (A.2), $\bar{\ }_i$ lies between $\hat{\ }_i$ and ${}^0_i$ elementwise.

Let $\varkappa_{1NT}=(\ {}^{-1/2}(\ln\ )^3+\ {}_1)(\ln\ )^{\nu}$ Let denote a generic constant that may vary across lines. By (A.4) and Lemmas S1.6-S1.7 in the Supplement, we can readily show that

$$
\left(\max_i\left\|\hat{\ }_i-{}^0_i\right\|\geq\ \varkappa_{1NT}\right)=(\ {}^{-1})\ \text{ for some }\ 0 \qquad \text{(A.10)}
$$

which in conjunction with the proof of Theorem 2.1(iii), implies that

$$
\left(\sqrt{\ }\left\|\hat{\ }_k-{}^0_k\right\|\geq(\ln\ )^{\nu}\right)=(\ {}^{-1})\ \text{ and }\ \left(\max_{i\in G_k^0}\left|\hat{\ }_{ki}-{}^0_k\right|\geq{}^0_k\ 2\right)=(\ {}^{-1}) \qquad \text{(A.11)}
$$

By (A.10)-(A.11), $\left(\max_{i\in G_k^0}\left\|\hat{\ }_{i5}\right\|\geq\sqrt{\ }\ {}_1\varkappa_{1NT}\right)=(\ {}^{-1})$ Combining these results with those in Lemmas S1.6(v) and S1.11(i), we have $(\Xi_{kNT})=1-(\ {}^{-1})$ where

$$
\begin{aligned}
\Xi_{kNT} \equiv& \left\{\max_{i\in G_k^0}\left|\hat{\ }_{ki}-{}^0_k\right|\leq{}^0_k\ 2\right\}\cap\left\{\max_{i\in G_k^0}\left\|\bar{\ }_{i\beta\beta}-\ {}_{i\beta\beta}(\ {}^0_i)\right\|\leq\ {}_H\ 2\right\}\cap\left\{\max_{i\in G_k^0}\left\|\hat{\ }_{i3}\right\|\leq(\ln\ )^{3+\nu}\right\} \\
&\cap\left\{\max_{i\in G_k^0}\left\|\hat{\ }_{i4}\right\|\leq(\ln\ )^{\nu}\right\}\cap\left\{\max_{i\in G_k^0}\left\|\hat{\ }_{i5}\right\|\leq\sqrt{\ }\ {}_1\varkappa_{1NT}\right\}
\end{aligned}
$$

Then conditional on Ξ_{kNT} we have uniformly in $\in{}^0_k$

$$
\begin{aligned}
\left\|(\hat{\ }_i-\hat{\ }_k)'(\hat{\ }_{i2}+\hat{\ }_{i3}+\hat{\ }_{i4}+\hat{\ }_{i5})\right\| &\geq \left\|(\hat{\ }_i-\hat{\ }_k)'\hat{\ }_{i2}\right\|-\left\|(\hat{\ }_i-\hat{\ }_k)'(\hat{\ }_{i3}+\hat{\ }_{i4}+\hat{\ }_{i5})\right\| \\
&\geq \sqrt{\ }\ {}_1\hat{\ }_{ki}\left\|\hat{\ }_i-\hat{\ }_k\right\|-\left\|\hat{\ }_i-\hat{\ }_k\right\|\left[2(\ln\ )^{3+\nu}+\sqrt{\ }\ {}_1\varkappa_{1NT}\right] \\
&\geq \sqrt{\ }\ {}_1\ {}^0_k\left\|\hat{\ }_i-\hat{\ }_k\right\|\ 4\ \text{ for sufficiently large }(\ )
\end{aligned}
$$

where the last inequality follows because $\sqrt{\ }\ {}_1\gg 2(\ln\ )^{3+\nu}+\sqrt{\ }\ {}_1\varkappa_{1NT}$ by Assumption A2(i). Then for all $\in{}^0_k$ we have

$$
\begin{aligned}
(\hat{\ }_{kNT,i}) &= \left(\ \in\hat{\ }_k\mid\ \in{}^0_k\right)=\left(-\hat{\ }_{i1}=\hat{\ }_{i2}+\hat{\ }_{i3}+\hat{\ }_{i4}+\hat{\ }_{i5}\right) \\
&\leq \left(\left|(\hat{\ }_i-\hat{\ }_k)'\hat{\ }_{i1}\right|\geq\left|(\hat{\ }_i-\hat{\ }_k)'(\hat{\ }_{i2}+\hat{\ }_{i3}+\hat{\ }_{i4}+\hat{\ }_{i5})\right|\right) \\
&\leq \left(\left\|\hat{\ }_{i1}\right\|\geq\sqrt{\ }\ {}_1\ {}^0_k\ 4\ \ \Xi_{kNT}\right)+(\Xi_{kNT}^c)\to 0\ \text{as}\ (\ )\to\infty
\end{aligned}
$$

where Ξ_{kNT}^c denotes the complement of Ξ_{kNT} and the convergence follows by Lemma S1.6(iv) and Assumption A2 . Consequently, we conclude that with probability $1 - \quad (\quad ^{-1})$ the difference $\hat{}_i - \hat{}_k$ must reach the point where $\| \ _i - \ _k\|$ is not differentiable with respect to $\ _i$ for any $\ \in \ ^0_k$. That is $\ (\| \hat{}_i - \hat{}_k \| = 0 \mid \ \in \ ^0_k) =$ $1 - \quad (\quad ^{-1})$

For uniform consistency, we have: $(\cup_{k=1}^{K_0} \hat{}_{kNT}) \leq \sum_{k=1}^{K_0} \ (\hat{}_{kNT}) \leq \sum_{k=1}^{K_0} \sum_{i\in G_k^0} \ (\hat{}_{kNT,i})$ and by Lemma S1.6(iv)

$$\sum_{k=1}^{K_0} \sum_{i\in G_k^0} \ (\hat{}_{kNT,i}) \leq \sum_{k=1}^{K_0} \sum_{i\in G_k^0} \ (\| \hat{}_{i1} \| \geq \sqrt{\ } \ _1 \ ^0_k \ 4 \quad \Xi_{kNT}) + \ (\Xi_{kNT}^c)$$

$$\leq \max_{1\leq i\leq N} \ \left(\left\| \frac{1}{\ } \sum_{t=1}^{T} \ _{it} \right\| \geq \ _1 \ ^{K_0-1}_{\alpha} \ 4\right) + \ (1) = \ (1) \qquad \text{(A.12)}$$

This completes the proof of (i).

(ii) Pretending each individual's membership is random, we have $\ (\ \in \ ^0_k) = \ _k \ \to \ _k \in (0 \ 1)$ for $\ = 1 \quad \ _0$ and can interpret previous results as conditional on the group membership assignment. By Bayes theorem,

$$\begin{aligned} \ (\hat{}_{kNT,i}) &= 1 - \ (\ \in \ ^0_k \mid \ \in \hat{}_k) \\ &= \frac{\sum_{l=1,l\neq k}^{K_0} \ (\ \in \hat{}_k \mid \ \in \ ^0_l) \ (\ \in \ ^0_l)}{\ (\ \in \hat{}_k \mid \ \in \ ^0_k) \ (\ \in \ ^0_k) + \sum_{l=1,l\neq k}^{K_0} \ (\ \in \hat{}_k \mid \ \in \ ^0_l) \ (\ \in \ ^0_l)} \end{aligned} \qquad \text{(A.13)}$$

For the numerator, we have by (A.12)

$$\sum_{l=1,l\neq k}^{K_0} \sum_{i\in \hat{G}_k} \ (\ \in \hat{}_k \mid \ \in \ ^0_l) \ (\ \in \ ^0_l) \leq (\ _0 - 1) \sum_{l=1}^{K_0} \sum_{i\in G_l^0} \ (\ \in \hat{}_l \mid \ \in \ ^0_l) = \ (1)$$

In addition, noting that $\ (\ \in \hat{}_k \mid \ \in \ ^0_k) = 1 - \ (\ \in \hat{}_k \mid \ \in \ ^0_k) = 1 - \ (1)$ uniformly in and by (i), we have that $\ (\ \in \hat{}_k \mid \ \in \ ^0_k) \ (\ \in \ ^0_k) + \sum_{l=1,l\neq k}^{K_0} \ (\ \in \hat{}_k \mid \ \in \ ^0_l) \ (\ \in \ ^0_l) \geq \ (\ \in \ ^0_k) \ 2$ w.p.a.1. It follows that

$$\begin{aligned} \ (\cup_{k=1}^{K_0} \hat{}_{kNT}) &\leq \sum_{k=1}^{K_0} \sum_{i\in \hat{G}_k} \ (\hat{}_{kNT,i}) \leq \frac{\sum_{l=1,l\neq k}^{K_0} \sum_{i\in \hat{G}_k} \ (\ \in \hat{}_k \mid \ \in \ ^0_l) \ (\ \in \ ^0_l)}{\min_{1\leq i\leq N} \min_{1\leq k\leq K_0} \ (\ \in \ ^0_k) \ 2} \\ &= \frac{\ (1)}{\min_{1\leq k\leq K_0} \ _k \ 2} = \ (1) \ \blacksquare \end{aligned}$$

. Noting that $\hat{}_k = \sum_{i=1}^{N} \mathbf{1}\{ \ \in \hat{}_k\}$, $\ _k = \sum_{i=1}^{N} \mathbf{1}\{ \ \in \ ^0_k\}$ and $\mathbf{1}\{ \ \in \hat{}_k\} - \mathbf{1}\{ \ \in \ ^0_k\} = \mathbf{1}\{ \ \in \hat{}_k \setminus \ ^0_k\} - \mathbf{1}\{ \ \in \ ^0_k \setminus \hat{}_k\}$ we have $\hat{}_k - \ _k = \sum_{i=1}^{N} [\mathbf{1}\{ \ \in \hat{}_k \setminus \ ^0_k\} - \mathbf{1}\{ \ \in \ ^0_k \setminus \hat{}_k\}]$ Then by the implication rule and the Markov inequality, for any $\quad 0$

$$\begin{aligned} \ (| \hat{}_k - \ _k | \geq 2 \ ) &\leq \ \left(\sum_{i=1}^{N} \mathbf{1}\{ \ \in \hat{}_k \setminus \ ^0_k\} \geq \ \right) + \ \left(\sum_{i=1}^{N} \mathbf{1}\{ \ \in \ ^0_k \setminus \hat{}_k\} \geq \ \right) \\ &= \frac{1}{\ } \sum_{i=1}^{N} \ (\hat{}_{kNT,i}) + \frac{1}{\ } \sum_{i=1}^{N} \ (\hat{}_{kNT,i}) \end{aligned}$$

By (A.12), $\sum_{i=1}^{N} (\hat{}_{kNT,i}) = \sum_{k=1}^{K_0} \sum_{i \in G_k^0} (\hat{}_{kNT,i}) =$ (1) By the proof of Theorem 2.2(i), $\sum_{i=1}^{N} (\hat{}_{kNT,i})$ $= \sum_{k=1}^{K_0} \sum_{i \in \hat{} P^{\hat{}}}$

where $\hat{\mathcal{R}}$

We start by proving a useful technical result and then proceed to prove the main results. Let $_{iNT}(\ _i) \equiv [\frac{1}{T} \sum_{t=1}^{T} (\ _{it}\ \ _i)]' \ _{iNT} [\frac{1}{T} \sum_{t=1}^{T} (\ _{it}\ \ _i)]$ and $\bar{\ }_i(\ _i) \equiv \{\frac{1}{T} \sum_{t=1}^{T} \mathbb{E}[\ ($

where $_{-1NT} \equiv \min_{1\le i\le N} \; _{\min}\left(\tilde{}'_{i,z\Delta x} \; _i \; \tilde{}_{i,z\Delta x}\right)$ satisfies that $\liminf_{(N,T)\to\infty} {}_{-1NT} \geq {}_{-W-\bar{Q}}^{\;\;2} \quad 0$ by Assumptions B1(iii)-(iv). Combining (B.4)-(B.6) yields

$$_{-1NT} \left\| \tilde{}_i - {}_i^0 \right\|^2 \leq \frac{2}{-} \left[2^- \; {}^0_{i,T} + _- \; \tilde{}_{i,T} + \; _2 \; \tilde{}_{K_0} \left(\left\| \tilde{}_i - {}_i^0 \right\| + 2 \left\| \tilde{}_i - {}_i^0 \right\|^2 \right) \right]$$

or, $\left(_{-1NT} - \frac{4}{c} \; _2 \; \tilde{}_{K_0}\right) \left\| \tilde{}_i - {}_i^0 \right\|^2 \leq \frac{2}{c} \; 2^- \; {}^0_{i,T} + _- \; \tilde{}_{i,T} + \; _2 \; \tilde{}_{K_0} \left\| \tilde{}_i - {}_i^0 \right\|$ where $\tilde{}_{K_0} = \; _{K_0}(\tilde{\boldsymbol{\alpha}})$ Then

$$\begin{aligned} \left\| \tilde{}_i - {}_i^0 \right\| &\leq \frac{\frac{2}{c} \; _2 \; \tilde{}_{K_0} + \left[\left(\frac{2}{c} \; _2 \; \tilde{}_{K_0} \right)^2 + \frac{8}{c} \left(_{-1NT} - \frac{4}{c} \; _2 \; \tilde{}_{K_0}\right) \left(2^- \; {}^0_{i,T} + _- \; \tilde{}_{i,T} \right) \right]^{1/2}}{2\left(_{-1NT} - \frac{4}{c} \; _2 \; \tilde{}_{K_0} \right)} \\ &= \; _P\left(\; ^{-1/2} + \; _2 \right) \end{aligned} \tag{B.7}$$

As in the proof of Theorem 2.1(ii), we can further demonstrate that $\frac{1}{N} \sum_{i=1}^N \left\| \tilde{}_i - {}_i^0 \right\|^2 = \; _P\left(\; ^{-1} \right)$

The proof of (iii) is completely analogous to that of Theorem 2.1(iii), now using the facts that $| \; _{NT}(\tilde{\boldsymbol{\beta}} \; \boldsymbol{\alpha}) - \; _{NT}\left(\boldsymbol{\beta}^0 \; \boldsymbol{\alpha}\right) | = \; _P\left(\; ^{-1/2}\right)$ and that $0 \geq \; _{NT}(\tilde{\boldsymbol{\beta}} \; \tilde{\boldsymbol{\alpha}}) - \; _{NT}(\tilde{\boldsymbol{\beta}} \; \boldsymbol{\alpha}^0)$ ■

(i) First, we fix $\; \in \{1 \;\; _0\}$ By the consistency of $\tilde{}_k$ and $\tilde{}_i$ in Theorem 3.1 and Assumptions B1(v)-(vi), we have $\tilde{}_i - \tilde{}_l \xrightarrow{P} {}^0_k - {}^0_l \neq 0$ for all $\; \in \; {}^0_k$ and $\; \neq \;$ and $\tilde{}_{ki} \equiv \Pi_{l=1,l\neq k}^{K_0} || \tilde{}_i - \tilde{}_l || \xrightarrow{P} {}^0_k \equiv \Pi_{l=1,l\neq k}^{K_0} || {}^0_k - {}^0_l || \geq \; {}^{K_0-1}_{\alpha} \quad 0$ for any $\; \in \; {}^0_k$ Now, suppose that $\left\| \tilde{}_i - \tilde{}_k \right\| \neq 0$ for some $\; \in \; {}^0_k$ Then the first order condition (with respect to $\; _i$) for the minimization problem in (3.2) implies that

$$\begin{aligned} \mathbf{0}_{p\times 1} &= -2 \; \tilde{}'_{i,z\Delta x} \; _{iNT} \frac{1}{\sqrt{\;}} \sum_{t=1}^T \; _{it} \left(\Delta \; _{it} - \tilde{}'_i \Delta \; _{it} \right) + \sqrt{\;} \; _2 \sum_{j=1}^{K_0} \tilde{}_{ij} \Pi_{l=1,l\neq j}^{K_0} \left\| \tilde{}_i - \tilde{}_l \right\| \\ &= -2 \; \tilde{}'_{i,z\Delta x} \; _{iNT} \frac{1}{\sqrt{\;}} \sum_{t=1}^T \; _{it} \Delta \; _{it} + \left\{ \frac{_2 \; \tilde{}_{ki}}{\left\| \tilde{}_i - \tilde{}_k \right\|} \; _p + 2 \; \tilde{}'_{i,z\Delta x} \; _{iNT} \; \tilde{}_{i,z\Delta x} \right\} \sqrt{\;} \left(\tilde{}_i - \tilde{}_k \right) \\ &\quad + 2 \; \tilde{}'_{i,z\Delta x} \; _{iNT} \; \tilde{}_{i,z\Delta x} \sqrt{\;} \left(\tilde{}_k - {}^0_k \right) + \sqrt{\;} \; _2 \sum_{j=1,j\neq k}^{K_0} \tilde{}_{ij} \Pi_{l=1,l\neq j}^{K_0} \left\| \tilde{}_i - \tilde{}_l \right\| \\ &\equiv -\tilde{}_{i1} + \tilde{}_{i2} + \tilde{}_{i3} + \tilde{}_{i4} \quad \text{say,} \end{aligned} \tag{B.8}$$

where $\tilde{}_{ij} = \frac{\tilde{\beta}_i - \tilde{\alpha}_j}{||\tilde{\beta}_i - \tilde{\alpha}_j||}$ if $|| \tilde{}_i - \tilde{}_j || \neq 0$ and $|| \tilde{}_{ij} || \leq 1$ if $|| \tilde{}_i - \tilde{}_j || = 0$ Following the proof of Lemma S1.7, we can show that $\left(\max_i || \tilde{}_i - {}^0_i || \geq \; \right) = \left(\; ^{-1} \right)$ for any given $\quad 0$ With this, by (B.7) and Assumptions B2(ii)-(iv), we can readily show that

$$\left(\max_i || \tilde{}_i - {}^0_i || \geq \varkappa_{2NT} \right) = \left(\; ^{-1} \right) \text{ for some } \quad 0 \tag{B.9}$$

where $\varkappa_{2NT} = (\; ^{-1/2} (\ln \;)^3 + \; _2) (\ln \;)^\nu$ This, in conjunction with the proof of Theorem 3.1(iii), implies that

$$\left(\sqrt{\;} || \tilde{}_k - {}^0_k || \geq (\ln \;)^\nu \right) = \left(\; ^{-1} \right) \text{ and } \left(\max_{i\in G^0_k} | \tilde{}_{ki} - {}^0_k | \geq {}^0_k \; 2 \right) = \left(\; ^{-1} \right) \tag{B.10}$$

By (B.9)-(B.10), $(\max_{i\in G_k^0}||\tilde{\ }_{i4}|| \geq \sqrt{\ }\ _2\varkappa_{2NT}) = \ ^{(\ \ -1)}$ By Assumptions B1(iii)-(iv), we have $(\max_{i\in G_k^0}||\tilde{\ }'_{i,z\Delta x}\ \ _{iNT}\tilde{\ }_{i,z\Delta x} - \bar{\ }'_{i,z\Delta x}\ \ _i\ \bar{\ }_{i,z\Delta x}|| \geq\ ) = \ ^{(\ \ -1)}$ for any $\ 0$ This result, in conjunction with (B.10), implies that $(\max_{i\in G_k^0}||\tilde{\ }_{i3}|| \geq (\ln\ )^{\nu}) = \ ^{(\ \ -1)}$ for some $\ 0$ It follows that $(\Gamma_{kNT}) = 1 - \ ^{(\ \ -1)}$ where

$$\Gamma_{kNT} \equiv \left\{\max_{i\in G_k^0}\left|\tilde{\ }_{ki} - \ ^0_k\right| \leq \ ^0_k\ 2\right\} \cap \left\{\max_{i\in G_k^0}\|\ _{iNT} - \ _i\| \leq \ _W\ 2\right\} \cap \left\{\max_{i\in G_k^0}\left\|\tilde{\ }_{i,z\Delta x} - \bar{\ }_{i,z\Delta x}\right\| \leq \ _{\bar{Q}}\ 2\right\}$$

$$\cap\left\{\max_{i\in G_k^0}\left\|\tilde{\ }_{i3}\right\| \leq (\ln\ )^{\nu}\right\} \cap \left\{\max_{i\in G_k^0}\left\|\tilde{\ }_{i4}\right\| \leq \sqrt{\ }\ _2\varkappa_{2NT}\right\}$$

Then conditional on Γ_{kNT} we have uniformly in $\ \in\ ^0_k$

$$\begin{aligned}(\tilde{\ }_i - \tilde{\ }_k)'(\tilde{\ }_{i2} + \tilde{\ }_{i3} + \tilde{\ }_{i4}) &\geq \left\|(\tilde{\ }_i - \tilde{\ }_k)'\tilde{\ }_{i2}\right\| - \left\|(\tilde{\ }_i - \tilde{\ }_k)'(\tilde{\ }_{i3} + \tilde{\ }_{i4})\right\| \\ &\geq \sqrt{\ }\ _2\tilde{\ }_{ki}\left\|\tilde{\ }_i - \tilde{\ }_k\right\| - \left\|\tilde{\ }_i - \tilde{\ }_k\right\|\left[(\ln\ )^{\nu} + \sqrt{\ }\ _2\varkappa_{2NT}\right] \\ &\geq \sqrt{\ }\ _2\ ^0_k\left\|\tilde{\ }_i - \tilde{\ }_k\right\|\ 4 \text{ for sufficiently large } (\ \ \ )\end{aligned}$$

because $\sqrt{\ }\ _2 \gg (\ln\ )^{\nu} + \sqrt{\ }\ _2\varkappa_{2NT}$ by Assumption B2(i). Then by Assumptions B2(i)-(ii)

$$\begin{aligned}(\tilde{\ }_{kNT,i}) &= \left(\ \in \tilde{\ }_k \mid\ \in\ ^0_k\right) = \left(\tilde{\ }_{i1} = \tilde{\ }_{i2} + \tilde{\ }_{i3} + \tilde{\ }_{i4}\right) \\ &\leq \left(\left|(\tilde{\ }_i - \tilde{\ }_k)'\tilde{\ }_{i1}\right| \geq \left|(\tilde{\ }_i - \tilde{\ }_k)'(\tilde{\ }_{i2} + \tilde{\ }_{i3} + \tilde{\ }_{i4})\right|\right) \\ &\leq \left(\left\|\tilde{\ }_{i1}\right\| \geq \sqrt{\ }\ _2\ ^0_k\ 4\ \ \Gamma_{kNT}\right) + (\Gamma^c_{kNT}) \to 0 \text{ as } (\ \ \ ) \to \infty\end{aligned}$$

It follows that $(||\tilde{\ }_i - \tilde{\ }_k|| = 0 \mid\ \in\ ^0_k) \to 1$ as $(\ \ \ ) \to \infty$ Now, observe that $(\cup_{k=1}^{K_0}\hat{\ }_{kNT}) \leq \sum_{k=1}^{K_0}(\hat{\ }_{kNT}) \leq \sum_{k=1}^{K_0}\sum_{i\in G_k^0}(\hat{\ }_{kNT,i})$ and by Assumption B2(ii)

$$\begin{aligned}\sum_{k=1}^{K_0}\sum_{i\in G_k^0}(\tilde{\ }_{kNT,i}) &\leq \sum_{k=1}^{K_0}\sum_{i\in G_k^0}\left(\left\|\tilde{\ }_{i1}\right\| \geq \sqrt{\ }\ _2\ ^0_k\ 4\ \ \Gamma_{kNT}\right) + (\Gamma^c_{kNT}) \\ &\leq \max_{1\leq i\leq N}\left(\left\|\tilde{\ }'_{i,z\Delta x}\ \ _{iNT}\frac{1}{\ }\sum_{t=1}^{T}\ _{it}\Delta\ _{it}\right\| \geq \ _2\ ^0_k\ 4\ \Gamma_{kNT}\right) + (1) \\ &\leq \max_{1\leq i\leq N}\left(\left\|\frac{1}{\ }\sum_{t=1}^{T}\ _{it}\Delta\ _{it}\right\| \geq \ _2\ ^{K_0-1}_{\alpha}\ (16\ _{\bar{Q}-W})\right) + (1) = (1)\end{aligned}$$

where we use the fact that $\left\|\tilde{\ }_{i,z\Delta x}\right\|\ \|\ _{iNT}\| \geq \left(\left\|\bar{\ }_{i,z\Delta x}\right\| - \left\|\tilde{\ }_{i,z\Delta x} - \bar{\ }_{i,z\Delta x}\right\|\right)(\|\ _i\| - \|\ _{iNT} - \ _i\|) \geq \ _{\bar{Q}-W}\ 4$ on the set Γ_{kNT}. Consequently, we have shown (i).

(ii) The proof of (i) is almost identical to that of Theorem 2.2(ii) and is omitted. ■

The proof follows closely from that of Theorem 2.4. Based on the subdifferential calculus, the KKT conditions for the minimization of (3.2) are that for each $\ = 1$ and $\ = 1\ \ \ _0$

$$\mathbf{0}_{p\times 1} = -2\tilde{\ }'_{i,z\Delta x}\ \ _{iNT}\frac{1}{\ }\sum_{t=1}^{T}\ _{it}\left(\Delta\ _{it} - \tilde{\ }'_i\Delta\ _{it}\right) + \frac{\ _2}{\ }\sum_{j=1}^{K_0}\tilde{\ }_{ij}\Pi_{l=1,l\neq j}^{K_0}\left\|\tilde{\ }_i - \tilde{\ }_l\right\| \quad \text{and}$$

$$\mathbf{0}_{p\times 1} = \frac{1}{\ }\sum_{i=1}^{N}\tilde{\ }_{ik}\Pi_{l=1,l\neq k}^{K_0}\left\|\tilde{\ }_i - \tilde{\ }_l\right\|$$

where $\tilde{}_{ij}$ is defined after (B.8). Fix $\in \{1 \quad {}_0\}$ As in the proof of Theorem 2.4, we can show that $\frac{2}{NT} \sum_{i\in\tilde{G}_k} \tilde{}'_{i,z\Delta x} \; {}_{iNT} \sum_{t=1}^{T} {}_{it} (\Delta \; {}_{it} - \tilde{}'_k \Delta \; {}_{it}) + \frac{\lambda_2}{N} \sum_{i\in\tilde{G}_0} \tilde{}_{ik} \Pi_{l=1,l\neq k}^{K_0} || \tilde{}_i - \tilde{}_l || = 0_{p\times 1}$ w.p.a.1. It follows that $\tilde{}_k = \tilde{}_{1k} + \tilde{\mathcal{R}}_k$ where $\tilde{}_{1k} = (\frac{1}{N} \sum_{i\in\tilde{G}_k} \tilde{}'_{i,z\Delta x} \; {}_{iNT} \tilde{}_{i,z\Delta x})^{-1} \times \frac{1}{NT} \sum_{i\in\tilde{G}_k} \tilde{}'_{i,z\Delta x} \; {}_{iNT} \sum_{t=1}^{T} {}_{it} \Delta \; {}_{it}$ and $\tilde{\mathcal{R}}_k = (\frac{1}{N} \sum_{i\in\tilde{G}_k} \tilde{}'_{i,z\Delta x} \; {}_{iNT} \tilde{}_{i,z\Delta x})^{-1} \frac{\lambda_2}{2N} \sum_{i\in\tilde{G}_0} \tilde{}_{ik} \Pi_{l=1,l\neq k}^{K_0} || \tilde{}_i - \tilde{}_l ||$ By Theorem 3.2, we can readily show that $(\sqrt{\quad} \| \tilde{\mathcal{R}}_k \| \geq \;) = \; (1)$ for any $\quad 0$, and

$$\sqrt{_k} (\tilde{}_{1k} - {}^0_k) = \left(\frac{1}{_k} \sum_{i\in G_k^0} \tilde{}'_{i,z\Delta x} W_{iNT} \right.$$

BROWNING, M., AND J. M. CARRO (2007): "Heterogeneity and Microeconometrics Modelling," In R. Blundell, W. K. Newey and T. Persson (Eds.), *Advances in Economics and Econometrics, Theory and Applications: Ninth World Congress of the Econometric Society*, Volume 3, pp. 45-74. Cambridge University Press, New York.

BROWNING, M., AND J. M. CARRO (2010): "Heterogeneity in Dynamic Discrete Choice Models," *Econometrics Journal* 13, 1-39.

BROWNING, M., AND J. M. CARRO (2014): "Dynamic Binary Outcome Models with Maximal Heterogeneity," *Journal of Econometrics* 178, 805-823.

CARROLL, C., AND D.N. WEIL (1994): "Saving and Growth: a Reinterpretation," *Carnegie-Rochester Conference Series on Public Policy* 40, 133-192

CHAN, N.H., C.Y. YAU, AND R-M. ZHANG (2014): "Group Lasso for Structural Break Time Series," *Journal of the American Statistical Association* 109, 590-599.

COLLIER, P. AND A. HOEFFLER (2004): "Greed and Grievance in Civil Wars," *Oxford Economic Papers* 56, 563-595.

DEATON, A. (1990): "Saving in Developing Countries: Theory and Review," *Proceedings of the World Bank Annual Conference on Development Economics* 61-96, The World Bank, Washington, DC.

DHAENE, G., AND K. JOCHMANS (2015): "Split-panel Jackknife Estimation of Fixed-effect Models," *Review of Economic Studies* 82, 991-1030.

DJANKOV, S. AND M. REYNAL-QUEROL (2010): "Poverty and Civil War: Revisiting the Evidence," *Review of Economics and Statistics* 92, 1035-1041.

DURLAUF, S. N., P. A. JOHNSON, AND J. R. TEMPLE (2005): "Growth Econometric," in P. Aghion and S. Durlauf (eds), *Handbook of Economic Growth*, Vol 1, pp. 555-677. Amsterdam: Elsevier.

EDWARDS, S (1996): "Why Are Latin America's Savings Rates So Low? An International Comparative Analysis," *Journal of Development Economics* 51, 5-44.

ESTEBAN, J., L. MAYORAL AND D. RAY (2012): "Ethnicity and Conflict: An Empirical Study," *American Economic Review* 102, 1310-1342.

FEARON, J.D. AND D. D. LAITIN (2003): "Ethnicity, Insurgency, and Civil War," *The American Political Science Review* 97, 75-90.

FELDSTEIN, M. (1980): "International Differences in Social Security and Saving," *Journal of Public Economics* 14, 225-244.

FERNÁNDEZ-VAL, I., AND L. LEE (2013): "Panel Data Models with Nonadditive Unobservable Heterogeneity: Estimation and Inference," *Quantitative Economics* 4, 453–481.

HAHN, J., AND G. KUERSTEINER (2002): "Asymptotically Unbiased Inference for a Dynamic Panel Model with Fixed Effects When Both and are Large," *Econometrica* 70, 1639-1657.

HAHN, J., AND G. KUERSTEINER (2011): "Bias Reduction for Dynamic Nonlinear Panel Models with Fixed Effects," *Econometric Theory* 27, 1152-1191.

HAHN, J., AND H.R. MOON (2010): "Panel Data Models with Finite Number of Multiple Equilibria," *Econometric Theory* 26, 863-881.

HAHN, J., AND W. NEWEY (2004): "Jackknife and Analytical Bias Reduction for Nonlinear Panel Models," *Econometrica* 72, 1295-1319.

HARCHAOUI, Z., AND C. LÉVY-LEDUC (2010): "Multiple Change-point Estimation with a Total Variation Penalty," *Journal of the American Statistical Association* 105, 1481-1493.

HSIAO, C. (2014): *Analysis of Panel Data*, 3rd edition. Cambridge University Press, New York.

HSIAO, C., AND H. PESARAN (2008): "Random Coefficient Panel Data Models," in L. Matyas and P. Sevestre (Eds.), *The Econometrics of Panel Data: Fundamentals and Recent Developments in Theory and Practice*, pp. 187-216, 3rd Edition. Springer-Verlag, Berlin.

HSIAO, C., AND A. K. TAHMISCIOGLU (1997): "A Panel Analysis of Liquidity Constraints and Firm Investment," *Journal of the American Statistical Association 92*, 455-465.

KASAHARA, H., AND K. SHIMOTSU (2009): "Nonparametric Identification of Finite Mixture Models of Dynamic Discrete Choices," *Econometrica* 77, 135-175.

KATO, K., A.F. GAVAO, AND G.V. MONTES-ROJAS (2012): "Asymptotics for Panel Quantile Regression Models with Individual Effects," *Journal of Econometrics* 170, 76-91.

KIVIET, J. F. (1995): "On Bias, Inconsistency, and Efficiency of Various Estimators in Dynamic Panel Data Models," *Journal of Econometrics* 68, 53-78.

LEE, K., M. H. PESARAN, AND R. SMITH (1997): "Growth and Convergence in a Multi-country Empirical Stochastic Growth Model," *Journal of Applied Econometrics* 12, 357-392.

LEE, Y. (2012): "Bias in Dynamic Panel Models under Time Series Misspecification," *Journal of Econometrics* 169, 54-60.

LEE, Y., AND P.C.B. PHILLIPS (2015): "Model Selection in the Presence of Incidental Parameters," *Journal of Econometrics* 188, 474-489.

LEEB, H., AND P.M. PÖTSCHER (2008): "Sparse Estimators and the Oracle Property, or the Return of Hodges' Estimator," *Journal of Econometrics* 142, 201–211.

LEEB, H., AND P.M. PÖTSCHER (2009): " On the Distribution of Penalized Maximum Likelihood Estimators: the LASSO, SCAD, and Thresholding," *Journal of Multivariate Analysis* 100, 2065–2082.

LI, H., J. ZHANG, AND J. ZHANG (2007): "Effects of Longevity and Dependency Rates on Saving and Growth," *Journal of Development Economics* 84, 138-154.

LIAO, Z. (2013): "Adaptive GMM Shrinkage Estimation with Consistent Moment Selection," *Econometric Theory* 29, 857-904.

LIN, C-C., AND S. Ng (2012): "Estimation of Panel Data Models with Parameter Heterogeneity when Group Membership Is Unknown," *Journal of Econometric Methods* 1, 42-55.

LOAYZA, N., K. SCHMIDT-HEBBEL, AND L. SERVÉN (2000): "Saving in Developing Countries: An Overview," *The World Bank Economic Review* 14, 393-414.

LU, X., AND L. SU (2016): "Shrinkage Estimation of Dynamic Panel Data Models with Interactive Fixed Effects," *Journal of Econometrics* 190, 148-175.

MIGUEL, E., S. SATYANATH AND E. SERGENTI (2004): "Economic Shocks and Civil Conflict: an Instrumental Variables Approach," *Journal of Political Economy* 112, 725-753.

NUNN, N. AND N. QIAN (2014): "US Food Aid and Civil Conflict," *American Economic Review* 104, 1630-1666.

PESARAN, H., Y. SHIN, AND R. SMITH (1999): "Pooled Mean Group Estimation of Dynamic Heterogeneous Panels," *Journal of the American Statistical Association* 94, 621-634.

PHILLIPS, P. C. B., AND D. SUL (2007a): "Transition Modeling and Econometric Convergence Tests," *Econometrica* 75, 1771-1855.

PHILLIPS, P. C. B., AND D. SUL (2007b): "Bias in Dynamic Panel Estimation with Fixed Effects, Incidental Trends and Cross Section Dependence," *Journal of Econometrics* 137, 162-188.

QIAN, J., AND L. SU (2015): "Shrinkage Estimation of Regression Models with Multiple Structural Changes," *Econometric Theory*, forthcoming.

RODRIK, D (2000): "Saving Transitions," *The World Bank Economic Review* 14, 481-507.

SARAFIDIS, V., AND N. WEBER (2015): "A Partially Heterogenous Framework for Analyzing Panel Data," *Oxford Bulletin of Economics and Statistics* 77, 274-296.

SU, L., AND Q. CHEN (2013): "Testing Homogeneity in Panel Data Models with Interactive Fixed Effects," *Econometric Theory*

Online Supplement to
"Identifying Latent Structures in Panel Data"[1]

Liangjun Su[a], Zhentao Shi[b], and Peter C. B. Phillips[c]

[a] *School of Economics, Singapore Management University*
[b] *Department of Economics, Chinese University of Hong Kong*
[c] *Yale University, University of Auckland,*
University of Southampton & Singapore Management University

This supplement is composed of four parts. Section S1 contains the proofs of some technical lemmas for the proofs of the main results in Section 2. Section S2 gives bias correction formulae in linear panel data models for both PPL and PGMM estimation. Sections S3 and S4 contain some additional simulation and applications results, respectively.

In this appendix, we state and prove some technical lemmas that are used in the proofs of the main results in Section 2. We first state an exponential inequality for strong mixing processes.

Let $\{ _t \quad = 1\ 2 \quad \}$ *be a zero-mean strong mixing process, not necessarily stationary, with the mixing coefficients satisfying* $( ) \leq \ _\alpha \ ^\tau$ *for some* $_\alpha \quad 0$ *and* $\in (0\ 1)$ *If* $\sup_{1\leq t\leq T} | \ _t| \leq \quad _T$ *then there exists a constant* $_0$ *depending on* $_\alpha$ *and* *such that for any* ≥ 2 *and* 0

$$\left(\left| \sum_{t=1}^{T} \ _t \right| \right) \leq \exp \left(- \frac{ \ _0 \ ^2}{ \ ^2_0 + \ ^2_T + \ _T (\ln \ )^2} \right)$$

where $^2_0 = \sup_{t\geq 1} \left[Var(\ _t) + 2 \sum_{s=t+1}^{\infty} |Cov(\ _t \ \ _s)| \right]$

Merlevède, Peilgrad, and Rio (2009, Theorem 2) prove (i) under the condition $( ) \leq \exp(-2 \quad)$ for some 0 If $_\alpha = 1$ we can take $= \exp(-2 \ )$ and apply the theorem to obtain the claim. ■

The above lemma is used in the proof of the following lemma.

(i) Let $(\ _{it}; \ )$ *be a* $\mathbb{R}^{d_\xi}$*-valued function indexed by the parameter* $\in \Phi$ *where* Φ *is a convex compact set in* $\mathbb{R}^{p+1}$ *and* $\mathbb{E}[\ (\ _{it}; \ )] = 0$ *for all* *and* $\in \Phi$ *Assume that there exists a function* $(\ _{it})$ *such that* $\| \ (\ _{it}; \ ) - \ (\ _{it}; \ ^-) \| \leq \ (\ _{it}) \| \ - \ ^- \|$ *for all* $^- \in \Phi$ *and* $sup_\phi \| \ (\ _{it}; \ )\| \leq \ (\ _{it})$ *Assume that* $\mathbb{E}| \ (\ _{it})|^q \quad \infty$ *for some* ≥ 6 *such that* $= \ (\ ^{q/2-1})$ *Let* $\{ \ _i\}$ *be a nonstochastic sequence in* Φ *Then*

(i) $\max_{1\leq i\leq N} \left\| \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \ (\ _{it}; \ _i) \right\| = \ _P((\ln \ )^3)$

(ii) $\max_{1\leq i\leq N} \left(\left\| \frac{1}{T} \sum_{t=1}^{T} \ (\ _{it}; \ _i) \right\| \geq \ \right) = \ (\ ^{-1})$ *for any given* 0

(iii) $\left(\max_{1\leq i\leq N} \left\| \frac{1}{T} \sum_{t=1}^{T} \ (\ _{it}; \ _i) \right\| \geq \ \right) = \ (\ ^{-1})$ *for any given* 0 *if* $^2 = \ (\ ^{q/2-1})$

where $= \ _{NT}$ *satisfies* $(\ln \ )^3 = \ (\ ^{1/2} \ )$

[1] Acknowledgements made in the leading footnote of the main paper apply also to this Online Supplement. In particular, Su acknowledges support from the Singapore Ministry of Education for Academic Research Fund (AcRF) under the Tier-2 grant number MOE2012-T2-2-021. Phillips acknowledges NSF support under Grant Nos. SES-0956687 and SES-1285258.

Address Correspondence to: Liangjun Su, School of Economics, Singapore Management University, 90 Stamford Road, Singapore 178903; E-mail: ljsu@smu.edu.sg, Phone: +65 6828 0386.

(i) Let $_{NT} = \ ^{1/2}$ Let $_\xi$ be an arbitrary $_\xi \times 1$ nonrandom vector with $\| \ _\xi\| = 1$. Let $\mathbf{1}_{it} = \mathbf{1}\{\| \ (\ _{it}; \ _i)\| \leq \ _{NT}\}$ and $\bar{\mathbf{1}}_{it} = 1 - \mathbf{1}_{it}$ Define

$$
\begin{aligned}
_1(\ _{it}; \ _i) &= \ '_\xi\{ \ (\ _{it}; \ _i)\mathbf{1}_{it} - \mathbb{E}[\ (\ _{it}; \ _i)\mathbf{1}_{it}]\} \\
_2(\ _{it}; \ _i) &= \ '_\xi \ (\ _{it}; \ _i)\bar{\mathbf{1}}_{it} \quad \text{and} \quad _{3it} = - \ '_\xi\mathbb{E}[\ (\ _{it}; \ _i)\bar{\mathbf{1}}_{it}]
\end{aligned}
$$

Apparently $_1(\ _{it}; \ _i) + \ _2(\ _{it}; \ _i) + \ _{3it} = \ '_\xi \ (\ _{it}; \ _i)$ as $\mathbb{E}[\ (\ _{it}; \ _i)] = 0$ We prove the lemma by showing that (i1) $\max_{1\leq i\leq N}\left\| \frac{1}{\sqrt{T}} \sum_{t=1}^T \ _1(\ _{it}; \ _i)\right\| = \ _P((\ln \ )^3)$ (i2) $\max_{1\leq i\leq N}\left\| \frac{1}{\sqrt{T}} \sum_{t=1}^T \ _2(\ _{it}; \ _i)\right\| \geq \ (\ln \ )^3 = \ (1)$ for any given $\ 0$ and (i3) $\max_{1\leq i\leq N}\left\| \frac{1}{\sqrt{T}} \sum_{t=1}^T \ _{3it}\right\| = \ ((\ln \ )^3)$

First, we prove (i3). By the Hölder and Markov inequalities

$$
\begin{aligned}
\max_{1\leq i\leq N}\left\| \frac{1}{\sqrt{\ }} \sum_{t=1}^T \ _{3it}\right\| &\leq \ ^{1/2} \max_{1\leq i\leq N}\max_{1\leq t\leq T} \|\mathbb{E}[\ (\ _{it}; \ _i)\bar{\mathbf{1}}_{it}]\| \\
&\leq \ ^{1/2} \max_{1\leq i\leq N}\max_{1\leq t\leq T} \ \mathbb{E}\| \ (\ _{it}; \ _i)\|^{q/2} \ ^{2/q} \ (\| \ (\ _{it}; \ _i)\| \ \ ^{1/2}) \ ^{(q-2)/q} \\
&\leq \ ^{1/2} \ _{1q} \max_{1\leq i\leq N}\max_{1\leq t\leq T} \ (\| \ (\ _{it}; \ _i)\| \ \ ^{1/2}) \ ^{(q-2)/q} \\
&\leq \ ^{1/2} \ _{1q} \max_{1\leq i\leq N}\max_{1\leq t\leq T} \ \ ^{-q/2}\mathbb{E}(\| \ (\ _{it}; \ _i)\|^q) \ ^{(q-2)/q} \\
&= \ _{1q} \ _{2q} \ ^{(3-q)/2} = \ ((\ln \ )^3) \text{ for any } \ \geq 3
\end{aligned}
$$

where $_{1q} \equiv \max_{1\leq i\leq N}\max_{1\leq t\leq T} \ \mathbb{E}\| \ (\ _{it}; \ _i)\|^{q/2} \ ^{2/q}$ and $_{2q} \equiv \max_{1\leq i\leq N}\max_{1\leq t\leq T}\{\mathbb{E}(\| \ (\ _{it}; \ _i)\|^q)\}^{(q-2)/q}$

Next, we prove (i2). Noting that $\left\| \frac{1}{\sqrt{T}} \sum_{t=1}^T \ _2(\ _{it}; \ _i)\right\| \geq (\ln \ )^3$ implies that $\max_{1\leq t\leq T}\| \ (\ _{it}; \ _i)\| \ \ _{NT}$ by the Boole and Markov inequalities, the dominated convergence theorem, and the stated conditions, we have

$$
\begin{aligned}
\left[\max_{1\leq i\leq N}\left\| \frac{1}{\sqrt{\ }} \sum_{t=1}^T \ _2(\ _{it}; \ _i)\right\| \geq \ (\ln \ )^3\right] &\leq \ \left[\max_{1\leq i\leq N}\max_{1\leq t\leq T}\| \ (\ _{it}; \ _i)\| \ \ _{NT}\right] \\
&\leq \ \max_{1\leq i\leq N}\max_{1\leq t\leq T} \ (\ (\ _{it}) \ \ _{NT}) \\
&\leq \frac{\ }{\ ^{q/2}} \max_{1\leq i\leq N}\max_{1\leq t\leq T}\mathbb{E} \ | \ (\ _{it})|^q \mathbf{1} \ (\ _{it}) \ \ ^{1/2} \ \ \\
&= \ (\ ^{1-q/2}) = \ (1)
\end{aligned}
$$

Now, we prove (i1). We observe that for any $\ 0$

$$
\left[\max_{1\leq i\leq N}\left\| \frac{1}{\sqrt{\ }} \sum_{t=1}^T \ _1(\ _{it}; \ _i)\right\| \geq \ (\ln \ )^3\right] \leq \sum_{i=1}^N \ \left[\left\| \frac{1}{\sqrt{\ }} \sum_{t=1}^T \ _1(\ _{it}; \ _i)\right\| \geq \ (\ln \ )^3\right]
$$

We choose $\ 0$ and divide Φ into subsets $\Phi_j \ = 1 \ \ _\varepsilon$ such that $\| \ - \bar{\ }\| \ \sqrt{\ }$ for all $\bar{\ } \in \Phi_j$, where $_\varepsilon = \ (\ ^{(p+1)/2})$ Then

$$
\begin{aligned}
\sum_{i=1}^N \ \left[\left\| \frac{1}{\sqrt{\ }} \sum_{t=1}^T \ _1(\ _{it}; \ _i)\right\| \geq \ (\ln \ )^3\right] &\leq \sum_{i=1}^N \ \left[\sup_{\phi\in\Phi}\left\| \frac{1}{\sqrt{\ }} \sum_{t=1}^T \ _1(\ _{it}; \ )\right\| \geq \ (\ln \ )^3\right] \\
&\leq \sum_{i=1}^N\sum_{j=1}^{n_\varepsilon} \ \left[\sup_{\phi\in\Phi_j}\left\| \frac{1}{\sqrt{\ }} \sum_{t=1}^T \ _1(\ _{it}; \ )\right\| \geq \ (\ln \ )^3\right]
\end{aligned}
$$

Let $_j \in \Phi_j$ Then for any $\in \Phi_j$ we have

$$\left\| \frac{1}{\sqrt{}} \sum_{t=1}^{T} {}_1 ({}_{it};\ ) \right\| \leq \left\| \frac{1}{\sqrt{}} \sum_{t=1}^{T} {}_1 ({}_{it};\ {}_j) \right\| + \left\| \frac{1}{\sqrt{}} \sum_{t=1}^{T} \left[{}_1 ({}_{it};\ {}_j) - {}_1 ({}_{it};\ ) \right] \right\|$$
$$\leq \left\| \frac{1}{\sqrt{}} \sum_{t=1}^{T} {}_1 ({}_{it};\ {}_j) \right\| + \frac{2}{\sqrt{}} \sum_{t=1}^{T} ({}_{it}) \| - {}_j \|$$
$$\leq \left\| \frac{1}{\sqrt{}} \sum_{t=1}^{T} {}_1 ({}_{it};\ {}_j) \right\| + \left| \frac{2}{} \sum_{t=1}^{T} \{ ({}_{it}) - \mathbb{E} [({}_{it})] \} \right| + \frac{2}{} \sum_{t=1}^{T} \mathbb{E} [({}_{it})]$$

It follows that

$$\left[\sup_{\phi \in \Phi_j} \left\| \frac{1}{\sqrt{}} \sum_{t=1}^{T} {}_1 ({}_{it};\ ) \right\| \geq (\ln\ )^3 \right] \leq \left[\left\| \frac{1}{\sqrt{}} \sum_{t=1}^{T} {}_1 ({}_{it};\ {}_j) \right\| \geq (\ln\ )^3\ 3 \right]$$
$$+ \left[\left| \frac{2}{} \sum_{t=1}^{T} ({}_{it}) - \mathbb{E} [({}_{it})] \right| \geq (\ln\ )^3\ 3 \right]$$

as $\frac{\varepsilon}{T} \sum_{t=1}^{T} \mathbb{E} [({}_{it})] \geq (\ln\ )^3\ 3 = 0$ Then

$$\left[\max_{1 \leq i \leq N} \left\| \frac{1}{\sqrt{}} \sum_{t=1}^{T} ({}_{it};\ {}_i) \right\| \geq (\ln\ )^3 \right] \leq \sum_{i=1}^{N} \sum_{j=1}^{n_\varepsilon} \left[\left\| \frac{1}{\sqrt{}} \sum_{t=1}^{T} {}_1 ({}_{it};\ {}_j) \right\| \geq (\ln\ )^3\ 3 \right]$$
$$+ \sum_{i=1}^{N} \sum_{j=1}^{n_\varepsilon} \left[\left| \frac{2}{} \sum_{t=1}^{T} ({}_{it}) - \mathbb{E} [({}_{it})] \right| \geq (\ln\ )^3\ 3 \right]$$

For the first term, we have by Lemma S1.1

$$\sum_{i=1}^{N} \sum_{j=1}^{n_\varepsilon} \left[\left\| \frac{1}{\sqrt{}} \sum_{t=1}^{T} {}_1 ({}_{it};\ {}_j) \right\| \geq (\ln\ )^3\ 3 \right] \leq {}_\varepsilon \exp \left(- \frac{{}^2\ {}_0\ (\ln\ )^6\ 9}{{}^{-2} + 4\ {}^2_{NT} + \frac{2C}{3}\ {}_{NT}\ {}^{1/2} (\ln\ )^{3+2}} \right)$$
$$= \exp \left(- \frac{{}^2\ {}_0\ (\ln\ )^6\ 9}{{}^{-2} + 4 + \frac{2C}{3} (\ln\ )^5} + \ln\ + \ln\ {}_\varepsilon \right)$$
$$\to 0 \text{ for sufficiently large}$$

Similarly, we can show that $\sum_{i=1}^{N} \sum_{j=1}^{n_\varepsilon} \left\| \frac{2\varepsilon}{T} \sum_{t=1}^{T} ({}_{it}) - \mathbb{E} [({}_{it})] \right\| \geq (\ln\ )^3\ 3 = (1)$ Then (i1) follows. This completes the proof of (i).

(ii) Let $_1$ $_2$ and $_{3it}$ be as defined in (i). Noting that $_{3it}$ is nonrandom, it suffices to show that for any given 0 we have (ii1) $\max_{1 \leq i \leq N} (\| \frac{1}{T} \sum_{t=1}^{T} {}_1 ({}_{it};\ {}_i) \| \geq\ ) = (1)$ (ii2) $\max_{1 \leq i \leq N} (\| \frac{1}{T} \sum_{t=1}^{T} {}_2 ({}_{it};\ {}_i) \| \geq\ ) = (1)$ and (ii3) $\max_{1 \leq i \leq N} \left\| \frac{1}{T} \sum_{t=1}^{T} {}_{3it} \right\| = ( )$. Following the analysis of $\frac{1}{\sqrt{T}} \sum_{t=1}^{T} {}_{3it}$ in (i), we have

$$\max_{1 \leq i \leq N} \left\| \frac{1}{} \sum_{t=1}^{T} {}_{3it} \right\| \leq {}_{1q}\ {}_{2q}\ {}^{(2-q)/2} = ( )$$

where we use the fact that $\gg {}^{-1/2}(\ln\ )^3$ and ≥ 3 by the stated conditions. Thus, (ii3) follows. Following the analysis of $\frac{1}{\sqrt{T}} \sum_{t=1}^{T} {}_3(\ {}_{it};\ {}_i)$ in (i2), we have

$$\max_{1\leq i\leq N} \left(\left\| \frac{1}{} \sum_{t=1}^{T} {}_2(\ {}_{it};\ {}_i) \right\| \geq \right) \leq \max_{1\leq i\leq N} \left(\max_{1\leq t\leq T} \| (\ {}_{it};\ {}_i)\| \quad {}_{NT} \right)$$

$$\leq \max_{1\leq i\leq N} \max_{1\leq t\leq T} (\ (\ {}_{it}) \quad {}_{NT})$$

$$= (\quad {}^{1-q/2}) = (1)$$

That is, (ii2) follows. For (ii1), the analysis is similar to that of $\max_{1\leq i\leq N} \left\| \frac{1}{\sqrt{T}} \sum_{t=1}^{T} {}_1(\ {}_{it};\ {}_i) \right\|$ in (i1) with $(\ln\ )^3$ replaced by ${}^{1/2}$ We now require ${}^{1/2}\ (\ln\ )^3 \to \infty$ as $(\quad) \to \infty$ This completes the proof of (ii).

(iii) Let ${}_1\ {}_2$ and ${}_{3it}$ be as defined in (i). Noting that ${}_{3it}$ is nonrandom, it suffices to show that for any given 0 we have (iii1) $\cdot\ (\max_{1\leq i\leq N} \|\frac{1}{T} \sum_{t=1}^{T} {}_1(\ {}_{it};\ {}_i)\| \geq\ ) = (1)$ (iii2) $\cdot\ (\max_{1\leq i\leq N} \|\frac{1}{T} \sum_{t=1}^{T} {}_2(\ {}_{it};\ {}_i)\| \geq\ ) = (1)$ and (iii3) $\max_{1\leq i\leq N} \left\| \frac{1}{T} \sum_{t=1}^{T} {}_{3it} \right\| = (\ )$. Following the analysis of $\frac{1}{\sqrt{T}} \sum_{t=1}^{T} {}_{3it}$ in (i), we have

$$\max_{1\leq i\leq N} \left\| \frac{1}{} \sum_{t=1}^{T} {}_{3it} \right\| \leq {}_{1q}\ {}_{2q}\ {}^{(2-q)/2} = (\ )$$

where we use the fact that $\gg {}^{-1/2}(\ln\ )^3$ and ≥ 6 by the stated conditions. Thus, (iii3) follows. Following the analysis of $\frac{1}{\sqrt{T}} \sum_{t=1}^{T} {}_3(\ {}_{it};\ {}_i)$ in (i2), we have

$$\cdot \left(\max_{1\leq i\leq N} \left\| \frac{1}{} \sum_{t=1}^{T} {}_2(\ {}_{it};\ {}_i) \right\| \geq \right) \leq \cdot \left(\max_{1\leq i\leq N} \max_{1\leq t\leq T} \| (\ {}_{it};\ {}_i)\| \quad {}_{NT} \right)$$

$$\leq {}^2 \max_{1\leq i\leq N} \max_{1\leq t\leq T} (\ (\ {}_{it}) \quad {}_{NT})$$

$$= (\ {}^2\ {}^{1-q/2}) = (1)$$

That is, (iii2) follows. For (iii1), the analysis is similar to that of $\max_{1\leq i\leq N} \left\| \frac{1}{\sqrt{T}} \sum_{t=1}^{T} {}_1(\ {}_{it};\ {}_i) \right\|$ in (i1) with $(\ln\ )^3$ replaced by ${}^{1/2}$ We now require ${}^{1/2}\ (\ln\ )^3 \to \infty$ as $(\quad) \to \infty$ This completes the proof of (ii). ■

Recall that $\hat{\Psi}_i(\quad) = \frac{1}{T} \sum_{t=1}^{T} (\ {}_{it};\quad)$ and $\Psi_i(\quad) = \frac{1}{T} \sum_{t=1}^{T} \mathbb{E}[\ (\ {}_{it};\quad)]$ Recall that $\hat{}_i(\ {}_i) = \arg\min_{\mu_i} \frac{1}{T} \sum_{t=1}^{T} (\ {}_{it};\ {}_i\ {}_i)$ The following three lemmas study the properties of $\hat{\Psi}_i(\quad)$ and $\hat{}_i(\ {}_i)$

For any 0 *we have* $\max_{1\leq i\leq N} \sup_{(\beta,\mu)} \left| \hat{\Psi}_i(\quad) - \Psi_i(\quad) \right| \geq\ = (\ {}^{-1})$

The proof is analogous to that of Lemma S1.2(iii). ■

For any 0 *we have* $[\max_{1\leq i\leq N} |\hat{}_i(\ {}_i) - {}_i(\ {}_i)| \geq\ ] = (\ {}^{-1})$

Let $= \min_i \left[\inf_{\mu_i:|\mu_i-\mu_i(\beta_i)|>\eta} \Psi_i(\ {}_i\ {}_i) - \Psi_i(\ {}_i\ {}_i(\ {}_i)) \right]$ Then 0 by Assumptions A1(ii) and (v). Then conditional on the event $\equiv \max_{1\leq i\leq N} \sup_{(\beta,\mu)} \left| \hat{\Psi}_i(\quad) - \Psi_i(\quad) \right| \leq \frac{1}{3}$ we have

$$\inf_{|\mu_i-\mu_i(\beta_i)|>\eta} \hat{\Psi}_i(\ {}_i\ {}_i) \geq \inf_{|\mu_i-\mu_i(\beta_i)|>\eta} \Psi_i(\ {}_i\ {}_i) - \frac{1}{3}$$

$$\geq \Psi_i(\ {}_i\ {}_i(\ {}_i)) + \frac{2}{3}$$

$$\geq \hat{\Psi}_i(\ {}_i\ {}_i(\ {}_i)) + \frac{1}{3}$$

On the other hand, $\hat{\Psi}_i(\ _i\ \hat{}_i(\ _i)) \leq \hat{\Psi}_i(\ _i\ \ _i(\ _i))$ It follows that $(\max_{1\leq i\leq N}|\hat{}_i(\ _i) - \ _i(\ _i)| \leq \ ) \leq$ $(\ ) = \ ^{(\ \ -1)}$ by Lemma S1.3. ■

(i) $\hat{}_i(\ _i) - \ _i(\ _i) = \ _P(\ ^{-1/2})$ *for each*

(ii) $\max_{1\leq i\leq N}|\hat{}_i(\ _i) - \ _i(\ _i)| = \ _P(\ ^{-1/2}(\ln\ )^3)$

(iii) $\max_{1\leq i\leq N}|\Psi_i(\ _i\ \hat{}_i(\ _i)) - \Psi_i(\ _i\ \ _i(\ _i))| = \ _P(\ ^{-1/2}(\ln\ )^3)$

(iv) $(\max_{1\leq i\leq N}|\hat{}_i(\ _i) - \ _i(\ _i)| \geq \ ^{-1/2}(\ln\ )^{3+\nu}) = \ ^{(\ \ -1)}$ *for any* $\ 0$ *and* $\ 0$

(v) $(\max_{1\leq i\leq N}|\Psi_i(\ _i\ \hat{}_i(\ _i)) - \Psi_i(\ _i\ \ _i(\ _i))| \geq \ ^{-1/2}(\ln\ )^{3+\nu}) = \ ^{(\ \ -1)}$ *for any* $\ 0$ *and* 0

(i)-(ii) Noting that $\hat{}_i(\ _i) = \arg\min_{\mu_i} \frac{1}{T}\sum_{t=1}^{T}(\ _{it};\ _i\ _i)$ we have

$$\begin{aligned} 0 &= \frac{1}{\ }\sum_{t=1}^{T}\ _i(\ _{it};\ _i\ \hat{}_i(\ _i)) \\ &= \frac{1}{\ }\sum_{t=1}^{T}\ _i(\ _{it};\ _i\ \ _i(\ _i)) + \frac{1}{\ }\sum_{t=1}^{T}\ _i^{\mu_i}(\ _{it};\ _i\ \breve{}_i(\ _i))[\hat{}_i(\ _i) - \ _i(\ _i)] \end{aligned}$$

where $\breve{}_i(\ _i)$ lies between $\hat{}_i(\ _i)$ and $\ _i(\ _i)$ for each It follows that

$$\hat{}_i(\ _i) - \ _i(\ _i) = -\left[\frac{1}{\ }\sum_{t=1}^{T}\ _i^{\mu_i}(\ _{it};\ _i\ \breve{}_i(\ _i))\right]^{-1}\frac{1}{\ }\sum_{t=1}^{T}\ _i(\ _{it};\ _i\ \ _i(\ _i)) \tag{S1}$$

provided $\frac{1}{T}\sum_{t=1}^{T}\ _i^{\mu_i}(\ _{it};\ _i\ \breve{}_i(\ _i))$ is asymptotically nonvanishing. Let $\ _{it}(\ _i) = \ _i(\ _{it};\ _i\ \ _i(\ _i))$ Noting that $\mathbb{E}[\ _{it}(\ _i)] = 0$ and

$$\begin{aligned} \mathrm{Var}\left(\frac{1}{\ }\sum_{t=1}^{T}\ _{it}(\ _i)\right) &= \frac{1}{\ ^2}\sum_{t=1}^{T}\sum_{s=1}^{T}\mathrm{Cov}(\ _{it}(\ _i)\ \ _{is}(\ _i)) \\ &\leq 8\max_{i,t}\{\mathbb{E}|\ _{it}(\ _i)|^q\}^{2/q}\frac{1}{\ ^2}\sum_{t=1}^{T}\sum_{s=1}^{T}(|\ -\ |)^{1-2/q} \\ &\leq 8\max_{i,t}\{\mathbb{E}|\ _{it}(\ _i)|^q\}^{2/q}\frac{1}{\ }\sum_{\tau=1}^{\infty}(\ )^{1-2/q} = \ \left(\frac{1}{\ }\right) \end{aligned}$$

by the Davydov inequality (e.g., Corollary A.2 in Hall and Heyde (1980)), we have $\frac{1}{T}\sum_{t=1}^{T}\ _{it}(\ _i) = \ _P(\ ^{-1/2})$ by the Chebyshev inequality. In addition, by a simple application of Lemma S1.2(i), we can show that $\max_{1\leq i\leq N}\left|\frac{1}{T}\sum_{t=1}^{T}\ _{it}(\ _i)\right| = \ _P(\ ^{-1/2}(\ln\ )^3)$

For $\frac{1}{T}\sum_{t=1}^{T}\ _i^{\mu_i}(\ _{it};\ _i\ \breve{}_i(\ _i))$ we make the following decomposition:

$$\begin{aligned} \frac{1}{\ }\sum_{t=1}^{T}\ _i^{\mu_i}(\ _{it};\ _i\ \breve{}_i(\ _i)) &= \frac{1}{\ }\sum_{t=1}^{T}\mathbb{E}\left[\ _i^{\mu_i}(\ _{it};\ _i\ \ _i(\ _i))\right] \\ &+\frac{1}{\ }\sum_{t=1}^{T}\left\{\ _i^{\mu_i}(\ _{it};\ _i\ \ _i(\ _i)) - \mathbb{E}\left[\ _i^{\mu_i}(\ _{it};\ _i\ \ _i(\ _i))\right]\right\} \\ &+\frac{1}{\ }\sum_{t=1}^{T}\left\{\ _i^{\mu_i}(\ _{it};\ _i\ \breve{}_i(\ _i)) - \ _i^{\mu_i}(\ _{it};\ _i\ \ _i(\ _i))\right\} \end{aligned} \tag{S2}$$

By Assumption A1(v), $\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}\left[\ {}_i^{\mu_i}(\ {}_{it};\ {}_i\ {}_i(\ {}_i))\right] = \ {}_{i\mu\mu}(\ {}_i) \geq\ {}_H\ 0$ uniformly in By a simple application of Lemma S1.2(i), we have

$$\max_{1\leq i\leq N}\left|\frac{1}{T}\sum_{t=1}^{T}\left\{\ {}_i^{\mu_i}(\ {}_{it};\ {}_i\ {}_i(\ {}_i)) - \mathbb{E}\left[\ {}_i^{\mu_i}(\ {}_{it};\ {}_i\ {}_i(\ {}_i))\right]\right\}\right| = \ {}_P(1)$$

Next, by Assumption A1, and Lemmas S1.2(i) and S1.4, we have

$$\begin{aligned}
&\max_{1\leq i\leq N}\left|\frac{1}{T}\sum_{t=1}^{T}\left[\ {}_i^{\mu_i}(\ {}_{it};\ {}_i\ \breve{}_i(\ {}_i)) - \ {}_i^{\mu_i}(\ {}_{it};\ {}_i\ {}_i(\ {}_i))\right]\right| \\
\leq\ & \max_{1\leq i\leq N}\frac{1}{T}\sum_{t=1}^{T}(\ {}_{it})\,|\breve{}_i(\ {}_i) - \ {}_i(\ {}_i)| \\
\leq\ & \left\{\max_{1\leq i\leq N}\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}[\ (\ {}_{it})] + \max_{1\leq i\leq N}\left|\frac{1}{T}\sum_{t=1}^{T}\{\ (\ {}_{it}) - \mathbb{E}[\ (\ {}_{it})]\}\right|\right\}\max_{1\leq i\leq N}|\hat{}_i(\ {}_i) - \ {}_i(\ {}_i)| \\
\leq\ & \ {}_M^{1/q} + \ {}_P(1)\ \ {}_P(1) = \ {}_P(1)
\end{aligned} \tag{S3}$$

It follows that $\frac{1}{T}\sum_{t=1}^{T}\ {}_i^{\mu_i}(\ {}_{it};\ {}_i\ \breve{}_i(\ {}_i)) = \ {}_{i\mu\mu}(\ {}_i) + \ {}_P(1)$ uniformly in $\hat{}_i(\ {}_i) - \ {}_i(\ {}_i) = \ {}_P(\ {}^{-1/2})$ for each and $\max_{1\leq i\leq N}|\hat{}_i(\ {}_i) - \ {}_i(\ {}_i)| = \ {}_P(\ {}^{-1/2}(\ln\ )^3)$

(iii) In view of the definition that $\Psi_i(\ {}_i\ {}_i) = \mathbb{E}[\ (\ {}_{it};\ {}_i\ {}_i)]$ we have $\max_{1\leq i\leq N}|\Psi_i(\ {}_i\ \hat{}_i(\ {}_i)) - \Psi_i(\ {}_i\ {}_i(\ {}_i))| = \max_{i,t}\mathbb{E}|\ (\ {}_{it})|\,|\hat{}_i(\ {}_i) - \ {}_i(\ {}_i)| = \ {}_P(\ {}^{-1/2}(\ln\ )^3)$

(iv) We define the following events:

$$\begin{aligned}
{}_1 &\equiv \left\{\max_{1\leq i\leq N}|\hat{}_i(\ {}_i) - \ {}_i(\ {}_i)| \leq \ {}_H\ (6\ {}_M^{1/q})\right\} \\
{}_2 &\equiv \left\{\max_{1\leq i\leq N}\left|\frac{1}{T}\sum_{t=1}^{T}\{\ (\ {}_{it}) - \mathbb{E}[\ (\ {}_{it})]\}\right| \leq \ {}_M^{1/q}\ 2\right\} \\
{}_3 &\equiv \left\{\max_{1\leq i\leq N}\left|\frac{1}{T}\sum_{t=1}^{T}\left[\ {}_i^{\mu_i}(\ {}_{it};\ {}_i\ \breve{}_i(\ {}_i)) - \ {}_i^{\mu_i}(\ {}_{it};\ {}_i\ {}_i(\ {}_i))\right]\right| \leq \ {}_H\ 4\right\} \\
{}_4 &\equiv \left\{\min_{1\leq i\leq N}\left|\frac{1}{T}\sum_{t=1}^{T}\ {}_i^{\mu_i}(\ {}_{it};\ {}_i\ {}_i(\ {}_i))\right| \geq \ {}_H\ 2\right\} \\
{}_5 &\equiv \left\{\min_{1\leq i\leq N}\left|\frac{1}{T}\sum_{t=1}^{T}\ {}_i^{\mu_i}(\ {}_{it};\ {}_i\ \breve{}_i(\ {}_i))\right| \geq \ {}_H\ 4\right\}
\end{aligned}$$

Let ${}_j^c$ denote the complement of ${}_j$ for $= 1\ 2\ 3\ 4\ 5$ Let ${}_i = \hat{}_i(\ {}_i) - \ {}_i(\ {}_i)$ By Lemmas S1.4 and S1.2(iii), $(\ {}_1^c) = (\ {}^{-1})$ and $(\ {}_2^c) = (\ {}^{-1})$ Then by (S3)

$$\begin{aligned}
&(\ {}_3^c) \\
\leq\ & \left(\left\{\max_{1\leq i\leq N}\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}[\ (\ {}_{it})] + \max_{1\leq i\leq N}\left|\frac{1}{T}\sum_{t=1}^{T}\{\ (\ {}_{it}) - \mathbb{E}[\ (\ {}_{it})]\}\right|\right\}\max_{1\leq i\leq N}|\ {}_i| \geq \ {}_H\ 4\right) \\
\leq\ & \left(\left\{\max_{1\leq i\leq N}\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}[\ (\ {}_{it})] + \max_{1\leq i\leq N}\left|\frac{1}{T}\sum_{t=1}^{T}\{\ (\ {}_{it}) - \mathbb{E}[\ (\ {}_{it})]\}\right|\right\}\max_{1\leq i\leq N}|\ {}_i| \geq \ {}_H\ 4\ \ {}_2\right) \\
&+ (\ {}_2^c) \\
\leq\ & \left(3\ {}_M^{1/q}\max_{1\leq i\leq N}|\ {}_i| \geq \ {}_H\ 2\right) + (\ {}_2^c) \\
\leq\ & (\ {}_1^c) + (\ {}_2^c) = (\ {}^{-1})
\end{aligned}$$

where $\check{}_i(\ ^0_i)$ lies between $\hat{}_i(\ ^0_i)$ and $_i(\ ^0_i)$ By Assumptions A1, Lemma S1.5, and the Markov inequality, one can readily show that the first term is $_P(\ ^{-1/2})$ and the second is $_P(\ ^{-1})$ It follows that $\hat{}_i - \ _i = \ _P(\ ^{-1/2})$

(ii) By a simple application of Lemma S1.2(i), $\max_{1\le i\le N}\|\ _i\| = \ _P(\ ^{-1/2}(\ln\ )^3)$ Next,

$$\begin{aligned}\max_{1\le i\le N}\left\|\hat{}_i - \ _i\right\| &\le \max_{1\le i\le N}\frac{1}{}\sum_{t=1}^{T}(\ _{it})\left|\hat{}_i(\ ^0_i) - \ _i(\ ^0_i)\right| \\ &\le \left\{\max_{1\le i\le N}\frac{1}{}\sum_{t=1}^{T}\mathbb{E}[\ (\ _{it})] + \max_{1\le i\le N}\left|\frac{1}{}\sum_{t=1}^{T}\{\ (\ _{it}) - \mathbb{E}[\ (\ _{it})]\}\right|\right\} \\ &\quad\times\max_{1\le i\le N}\left|\hat{}_i(\ ^0_i) - \ _i(\ ^0_i)\right| \\ &= \{\ (1) + \ _P(1)\}\ _P(\ ^{-1/2}(\ln\ )^3) = \ _P(\ ^{-1/2}(\ln\ )^3)\end{aligned}$$

(iii) By the Cauchy-Schwarz inequality, $\frac{1}{N}\sum_{i=1}^{N}\left\|\hat{}_i\right\|^2 \le \frac{2}{N}\sum_{i=1}^{N}\|\ _i\|^2 + \frac{2}{N}\sum_{i=1}^{N}\left\|\hat{}_i - \ _i\right\|^2$ The first term in $_P(\ ^{-1})$ by the Markov inequality and the calculation in (i). Using the decomposition of $\hat{}_i - \ _i$ in (i), we can readily show that the second term is $_P(\ ^{-1})$ Then $\frac{1}{N}\sum_{i=1}^{N}\left\|\hat{}_i\right\|^2 = \ _P(\ ^{-1})$

(iv) The result follows by a simple application of Lemma S1.2(ii) and Assumption A2.

(v) The proof is similar to that of (ii) but we now apply Lemmas S1.2(iii) and S1.5(iv). ■

The next lemma establishes the uniform consistency of $\hat{}_i$

For any $\ 0$ *we have* $\left(\max_{1\le i\le N}\left\|\hat{}_i - \ ^0_i\right\|\ \right) = \ (\ ^{-1})$

Recall that $^{(K_0)}_{1NT,\lambda_1}(\boldsymbol{\beta}\ \boldsymbol{\alpha}) = \ _{1,NT}(\boldsymbol{\beta}) + \frac{\lambda_1}{N}\sum_{i=1}^{N}\Pi_{k=1}^{K_0}\|\ _i - \ _k\|$ where $_{1,NT}(\boldsymbol{\beta}) = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}(\ _{it};\ _i\ \hat{}_i(\ _i)) = \frac{1}{N}\sum_{i=1}^{N}\hat{\Psi}_i(\ _i\ \hat{}_i(\ _i))$ Noting that $(\hat{\boldsymbol{\beta}}\ \hat{\boldsymbol{\alpha}}) = \arg\min_{(\boldsymbol{\beta},\boldsymbol{\alpha})}\ ^{(K_0)}_{1NT,\lambda_1}(\boldsymbol{\beta}\ \boldsymbol{\alpha})$ we have $^{(K_0)}_{1NT,\lambda_1}(\hat{\boldsymbol{\beta}}\ \hat{\boldsymbol{\alpha}}) \le \ ^{(K_0)}_{1NT,\lambda_1}(\boldsymbol{\beta}^0\ \hat{\boldsymbol{\alpha}})$ and

$$\hat{\Psi}_i\left(\hat{}_i\ \hat{}_i(\hat{}_i)\right) + \ _1\Pi_{k=1}^{K_0}\left\|\hat{}_i - \hat{}_k\right\| \le \hat{\Psi}_i\left(\ ^0_i\ \hat{}_i(\ ^0_i)\right) + \ _1\Pi_{k=1}^{K_0}\left\|\ ^0_i - \hat{}_k\right\|\ \text{ for }\ = 1$$

Let $\ = \min_i\left[\inf_{\beta_i:\|\beta_i-\beta_i^0\|>\eta}\Psi_i(\ _i\ _i(\ _i)) - \Psi_i(\ ^0_i\ _i(\ ^0_i))\right]\ 0$ Define three events $_1 \equiv \{\max_{1\le i\le N}\sup_{(\beta,\mu)}|\hat{\Psi}_i(\ ) - \Psi_i(\ )| \le \frac{1}{6}\ \}$ and $_2 \equiv \{\max_{1\le i\le N}\sup_{\beta_i}|\Psi_i(\ _i\ \hat{}_i(\ _i)) - \Psi_i(\ _i\ _i(\ _i))| \le \frac{1}{6}\ \}$ and $_3 \equiv \{\ _1\max_{\beta_i:\beta_i\in\mathcal{B}}\Pi_{k=1}^{K_0}\|\ _i - \hat{}_k\| \le \frac{1}{6}\ \}$ By Lemmas S1.3, S1.5(v) and Assumption A2(i), $(\ _1\cap\ _2\cap\ _3) \ge 1 - \ (\ ^c_1) - \ (\ ^c_2) - \ (\ ^c_3) = 1 - \ (\ ^{-1})$ Then conditional on $_1\cap\ _2\cap\ _3$ we have uniformly in

$$\begin{aligned}&\inf_{\beta_i:\|\beta_i-\beta_i^0\|>\eta}\hat{\Psi}_i(\ _i\ \hat{}_i(\ _i)) + \ _1\Pi_{k=1}^{K_0}\|\ _i - \hat{}_k\| \\ \ge\ &\inf_{\beta_i:\|\beta_i-\beta_i^0\|>\eta}\Psi_i(\ _i\ \hat{}_i(\ _i)) - \frac{1}{6}\ + 0 \ge \inf_{\beta_i:\|\beta_i-\beta_i^0\|>\eta}\Psi_i(\ _i\ _i(\ _i)) - \frac{1}{6}\ - \frac{1}{6} \\ \ge\ &\Psi_i(\ ^0_i\ _i(\ ^0_i)) + \ - \frac{1}{6}\ - \frac{1}{6} \\ \ge\ &\Psi_i(\ ^0_i\ \hat{}_i(\ ^0_i)) - \frac{1}{6}\ + \ - \frac{1}{6}\ - \frac{1}{6} \\ \ge\ &\hat{\Psi}_i(\ ^0_i\ \hat{}_i(\ ^0_i)) - \frac{1}{6}\ - \frac{1}{6}\ + \ - \frac{1}{6}\ - \frac{1}{6} \\ =\ &\hat{\Psi}_i(\ ^0_i\ \hat{}_i(\ ^0_i)) + \frac{1}{3} \\ \ge\ &\hat{\Psi}_i(\ ^0_i\ \hat{}_i(\ ^0_i)) + \ _1\Pi_{k=1}^{K_0}\left\|\ ^0_i - \hat{}_k\right\| + \frac{1}{6}\end{aligned}$$

On the other hand, $\hat{\Psi}_i\left(\hat{}_i\ \hat{}_i(\hat{}_i)\right) + {}_1\Pi_{k=1}^{K_0}\left\|\hat{}_i - \hat{}_k\right\| \leq \hat{\Psi}_i\left({}^0_i\ \hat{}_i({}^0_i)\right) + {}_1\Pi_{k=1}^{K_0}\left\|{}^0_i - \hat{}_k\right\|$ It follows that $\left(\max_{1\leq i\leq N}\left\|\hat{}_i - {}^0_i\right\|\right) = (\ )$ ■

To state and prove the next lemma, we follow Hahn and Newey (2004) and introduce some notation. Let ${}_i$ and $\hat{}_i$ denote the cumulative and empirical distribution functions of ${}_{it}$ respectively. Let ${}_i(\ ) \equiv {}_i + \sqrt{\ }(\hat{}_i - {}_i)$ for $\in [0\ \ {}^{-1/2}]$ For fixed ${}_i$ and let ${}_i(\ {}_i\ {}_i(\ )) \equiv \arg\min_{\mu_i} \int (\cdot;\ {}_i\ {}_i)\ {}_i(\ )$ which is the solution to the estimating equation

$$0 = \int {}_i(\cdot;\ {}_i\ {}_i(\ {}_i\ {}_i(\ )))\ {}_i(\ ) \tag{S4}$$

Define ${}^{\beta_i}_i(\ ) = {}_i(\ {}_i\ {}_i(\ ))\ \ {}_i$ Apparently, ${}_i(0) = {}_i\ \ {}_i({}^{-1/2}) = \hat{}_i$

$$\begin{aligned}
{}_i(\ {}_i) &= {}_i(\ {}_i\ {}_i(0)) \\
\hat{}_i(\ {}_i) &= {}_i\left(\ {}_i\ {}_i(\ {}^{-1/2})\right) \\
\frac{{}_i(\ {}_i)}{{}_i} &= \frac{{}_i(\ {}_i\ {}_i(0))}{{}_i} = {}^{\beta_i}_i(0) \quad \text{and} \\
\frac{\hat{}_i(\ {}_i)}{{}_i} &= \frac{{}_i\left(\ {}_i\ {}_i(\ {}^{-1/2})\right)}{{}_i} = {}^{\beta_i}_i(\ {}^{-1/2})
\end{aligned}$$

We study the properties of ${}_i(\ {}_i\ {}_i(\ ))$ and ${}^{\beta_i}_i(\ )$ in the next two lemmas.

(i) $\left(\max_{1\leq i\leq N}\max_{0\leq\epsilon\leq T^{-1/2}}\left|{}_i(\ {}_i\ {}_i(\ )) - {}_i(\ {}_i)\right| \geq\ \right) = {}_P(\ {}^{-1})$ *for any* $\ 0$

(ii) $\max_{1\leq i\leq N, \max\|\beta_i - \beta_i^0\| = o(1)}\left|{}_i(\ {}_i) - {}_i(\ {}^0_i)\right| = (1)$

(iii) $\left(\max_{1\leq i\leq N, \max\|\beta_i - \beta_i^0\| = o(1)}\left|\hat{}_i(\ {}_i) - \hat{}_i(\ {}^0_i)\right| \geq\ \right) = (\ {}^{-1})$ *for any* $\ 0$

(i) Let $\ = \min_i\left[\inf_{\mu_i:|\mu_i - \mu_i(\beta_i)|>\eta}\Psi_i(\ {}_i\ {}_i) - \Psi_i(\ {}_i\ {}_i(\ {}_i))\right]\ \ 0$ Noting that

$$\int (\cdot;\ {}_i\ {}_i)\ {}_i(\ ) = \left(1 - \sqrt{\ }\right)\Psi_i(\ {}_i\ {}_i) + \sqrt{\ }\hat{\Psi}_i(\ {}_i\ {}_i)$$

we have

$$\begin{aligned}
\left|\int (\cdot;\ {}_i\ {}_i)\ {}_i(\ ) - \Psi_i(\ {}_i\ {}_i)\right| &\leq \sqrt{\ }\left|\hat{\Psi}_i(\ {}_i\ {}_i) - \Psi_i(\ {}_i\ {}_i)\right| \\
&\leq \left|\hat{\Psi}_i(\ {}_i\ {}_i) - \Psi_i(\ {}_i\ {}_i)\right|
\end{aligned}$$

By Lemma S1.3, we have $[\ ] = (\ {}^{-1})$ where

$$\equiv \left\{\max_{0\leq\epsilon\leq T^{-1/2}}\max_{1\leq i\leq N}\left|\int (\cdot;\ {}_i\ {}_i)\ {}_i(\ ) - \Psi_i(\ {}_i\ {}_i)\right| \geq\ 3\right\}$$

Therefore for every $\in [0\ \ {}^{-1/2}]$ and conditional on the event we have

$$\begin{aligned}
\inf_{\mu_i:|\mu_i - \mu_i(\beta_i)|>\eta}\int (\cdot;\ {}_i\ {}_i)\ {}_i(\ ) &\geq \inf_{\mu_i:|\mu_i - \mu_i(\beta_i)|>\eta}\Psi_i(\ {}_i\ {}_i) - \frac{1}{3} \\
&\geq \Psi_i(\ {}_i\ {}_i(\ {}_i)) + \frac{2}{3} \\
&\geq \int \Psi_i(\ {}_i\ {}_i(\ {}_i))\ {}_i(\ ) + \frac{1}{3}
\end{aligned}$$

By the triangle inequality,

$$\left\| \; {}_i^{\beta_i}(\cdot;\ {}_i\ {}_i(\ {}_i\ {}_i(\ ))) \ {}_i(\ ) - \; {}_i^{\beta_i}(\cdot;\ {}_i\ {}_i(\ {}_i)) \ {}_i \right\|$$

$$\leq \left\| \; {}_i^{\beta_i}(\cdot;\ {}_i\ {}_i(\ {}_i\ {}_i(\ ))) - {}_i^{\beta_i}(\cdot;\ {}_i\ {}_i(\ {}_i)) \ {}_i(\ ) \right\| + \left\| \; {}_i^{\beta_i}(\cdot;\ {}_i\ {}_i(\ {}_i)) \ [\ {}_i(\ ) - \ {}_i] \right\|$$

$$= \left\| \; {}_i^{\beta_i}(\cdot;\ {}_i\ {}_i(\ {}_i\ {}_i(\ ))) - {}_i^{\beta_i}(\cdot;\ {}_i\ {}_i(\ {}_i)) \ {}_i(\ ) \right\| + \sqrt{\ } \left\| \; {}_i^{\beta_i}(\cdot;\ {}_i\ {}_i(\ {}_i)) \ \hat{}_i - \ {}_i \right\|$$

Using Lemma S1.2(iii), we have

$$\left(\max_{1\leq i\leq N} \max_{0\leq\epsilon\leq T^{-1/2}} \sqrt{\ } \left\| \; {}_i^{\beta_i}(\cdot;\ {}_i\ {}_i(\ {}_i)) \ \hat{}_i - \ {}_i \right\| \geq \ 4 \right) = \ (\ ^{-1})$$

In addition, by Lemma S1.8(i),

$$\left(\max_{1\leq i\leq N} \max_{0\leq\epsilon\leq T^{-1/2}} \left\| \; {}_i^{\beta_i}(\cdot;\ {}_i\ {}_i(\ {}_i\ {}_i(\ ))) - {}_i^{\beta_i}(\cdot;\ {}_i\ {}_i(\ {}_i)) \ {}_i(\ ) \right\| \geq \ 4 \right)$$

$$\leq \left(\max_{1\leq i\leq N} \ (\cdot) \ {}_i(\ ) \max_{1\leq i\leq N} \max_{0\leq\epsilon\leq T^{-1/2}} |\ {}_i(\ {}_i\ {}_i(\ )) - \ {}_i(\ {}_i)| \geq \ 4 \right) = \ (\ ^{-1})$$

Then (S6) follows. Analogously we can prove (S7).

(ii) Recall that

$$\frac{{}_i(\ {}_i)}{{}_i} = -\frac{{}_i^{\beta_i}(\cdot;\ {}_i\ {}_i(\ {}_i)) \ {}_i}{{}_i^{\mu_i}(\cdot;\ {}_i\ {}_i(\ {}_i)) \ {}_i} \tag{S8}$$

To prove (ii), it suffices to show that

$$\max_{1\leq i\leq N, \max\|\beta_i-\beta_i^0\|=o(1)} \left\| \; {}_i^{\beta_i}(\cdot;\ {}_i\ {}_i(\ {}_i)) \ {}_i - \; {}_i^{\beta_i}(\cdot;\ {}_i^0\ {}_i(\ {}_i^0)) \ {}_i \right\| = \ (1)$$

and

$$\max_{1\leq i\leq N, \max\|\beta_i-\beta_i^0\|=o(1)} \left\| \; {}_i^{\mu_i}(\cdot;\ {}_i\ {}_i(\ {}_i)) \ {}_i - \; {}_i^{\mu_i}(\cdot;\ {}_i^0\ {}_i(\ {}_i^0)) \ {}_i \right\| = \ (1)$$

We only show the first result as the proof of the second one is similar. By Assumption A1(iv) and Lemma S1.8(ii),

$$\max_{1\leq i\leq N, \max\|\beta_i-\beta_i^0\|=o(1)} \left\| \; {}_i^{\beta_i}(\cdot;\ {}_i\ {}_i(\ {}_i)) \ {}_i - \; {}_i^{\beta_i}(\cdot;\ {}_i^0\ {}_i(\ {}_i^0)) \ {}_i \right\|$$

$$\leq \max_{i,t} \mathbb{E}[\ (\ {}_{it})] \max_{1\leq i\leq N, \max\|\beta_i-\beta_i^0\|=o(1)} \left\{ \|\ {}_i - \ {}_i^0\| + |\ {}_i(\ {}_i) - \ {}_i(\ {}_i^0)| \right\} = \ (1)$$

(iii) By the triangle inequality,

$$\max_{1\leq i\leq N} \left\| \frac{\hat{}_i(\ {}_i)}{{}_i} - \frac{\hat{}_i(\ {}_i^0)}{{}_i} \right\| \leq \max_{1\leq i\leq N} \left\| \frac{\hat{}_i(\ {}_i)}{{}_i} - \frac{{}_i(\ {}_i)}{{}_i} \right\| + \max_{1\leq i\leq N} \left\| \frac{\hat{}_i(\ {}_i^0)}{{}_i} - \frac{{}_i(\ {}_i^0)}{{}_i} \right\|$$

$$+ \max_{1\leq i\leq N} \left\| \frac{{}_i(\ {}_i)}{{}_i} - \frac{{}_i(\ {}_i^0)}{{}_i} \right\|$$

Noting that $\left(\max_{1\leq i\leq N} \left\| \frac{\partial\hat{\mu}_i(\beta_i)}{\partial\beta_i} - \frac{\partial\mu_i(\beta_i)}{\partial\beta_i} \right\| \geq \ 3 \right) = \ (\ ^{-1})$ by (i) and the last term in the above displayed equation is $\ (1)$ uniformly in the set $\max_i \|\ {}_i - \ {}_i^0\| = \ _P(1)$ by (ii), we have $\left(\max_{1\leq i\leq N, \max\|\beta_i-\beta_i^0\|=o(1)} \left\| \frac{\partial\hat{\mu}_i(\beta_i)}{\partial\beta_i} - \frac{\partial\hat{\mu}_i(\beta_i^0)}{\partial\beta_i} \right\| \geq \ \right) = \ (\ ^{-1})$ for any $\ 0$ ■

Recall from (A.2) that

$$\hat{}_{i\beta\beta}(\ _i) = \frac{1}{\ } \sum_{t=1}^{T} \Big[\ {}^{\beta_i}_{i}(\ {}_{it};\ {}_{i}\ \hat{}_{i}(\ _i)) + \ {}^{\mu_i}_{i}(\ {}_{it};\ {}_{i}\ \hat{}_{i}(\ _i))\ \frac{\hat{}_{i}(\ _i)}{{}'_{i}}\Big]$$

Let $\check{}_{i\beta\beta}(\ _i) = \frac{1}{T} \sum_{t=1}^{T} [\ {}^{\beta_i}_{i}(\ {}_{it};\ {}_{i}\ {}_{i}(\ _i)) + \ {}^{\mu_i}_{i}(\ {}_{it};\ {}_{i}\ {}_{i}(\ _i))\ \frac{\partial\mu_i(\beta_i)}{\partial\beta_i'}]$ Note that $\ {}_{i\beta\beta}(\ _i) = \mathbb{E}[\check{}_{i\beta\beta}(\ _i)]$ where $\ {}_{i\beta\beta}(\cdot)$ is defined in Section 2.3. The next lemma study the asymptotics of $\hat{}_{i\beta\beta}(\ _i)$

(i) $\left(\max_{1\leq i\leq N}\left\|\hat{}_{i\beta\beta}(\check{}_i) - \ {}_{i\beta\beta}(\ {}^{0)}_{i}\right\| \geq\ \right) = \ (\ \ ^{-1})$

(ii) $\ _{\hat{H}} \equiv \min_{1\leq i\leq N}\ {}_{\min}\left(\hat{}_{i\beta\beta}(\breve{}_i)\right) \geq \ {}_{H} - \ {}_{P}(1)$

(i) By the triangle inequality,

$$\begin{aligned}
&\max_{1\leq i\leq N}\left\|\hat{}_{i\beta\beta}(\breve{}_i) - \ {}_{i\beta\beta}(\ {}^{0)}_{i}\right\| \\
\leq\ &\max_{1\leq i\leq N}\left\|\hat{}_{i\beta\beta}(\breve{}_i) - \hat{}_{i\beta\beta}(\ {}^{0)}_{i}\right\| + \max_{1\leq i\leq N}\left\|\hat{}_{i\beta\beta}(\ {}^{0)}_{i} - \check{}_{i\beta\beta}(\ {}^{0)}_{i}\right\| + \max_{1\leq i\leq N}\left\|\check{}_{i\beta\beta}(\ {}^{0)}_{i} - \ {}_{i\beta\beta}(\ {}^{0)}_{i}\right\|
\end{aligned}$$

We prove (i) by showing that (i1) $(\max_{1\leq i\leq N}\|\hat{}_{i\beta\beta}(\breve{}_i) - \hat{}_{i\beta\beta}(\ {}^{0)}_{i}\| \geq\ 3) = \ (\ \ ^{-1})$ (i2) $(\max_{1\leq i\leq N}\|\hat{}_{i\beta\beta}(\ {}^{0)}_{i} - \check{}_{i\beta\beta}(\ {}^{0)}_{i}\| \geq\ 3) = \ (\ \ ^{-1})$ and (i3) $(\max_{1\leq i\leq N}\|\check{}_{i\beta\beta}(\ {}^{0)}_{i} - \ {}_{i\beta\beta}(\ {}^{0)}_{i}\| \geq\ 3) = \ (\ \ ^{-1})$ For (i1), we make the following decomposition:

$$\begin{aligned}
\hat{}_{i\beta\beta}(\breve{}_i) - \hat{}_{i\beta\beta}(\ {}^{0)}_{i} &= \frac{1}{\ } \sum_{t=1}^{T} \ {}^{\beta_i}_{i}(\ {}_{it};\breve{}_i\ \hat{}_{i}(\breve{}_i)) - \ {}^{\beta_i}_{i}(\ {}_{it};\ {}^{0}_{i}\ \hat{}_{i}(\ {}^{0)}_{i}) \\
&+ \frac{1}{\ } \sum_{t=1}^{T} \Big[\ {}^{\mu_i}_{i}(\ {}_{it};\breve{}_i\ \hat{}_{i}(\breve{}_i))\ \frac{\hat{}_{i}(\breve{}_i)}{{}'_{i}} - \ {}^{\mu_i}_{i}(\ {}_{it};\ {}^{0}_{i}\ \hat{}_{i}(\ {}^{0)}_{i})\ \frac{\hat{}_{i}(\ {}^{0)}_{i}}{{}'_{i}}\Big] \\
&\equiv \ {}_{11i} + \ {}_{12i} \text{ say}
\end{aligned}$$

For $\ {}_{11i}$ we have

$$\max_{1\leq i\leq N}\|\ {}_{11i}\| \leq \max_{1\leq i\leq N} \frac{1}{\ } \sum_{t=1}^{T} \ (\ {}_{it}) \left\|\breve{}_i - \ {}^{0}_{i}\right\| + \left\|\hat{}_{i}(\breve{}_i) - \hat{}_{i}(\ {}^{0)}_{i}\right\|$$

Using the arguments as used in the proof of Lemma S1.5(iv), we can show that

$$\left(\max_{1\leq i\leq N} \frac{1}{\ } \sum_{t=1}^{T} \ (\ {}_{it}) \leq 2\ {}^{1/q}_{M}\right) = 1 - \ (\ \ ^{-1})$$

Then by Lemmas S1.7 and S1.8(iii), we can readily show that $(\max_{1\leq i\leq N}\|\ {}_{11i}\| \geq\ 6) = \ (\ \ ^{-1})$ For $\ {}_{12i}$ we make the following decomposition:

$$\begin{aligned}
\ {}_{12i} &= \frac{1}{\ } \sum_{t=1}^{T} \ {}^{\mu_i}_{i}(\ {}_{it};\breve{}_i\ \hat{}_{i}(\breve{}_i)) - \ {}^{\mu_i}_{i}(\ {}_{it};\ {}^{0}_{i}\ \hat{}_{i}(\ {}^{0)}_{i})\ \frac{\hat{}_{i}(\breve{}_i)}{{}'_{i}} \\
&+ \frac{1}{\ } \sum_{t=1}^{T} \ {}^{\mu_i}_{i}(\ {}_{it};\ {}^{0}_{i}\ \hat{}_{i}(\ {}^{0)}_{i}) \left[\frac{\hat{}_{i}(\breve{}_i)}{{}'_{i}} - \frac{\hat{}_{i}(\ {}^{0)}_{i}}{{}'_{i}}\right] \\
&\equiv \ {}_{12i,1} + \ {}_{12i,2} \text{ say.}
\end{aligned}$$

Following the analysis of $_{11i}$ and applying Lemmas S1.8(i) and (iii) and Lemmas S1.9(i) and (iii), we can readily show that $(\max_{1\leq i\leq N} \| \ _{12i,s}\| \geq \ 12) = \ (\ -1)$ for $= 1 \ 2$ Then $(\max_{1\leq i\leq N} \| \ _{12i}\| \geq \ 6) = (\ -1)$ Consequently, we have $\left(\max_{1\leq i\leq N} || \hat{\ }_{i\beta\beta}(\check{\ }_i) - \hat{\ }_{i\beta\beta}(\ ^0_i) || \geq \ 3 \right) = (\ -1)$

To prove (i2), we make the following decomposition:

$$\hat{\ }_{i\beta\beta}(\ ^0_i) - \check{\ }_{i\beta\beta}(\ ^0_i) = \frac{1}{\ } \sum_{t=1}^{T} \ ^{\beta_i}_i(\ _{it}; \ ^0_i \ \hat{\ }_i(\ ^0_i)) - \ ^{\beta_i}_i(\ _{it}; \ ^0_i \ _i(\ ^0_i))$$

$$+\frac{1}{\ } \sum_{t=1}^{T} \left[\ ^{\mu_i}_i(\ _{it}; \ ^0_i \ \hat{\ }_i(\ ^0_i)) \frac{\hat{\ }_i(\ ^0_i)}{\ '_i} - \ ^{\mu_i}_i(\ _{it}; \ ^0_i \ _i(\ ^0_i)) \frac{\ _i(\ ^0_i)}{\ '_i} \right]$$

$$\equiv \ _{21i} + \ _{22i}$$

Following the analysis of $\max_{1\leq i\leq N} || \hat{\ }_{i\beta\beta}(\check{\ }_i) - \hat{\ }_{i\beta\beta}(\ ^0_i) ||$ and using Lemmas S1.2, S1.7, S1.8, and S1.9 and Assumption A1, we can show $(\max_{1\leq i\leq N} \| \ _{2si}\| \geq \ 6) = (\ -1)$ for $= 1 \ 2$. Then (i2) holds.

Next,

$$\check{\ }_{i\beta\beta}(\ ^0_i) - \ _{i\beta\beta}(\ ^0_i) = \frac{1}{\ } \sum_{t=1}^{T} \ ^{\beta_i}_i(\ _{it}; \ ^0_i \ _i(\ ^0_i)) - \mathbb{E} \ ^{\beta_i}_i(\ _{it}; \ ^0_i \ _i(\ ^0_i))$$

$$+\frac{1}{\ } \sum_{t=1}^{T} \{ \ ^{\mu_i}_i(\ _{it}; \ ^0_i \ _i(\ ^0_i)) - \mathbb{E}[\ ^{\mu_i}_i(\ _{it}$$

Let 0 be an arbitrary constant. By Theorem 2.2, $(\|\hat{}_{k,1}\| \geq) \leq (\hat{}_{kNT}) \to 0$ and $(\|\hat{}_{k,2}\| \geq) \leq (\hat{}_{kNT}) \to 0$ Thus $\hat{}_{\hat{G}_k} = \hat{}_{G_k^0} + {}_P(1)$ and it suffices to prove the lemma by showing that (i) $\hat{}_{G_k^0} + \mathbb{B}_{kNT} = \frac{1}{\sqrt{N_k T}} \sum_{i \in G_k^0} \sum_{t=1}^{T} \mathbb{U}_{it} + {}_P(1)$ and (ii) $\frac{1}{\sqrt{N_k T}} \sum_{i \in G_k^0} \sum_{t=1}^{T} \mathbb{U}_{it} \xrightarrow{D} (0\ \Omega_k)$

We prove $\hat{}_{G_k^0} + \mathbb{B}_{kNT} = \frac{1}{\sqrt{N_k T}} \sum_{i \in G_k^0} \sum_{t=1}^{T} \mathbb{U}_{it} + {}_P(1)$ By second order Taylor expansion,

$$\begin{aligned} \hat{}_{G_k^0} &= \frac{1}{\sqrt{\ \ _k}} \sum_{i \in G_k^0} \sum_{t=1}^{T} {}_{it} + \frac{1}{\sqrt{\ \ _k}} \sum_{i \in G_k^0} \sum_{t=1}^{T} {}_{it}^{\mu_i} [\hat{}_i(\ _k^0) - \ _i^0] \\ &\quad + \frac{1}{2\sqrt{\ \ _k}} \sum_{i \in G_k^0} \sum_{t=1}^{T} {}_{i}^{\mu_i \mu_i} (\ _{it};\ _k^0\ _i^*) [\hat{}_i(\ _k^0) - \ _i^0]^2 \\ &\equiv \ _{k,1} + \ _{k,2} + \ _{k,3} \text{ say,} \end{aligned} \tag{S9}$$

U $UQV!Q p"d6a6€ 7'7' G6 X , s&FC‰ SÀ añ •#(h,$P>

where $_i^*$ lies between $\hat{}_i(\ _k^0)$ and $_i^0$ We will show that $_{k,1}$ contributes to the asymptotic variance of $\hat{}_{G_k^0}$ $_{k,3}$ contributes to the asymptotic bias, and $_{k,2}$ contributes to both. We analyze $_{k,3}$ first. Let $_{k,3}^0 = \frac{1}{2\sqrt{N_k T}} \sum_{i \in G_k^0} \sum_{t=1}^{T} {}_{it}^{\mu_i \mu_i} [\hat{}_i(\ _k^0) - \ _i^0]^2$ By Assumption A1, the Markov inequality, and Lemma S1.5(ii), we have

$$\begin{aligned} \|\ _{k,3} - \ _{k,3}^0\| &= \frac{1}{2\sqrt{\ \ _k}} \sum_{i \in G_k^0} \sum_{t=1}^{T} \left\| {}_i^{\mu_i \mu_i} (\ _{it};\ _k^0\ _i^*) - {}_i^{\mu_i \mu_i} (\ _{it};\ _k^0\ _i^0) \right\| [\hat{}_i(\ _k^0) - \ _i^0]^2 \\ &\leq \left\{ \frac{1}{2\ \ _k} \sum_{i \in G_k^0} \sum_{t=1}^{T} (\ _{it}) \right\} \sqrt{\ \ _k} \end{aligned}$$

04790D0019151f70200702380045604 6m(2)805D13141f1002001

Now, we study $_{k,2}$ By Lemma S1.5(ii), (S1) in its proof, and the fact that $\max_{1\le i\le N}\left|\frac{1}{T}\sum_{t=1}^{T}\ _{it}\right| = {}_P(\ ^{-1/2}(\ln\ )^3)$ and $\max_{1\le i\le N}\left|\frac{1}{T}\sum_{t=1}^{T}\ ^{\mu_i}_{it} - \ _{iV}\right| = {}_P(\ ^{-1/2}(\ln\ )^3)$ we have

$$\hat{\ }_i(\ ^0_k) - \ ^0_i = -\frac{\frac{1}{T}\sum_{t=1}^{T}\ _{it}}{\frac{1}{T}\sum_{t=1}^{T}\ ^{\mu_i}_i(\ _{it};\ ^0_k\ \check{\ }_i(\ ^0_k))} = -\frac{\frac{1}{T}\sum_{t=1}^{T}\ _{it}}{\frac{1}{T}\sum_{t=1}^{T}\ ^{\mu_i}_{it}} + {}_P(\ ^{-1}(\ln\ )^6)$$

$$= -\ ^{-1}_{iV}\frac{1}{\ }\sum_{t=1}^{T}\ _{it} + {}_P(\ ^{-1}(\ln\ )^6) \text{ uniformly in } \ \in\ ^0_k$$

But the above expansion is not sufficient to study $_{k,2}$ and we need to get better control on the remainder term. Noting that $\hat{\ }_i(\ ^0_i) = \arg\min_{\mu_i}\frac{1}{T}\sum_{t=1}^{T}(\ _{it};\ ^0_i\ _i)$ we have

$$0 = \frac{1}{\ }\sum_{t=1}^{T}\ _i(\ _{it};\ ^0_i\ \hat{\ }_i(\ ^0_i))$$

$$= \frac{1}{\ }\sum_{t=1}^{T}\ _{it} + \frac{1}{\ }\sum_{t=1}^{T}\ ^{\mu_i}_{it}[\hat{\ }_i(\ ^0_i) - \ _i(\ ^0_i)] + \frac{1}{2\ }\sum_{t=1}^{T}\ ^{\mu_i\mu_i}_i(\ _{it};\ ^0_i\ \check{\ }_i(\ ^0_i))[\hat{\ }_i(\ ^0_i) - \ _i(\ ^0_i)]^2$$

where $\check{\ }_i(\ ^0_i)$ lies between $\hat{\ }_i(\ ^0_i)$ and $_i(\ ^0_i)$ for each It follows that

$$\hat{\ }_i(\ ^0_i) - \ _{i0}(\ ^0_i)$$

$$= -\left[\frac{1}{\ }\sum_{t=1}^{T}\ ^{\mu_i}_{it}\right]^{-1}\left\{\frac{1}{\ }\sum_{t=1}^{T}\ _{it} + \frac{1}{2\ }\sum_{t=1}^{T}\ ^{\mu_i\mu_i}_i(\ _{it};\ ^0_i\ \check{\ }_i(\ ^0_i))[\hat{\ }_i(\ ^0_i) - \ _i(\ ^0_i)]^2\right\}$$

$$= -\left[\frac{1}{\ }\sum_{t=1}^{T}\ ^{\mu_i}_{it}\right]^{-1}\left\{\frac{1}{\ }\sum_{t=1}^{T}\ _{it} + \frac{1}{2\ }\sum_{t=1}^{T}\ ^{\mu_i\mu_i}_{it}[\hat{\ }_i(\ ^0_i) - \ _i(\ ^0_i)]^2\right\} + {}_P(\ ^{-3}(\ln\ )^9)$$

$$= -\left[\frac{1}{\ }\sum_{t=1}^{T}\ ^{\mu_i}_{it}\right]^{-1}\left\{\frac{1}{\ }\sum_{t=1}^{T}\ _{it} + \frac{1}{2}\ ^{-2}_{iV}\frac{1}{\ }\sum_{t=1}^{T}\ ^{\mu_i\mu_i}_{it}\left(\frac{1}{\ }\sum_{t=1}^{T}\ _{it}\right)^2\right\} + {}_P(\ ^{-3}(\ln\ )^9)$$

$$= -\left[\frac{1}{\ }\sum_{t=1}^{T}\ ^{\mu_i}_{it}\right]^{-1}\left\{\frac{1}{\ }\sum_{t=1}^{T}\ _{it} + \frac{1}{2}\ ^{-2}_{iV}\ _{iV2}\left(\frac{1}{\ }\sum_{t=1}^{T}\ _{it}\right)^2\right\} + {}_P(\ ^{-3}(\ln\ )^9) \tag{S11}$$

where we use the fact $\max_{1\le i\le N}\left|\frac{1}{T}\sum_{t=1}^{T}[\ ^{\mu_i\mu_i}_i(\ _{it};\ _i\ \check{\ }_i(\ ^0_i)) - \ ^{\mu_i\mu_i}_{it}]\right| \le \max_{1\le i\le N}\frac{1}{T}\sum_{t=1}^{T}(\ _{it})$ $\times\max_{1\le i\le N}\left|\check{\ }_i(\ ^0_i) - \ _i(\ ^0_i)\right| = {}_P(\ ^{-1/2}(\ln\ )^3)$ and $\max_{1\le i\le N}\left|\frac{1}{T}\sum_{t=1}^{T}\ ^{\mu_i\mu_i}_{it} - \ _{iV2}\right| = {}_P(\ ^{-1/2}(\ln\ )^3)$ by Lemma S1.2(i). It follows that

$$_{k,2} = \frac{-1}{\sqrt{\ _k}}\sum_{i\in G^0_k}\sum_{t=1}^{T}\ ^{\mu_i}_{it}\left\{\left[\frac{1}{\ }\sum_{t=1}^{T}\ ^{\mu_i}_{it}\right]^{-1}\left\{\frac{1}{\ }\sum_{t=1}^{T}\ _{it} + \frac{1}{2}\ ^{-2}_{iV}\ _{iV2}\left(\frac{1}{\ }\sum_{t=1}^{T}\ _{it}\right)^2\right\} + {}_P(\ ^{-3}(\ln\ )^9)\right\}$$

$$= -\frac{1}{\sqrt{\ _k}}\sum_{i\in G^0_k}\sum_{t=1}^{T}\ _{it}\frac{\frac{1}{T}\sum_{t=1}^{T}\ ^{\mu_i}_{it}}{\frac{1}{T}\sum_{t=1}^{T}\ ^{\mu_i}_{it}}$$

$$-\frac{1}{2\sqrt{\ _k}}\sum_{i\in G^0_k}\sum_{t=1}^{T}\ ^{\mu_i}_{it}\frac{\ ^{-2}_{iV}\ _{iV2}\left(\frac{1}{T}\sum_{t=1}^{T}\ _{it}\right)^2}{\frac{1}{T}\sum_{t=1}^{T}\ ^{\mu_i}_{it}} + {}_P(1)$$

$$\equiv -\ _{k,21} - \ _{k,22} + {}_P(1) \tag{S12}$$

For $_{k,21}$ we make the [redacted]wing decomposition:

$$
\begin{aligned}
{}_{k,21} &= \frac{1}{\sqrt{\quad k}} \sum_{i \in G_k^0} \sum_{t=1}^{T} {}_{it} \frac{{}_{iU}}{{}_{iV}} + \frac{1}{\sqrt{\quad k}} \sum_{i \in G_k^0} \sum_{t=1}^{T} {}_{it} \frac{\frac{1}{T} \sum_{t=1}^{T} ({}^{\mu_i}_{it} - {}_{iU})}{{}_{iV}} \\
&+ \frac{1}{\sqrt{\quad k}} \sum_{i \in G_k^0} \sum_{t=1}^{T} {}_{it} \, {}_{iU} \left\{ \frac{1}{\frac{1}{T} \sum_{t=1}^{T} {}^{\mu_i}_{it}} - \frac{1}{{}_{iV}} \right\} \\
&+ \frac{1}{\sqrt{\quad k}} \sum_{i \in G_k^0} \sum_{t=1}^{T} {}_{it} \frac{1}{} \sum_{t=1}^{T} ({}^{\mu_i}_{it} - {}_{iU}) \left\{ \frac{1}{\frac{1}{T} \sum_{t=1}^{T} {}^{\mu_i}_{it}} - \frac{1}{{}_{iV}} \right\} \\
&\equiv {}_{k,21a} + {}_{k,21b} + {}_{k,21c} + {}_{k,21d}
\end{aligned}
\tag{S13}
$$

Apparently, $_{k,21b} = \frac{1}{\sqrt{N_k T^3}} \sum_{i \in G_k^0} {}_{iV}^{-1} \sum_{s=1}^{T} \sum_{t=1}^{T} {}_{is} [{}^{\mu_i}_{it} - \mathbb{E}({}^{\mu_i}_{it})]$ For $_{k,21c}$ we can use the fact that $\max_{1 \le i \le N} \left| \frac{1}{T} \sum_{t=1}^{T} {}_{it} \right| = {}_P \left({}^{-1/2} (\ln \quad)^3 \right)$ and $\max_{1 \le i \le N} \left| \frac{1}{T} \sum_{t=1}^{T} {}^{\mu_i}_{it} - {}_{iV} \right| = {}_P \left({}^{-1/2} (\ln \quad)^3 \right)$ to show that

$_{k,21c}$ b e„4 Q‘ G÷á”v%

Combining (S12)-(S17) yields

$$
\begin{aligned}
{}_{k,2} &= -\frac{1}{\sqrt{\quad k}} \sum_{i \in G_k^0} \sum_{t=1}^{T} {}_{it} \frac{{}_{iU}}{{}_{iV}} \\
&\quad - \left\{ \frac{1}{\sqrt{\quad k}^{\,3}} \sum_{i \in G_k^0} {}_{iV}^{-1} \sum_{t=1}^{T} \sum_{s=1}^{T} \right.
\end{aligned}
$$

second derivatives of $\mathbb{U}_i$ with respect to $\ _i$ Let $\mathbb{U}_{it}^{\mu_i} = \mathbb{U}_i^{\mu_i}(\ _{it};\ ^0_i\ ^0_i)$ and $\mathbb{U}_{it}^{\mu_i\mu_i} = \mathbb{U}_i^{\mu_i\mu_i}(\ _{it};\ ^0_i\ ^0_i)$ Following HK, the asymptotic bias term of $\hat{\ }_{\hat{G}_k}$ takes the form:

$$\mathbb{B}_{kNT}^{HK} = \frac{1}{\sqrt{\ _k}} \sum_{i \in G_k^0} \left[\ _{iV}^{-1} \frac{1}{\sqrt{\ }} \sum_{t=1}^{T} \ _{it} \right] \left[\frac{1}{\sqrt{\ }} \sum_{t=1}^{T} \left(\mathbb{U}_{it}^{\mu_i} - \frac{\ _{i\mathbb{U}2}}{2\ _{iV}} \ _{it} \right) \right]$$

where $\ _{i\mathbb{U}2} \equiv \frac{1}{T} \sum_{t=1}^{T} \mathbb{E}(\mathbb{U}_{it}^{\mu_i\mu_i})$ Note that

$$\mathbb{B}_{kNT}^{HK} = \frac{1}{\sqrt{\ _k}^{3}} \sum_{i\in G_k^0} \ _{iV}^{-1} \sum_{s=1}^{T} \sum_{t=1}^{T} \ _{is} \mathbb{U}_{it}^{\mu_i} - \frac{1}{2\sqrt{\ _k}} \sum_{i\in G_k^0} \ _{iV}^{-2} \ _{i\mathbb{U}2} \left(\frac{1}{\sqrt{\ }} \sum_{t=1}^{T} \ _{it} \right)^2$$

$$\equiv \mathbb{B}_{1kNT}^{HK} - \mathbb{B}_{2kNT}^{HK} \text{ say.}$$

Let $\mathbb{B}_{1kNT}$ and $\mathbb{B}_{2kNT}$ be as defined in Theorem 2.4. Apparently, $\mathbb{B}_{1kNT}^{HK} = \mathbb{B}_{1kNT}$ Noting that $\ _{i\mathbb{U}2} = \frac{1}{T} \sum_{t=1}^{T} \mathbb{E}(\ ^{\mu_i\mu_i}_{it} - \frac{m_{iU}}{m_{iV}}\ ^{\mu_i\mu_i}_{i}) = \ _{iU2} - \frac{m_{iU}}{m_{iV}}\ _{iV2}$ with $\ _{iU2} = \frac{1}{T} \sum_{t=1}^{T} \mathbb{E}(\ ^{\mu_i\mu_i}_{it})$ we have

$$\mathbb{B}_{2kNT}^{HK} = \frac{1}{2\sqrt{\ _k}} \sum_{i\in G_k^0} \ _{iV}^{-2} \left(\ _{iU2} - \frac{\ _{iU}}{\ _{iV}} \ _{iV2} \right) \left(\frac{1}{\sqrt{\ }} \sum_{t=1}^{T} \ _{it} \right)^2 = \mathbb{B}_{2kNT}$$

It follows that $\mathbb{B}_{kNT}^{HK} = \mathbb{B}_{kNT}$

Let $\hat{\ }_{(k)} \equiv \frac{1}{N_k T} \sum_{i\in \hat{G}_k} \sum_{t=1}^{T} [\ ^{\beta_i}_i(\ _{it};\breve{\ }_k\ \hat{\ }_i(\breve{\ }_k)) + \ ^{\mu_i}_i(\ _{it};\breve{\ }^0_k\ \hat{\ }_i(\breve{\ }_k)) \frac{\partial \hat{\mu}_i(\breve{\alpha}_k)}{\partial \alpha_k'}]$ *and* $\breve{\ }_k$ *lying between* $\hat{\ }_k$ *and* $\ ^0_k$ *elementwise. Then* $\hat{\ }_{(k)} = \mathbb{H}_{kNT} + \ _P(\ _{NT})$ *where* $\ _{NT} = \min(1\ \sqrt{\ _k})$

As in the proof of Lemma S1.12, we can readily show that $\hat{\ }_{(k)} = \hat{\ }_{G_k^0} + \ _P(1)$ where $\hat{\ }_{G_k^0} = \frac{1}{N_k T} \sum_{i\in G_k^0} \sum_{t=1}^{T} [\ ^{\beta_i}_i(\ _{it};\breve{\ }_k\ \hat{\ }_i(\breve{\ }_k)) + \ ^{\mu_i}_i(\ _{it};\breve{\ }^0_k\ \hat{\ }_i(\breve{\ }_k)) \frac{\partial \hat{\mu}_i(\breve{\alpha}_k)}{\partial \alpha_k'}]$ ... $\hat{\ }_{G_k^0}$ ... lowing decomposition

$$\hat{\ }_{G_k^0} \equiv \frac{1}{\ _k} \sum_{i\in G_k^0} \sum_{t=1}^{T} \left[\ ^{\beta_i}_i(\ _{it};\ ^0_k\ \ _i(\ ^0_k)) + \ ^{\mu_i}_i(\ _{it};\ ^0_k\ \ _i(\ ^0_k)) \ldots \right.$$

It follows that $\hat{\ }_{(k)} = \mathbb{H}_{kNT} + \ _{P}($

As before, we classify $\in \hat{}_k(\quad {}_1)$ if $\| \hat{}_i - \hat{}_k \| = 0$ for $= 1$ and $\in \hat{}_0(\quad {}_1)$ otherwise. Suppose that $\in \hat{}_k(\quad {}_1)$ for $\in \{ \ {}_0 + 1 \quad \}$ Then by the pointwise consistency of $\hat{}_i$ we know that the probability limit of $\hat{}_k$ must be given by one of the columns in $\boldsymbol{\alpha}^0 = ({}^0_1 \quad {}^0_{K_0})$ and it converges in probability to the true value at the rate ${}^{-1/2} + {}_1$ Apparently, if $\mathcal{C}$ contains ${}_k$ elements with probability limit given by ${}^0_k$ we can derive that $\| \hat{}_k - {}^0_k \| = {}_P(\min({}^{-1/(2n_k)} \quad {}^{-1/2} + {}_1))$ for $= 1 \quad {}_0$ Without loss of generality, assume that if

By Lemma S1.12 and the fact that $\mathbb{B}_k = \quad_P((\quad)^{1/2})$ we can show that $\quad_{1NT,2} = \quad_P(\quad_{NT}^{-2})$ Let $\bar{\quad}_{1NT,3} \equiv \frac{1}{NT} \sum_{k=1}^{K_0} \sum_{i \in G_k^0} \sum_{t=1}^{T} \quad_i(\quad_{it}; \quad_k^0 \hat{\quad}_i(\quad_k^0)) \frac{\partial \hat{\mu}_i(\alpha_k^0)}{\partial \alpha_k}$ Then

$$
\begin{aligned}
\bar{\quad}_{1NT,3} &= \frac{1}{\quad} \sum_{i=1}^{N} \sum_{t=1}^{T} \quad_{it} \frac{\quad_i(\quad_i^0)}{\quad_i} + \frac{1}{\quad} \sum_{i=1}^{N} \sum_{t=1}^{T} \quad_{it} \left(\frac{\hat{\quad}_i(\quad_i^0)}{\quad_i} - \frac{\quad_i(\quad_i^0)}{\quad_i} \right) \\
&\quad + \frac{1}{\quad} \sum_{i=1}^{N} \sum_{t=1}^{T} \quad_{it}^{\mu_i} (\hat{\quad}_i(\quad_i^0) - \quad_i^0) \frac{\hat{\quad}_i(\quad_k^0)}{\quad_k} + \quad_P(\quad^{-1}) \\
&\equiv \bar{\quad}_{1NT,31} + \bar{\quad}_{1NT,32} + \bar{\quad}_{1NT,33} + \quad_P(\quad^{-1})
\end{aligned}
$$

By the Chebyshev and Davydov inequalities, we can readily show that $\bar{\quad}_{1NT,31} = \quad_P((\quad)^{-1/2})$ By (S5),

$$
\begin{aligned}
\frac{\hat{\quad}_i(\quad_i^0)}{\quad_i} - \frac{\quad_i(\quad_i^0)}{\quad_i} &= \frac{\quad_i(\quad_i^0 \quad_i(\quad^{-1/2}))}{\quad_i} - \frac{\quad_i(\quad_i^0 \quad_i(0))}{\quad_i} \\
&= \frac{\sum \quad_i^{\beta_i}(\cdot; \quad_i^0 \quad_i^0) \quad_i}{\sum \quad_i^{\mu_i}(\cdot; \quad_i^0 \quad_i^0) \quad_i} - \frac{\sum \quad_i^{\beta_i}(\cdot; \quad_i^0 \hat{\quad}_i(\quad_i^0)) \hat{\quad}_i}{\sum \quad_i^{\mu_i}(\cdot; \quad_i^0 \hat{\quad}_i(\quad_i^0)) \hat{\quad}_i} \\
&= \frac{\quad_{iV}}{\quad_{iV}} - \frac{\hat{\quad}_{iV}}{\hat{\quad}_{iV}} = \frac{\quad_{iV} \hat{\quad}_{iV} - \hat{\quad}_{iV} \quad_{iV}}{\quad_{iV} \hat{\quad}_{iV}} \\
&= \frac{\quad_{iV}(\hat{\quad}_{iV} - \quad_{iV}) + (\quad_{iV} - \hat{\quad}_{iV}) \quad_{iV}}{\quad_{iV} \hat{\quad}_{iV}} \qquad \text{(S20)}
\end{aligned}
$$

where $\quad_{iV} \equiv \sum \quad_i^{\mu_i}(\cdot; \quad_i^0 \quad_i^0) \quad_i$ $\hat{\quad}_{iV} \equiv \sum \quad_i^{\beta_i}(\cdot; \quad_i^0 \hat{\quad}_i(\quad_i^0)) \hat{\quad}_i$ $\hat{\quad}_{iV} \equiv \sum \quad_i^{\mu_i}(\cdot; \quad_i^0 \hat{\quad}_i(\quad_i^0)) \hat{\quad}_i$ and recall $\quad_{iV} \equiv \sum \quad_i^{\mu_i}(\cdot; \quad_i^0 \quad_i^0) \quad_i$ Then by (S11) and Lemma S1.2(i), we can show that

$$
\bar{\quad}_{1NT,32} = \frac{1}{\quad} \sum_{i=1}^{N} \sum_{t=1}^{T} \quad_{it} \quad_{iV}^{-2} \quad_{iV} (\hat{\quad}_{iV} - \quad_{iV}) - \frac{1}{\quad} \sum_{i=1}^{N} \sum_{t=1}^{T} \quad_{it} \quad_{iV}^{-1} (\hat{\quad}_{iV} - \quad_{iV}) + \quad_P(\quad)
$$

For $\bar{\ }_{1NT,33}$ using (S11), (S20), and Lemma S1.2(i), we can readily show that

$$
\begin{aligned}
\bar{\ }_{1NT,33} &= \frac{1}{\ } \sum_{i=1}^{N} \sum_{t=1}^{T} {}_{it}^{\mu_i} (\hat{\ }_i(\ _i^0) - \ _i^0) \frac{\ _i(\ _k^0)}{\ _k} + \ _P(\ _{NT}^{-1}) \\
&= \frac{1}{\ } \sum_{i=1}^{N} \sum_{t=1}^{T} \ _{it} \frac{\ _i(\ _k^0)}{\ _k} + \ _P(\ _{NT}^{-1}) \\
&= \ _P\big((\ \ )^{-1/2}\big) + \ _P(\ _{NT}^{-1}) = \ _P(\ _{NT}^{-1})
\end{aligned}
$$

Then $\bar{\ }_{1NT,3} = \ _P(\ _{NT}^{-1})$ and $\ _{1NT,3} = \ _P(\ _{NT}^{-2})$

By Taylor expansion,

$$
\begin{aligned}
\ _{1NT,1} - \ _{G^0}^{-2} &= \frac{2}{\ } \sum_{i=1}^{N} \sum_{t=1}^{T} (\ _{it}; \ _i^0 \ \hat{\ }_i(\ _i^0)) - \ _{G^0}^{-2} \\
&= \frac{2}{\ } \sum_{i=1}^{N} \sum_{t=1}^{T} \ _i(\ _{it}; \ _i^0 \ _i^0) [\hat{\ }_i(\ _i^0) - \ _i^0] + \frac{2}{\ } \sum_{i=1}^{N} \sum_{t=1}^{T} \ _i^{\mu_i}(\ _{it}; \ _i^0 \ \breve{\ }_i(\ _i^0)) [\hat{\ }_i(\ _i^0) - \ _i^0]^2 \\
&\equiv \ _{1NT,11} + \ _{1NT,12}
\end{aligned}
$$

Using (S11), we can readily show that $\ _{1NT,11} = \ _P(\ ^{-1})$ and $\ _{1NT,12} = \ _P(\ ^{-1})$ Then $\ _{1NT} = \ _{G^0}^{-2} + \ _P(\ ^{-1})$ It follows that $\hat{\ }^2_{\hat{G}(K,\lambda_1)} - \ _{G^0}^{-2} = \ _P(\ ^{-1})$ for each $\ _0 \leq \ \leq \ _{\max}$ ■

Other choices of kernels are possible. So the bias-corrected PLS C-Lasso estimator is given by

$$\hat{\ }^{(c)}_{k} = \hat{\ }_{k} - \frac{1}{\hat{\ }_{k}} \mathbb{H}^{-1}_{kNT} \hat{\mathbb{B}}_{1kNT}$$

Similarly, we can obtain the bias-corrected estimator for the post-Lasso estimator $\hat{\ }_{\hat{G}_k}$

Let $_{i} \equiv (\ _{i1} \quad _{iT})'$ and $_{i} \equiv (\ _{i1} \quad _{iT})'$ Let $\| \quad \|_a = \{\mathbb{E}\| \quad \|^a\}^{1/a}$ for any ≥ 1 Let denote a generic positive constant that does not depend on and We add the following assumption.

ASSUMPTION D1. *(i) For each* $= 1$ $\{(\ _{it} \ _{it}) : \ = 1 \ 2 \quad \}$ *is strong mixing with mixing coefficients* $\{\ _i(\cdot)\}$ *such that* $_i(\ ) \leq \ _{\alpha,i} \ ^{\tau}$ *for some* $_{\alpha,i} \quad \infty$ *and* $\in (0 \ 1)$ $\frac{1}{N_k} \sum_{i \in G^0_k} \ ^{(2q-1)/q}_{\alpha,i} = \quad (1)$

(ii) $(\ _i \ _i)$ *are independent across* $\in \ ^0_k$ *where* $= 1 \quad _0$

(iii) $\max_{i,t} \mathbb{E}\| \ _{it}\|^{4q} \quad \infty$ *and* $\max_{i,t} \mathbb{E}\| \ _{it}\|^{4q} \quad \infty$ *for some* ≥ 1

(iv) As $(\quad) \to \infty \quad _T \to \infty \quad \frac{2}{T} \quad \to 0 \quad \frac{2}{T} \ _k \quad ^3 \to 0$ *and* $\ _k^{-1/2} \ ^{1/2} \sum_{i \in G^0_k} \ _i(\ \ _T)^{\frac{2q-1}{2q}} \to 0$ *for each* $= 1 \quad _0$

Assumption D1(i) assumes the usual mixing condition. D1(ii) assumes cross sectional independence to simplify the proof which can be relaxed at the cost of lengthy arguments. D1(iii) assumes moment conditions. The last condition in D1(iv) can be easily ensured under D1(i) because for any $_T \gg -\frac{2q}{(2q-1)\ln q} \ln(\ ^{1/2} \ ^{1/2})$ (e.g., $_T = ^{(}\ln(\ ^{1/2} \ ^{1/2})^{)}\ ^{1+\epsilon}$ for some $\quad 0$), we have

$$\begin{aligned} \ _k^{-1/2} \ ^{1/2} \sum_{i \in G^0_k} \ _i(\ \ _T)^{(2q-1)/(2q)} &\leq \left(\ _k^{-1} \sum_{i \in G^0_k} \ ^{(2q-1)/(2q)}_{\alpha,i} \right) \ ^{1/2}_k \ ^{1/2} \ ^{M_T(2q-1)/(2q)} \\ &= \quad (1) \exp\left(\ln\left(\ ^{1/2}_k \ ^{1/2}\right) + \frac{(2 \ -1) \quad _T}{2} \ln \ \right) \to 0 \end{aligned}$$

The first three requirements in D1(iv) can be easily satisfied too. For example, if $_k \quad ^a$ for some $\quad 3$ it suffices to set $_T \quad ^{1/b}$ for some $\quad \max\{2 \ 2 \ (3 - \ )\}$

Suppose Assumption D1 holds. Then $\mathbb{H}^{-1}_{kNT} \hat{\mathbb{B}}_{1kNT} - \mathbb{H}^{-1}_{kNT} \mathbb{B}_{1kNT} = \ _P(1)$

Noting that $\mathbb{H}^{-1}_{kNT} \hat{\mathbb{B}}_{1kNT} - \mathbb{H}^{-1}_{kNT} \mathbb{B}_{1kNT} = (\mathbb{H}^{-1}_{kNT} - \mathbb{H}^{-1}_{k})\mathbb{B}_{1kNT} + (\mathbb{H}^{-1}_{kNT} - \mathbb{H}^{-1}_{kNT})(\hat{\mathbb{B}}_{1kNT} - \mathbb{B}_{1kNT}) + \mathbb{H}^{-1}_{kNT}(\hat{\mathbb{B}}_{1kNT} - \mathbb{B}_{1kNT})$ $\mathbb{H}^{-1}_{kNT} = \quad (1)$ and $\mathbb{B}_{1kNT} = \ _P(\ \overline{\ _k}\ )$ it suffices to show that (i) $\mathbb{H}_{kNT} - \mathbb{H}_{kNT} = \ _P(\ _{NT})$ and (ii) $\hat{\mathbb{B}}_{1kNT} - \mathbb{B}_{1kNT} = \ _P(1)$ where $_{NT} = \min(1 \quad \overline{\ _k})$

We first prove (i). Let $\bar{\mathbb{H}}_{kNT} \equiv \frac{1}{N_k T} \sum_{i \in G^0_k} \sum_{t=1}^{T} \tilde{\ }_{it} \tilde{\ }'_{it}$ It suffices to show that (i1) $\mathbb{H}_{kNT} - \bar{\mathbb{H}}_{kNT} = \ _P(\ _{NT})$ and (i2) $\bar{\mathbb{H}}_{kNT} - \mathbb{H}_{kNT} = \ _P(\ _{NT})$ Note that

$$\begin{aligned} \mathbb{H}_{kNT} - \bar{\mathbb{H}}_{kNT} &= \frac{1}{\hat{\ }_k} \sum_{i \in \hat{G}_k} \sum_{t=1}^{T} \tilde{\ }_{it} \tilde{\ }'_{it} - \frac{1}{\ _k} \sum_{i \in G^0_k} \sum_{t=1}^{T} \tilde{\ }_{it} \tilde{\ }'_{it} \\ &= \frac{1}{\hat{\ }_k} \left(\sum_{i \in \hat{G}_k} - \sum_{i \in G^0_k} \right) \sum_{t=1}^{T} \tilde{\ }_{it} \tilde{\ }'_{it} + \frac{\ _k - \hat{\ }_k}{\hat{\ }_k \ _k} \sum_{i \in G^0_k} \sum_{t=1}^{T} \tilde{\ }_{it} \tilde{\ }'_{it} \\ &\equiv \ _{k,1} + \ _{k,2} \end{aligned}$$

By Corollary 2.3, we can readily show that $_{k,2} = \ _P(\ _k^{-1}) = \ _P(\ _{NT})$ For any $\quad 0$ we have by the proof of Theorem 2.2, $(\| \ _{k,1}\| \geq \ _{NT}\ ) \leq (\ \hat{\ }_{kNT}) + (\ \hat{\ }_{kNT}) = \quad (1)$ It follows that $\mathbb{H}_{kNT} - \bar{\mathbb{H}}_{kNT} =$

$_P(\ \ _{NT})$ Now,

$$\begin{aligned}\bar{\mathbb{H}}_{kNT} - \mathbb{H}_{kNT} &= \frac{1}{\ \ _k} \sum_{i \in G_k^0} \sum_{t=1}^{T} \{ \tilde{\ }_{it} \tilde{\ }_{it}' - \mathbb{E}\{ [\ \ _{it} - \mathbb{E}(\bar{\ }_{i\cdot})] [\ \ _{it} - \mathbb{E}(\bar{\ }_{i\cdot})]' \} \} \\ &= \frac{1}{\ \ _k} \sum_{i \in G_k^0} \sum_{t=1}^{T} \{ [\ \ _{it} - \mathbb{E}(\bar{\ }_{i\cdot})] [\ \ _{it} - \mathbb{E}(\bar{\ }_{i\cdot})] - \mathbb{E}\{ [\ \ _{it} - \mathbb{E}(\bar{\ }_{i\cdot})] [\ \ _{it} - \mathbb{E}(\bar{\ }_{i\cdot})]' \} \\ &\quad + \frac{1}{\ \ _k} \sum_{i \in G_k^0} \sum_{t=1}^{T} \{ \tilde{\ }_{it} \tilde{\ }_{it}' - [\ \ _{it} - \mathbb{E}(\bar{\ }_{i\cdot}\end{aligned}$$

For the first term, we can apply Lemma A.2(ii) in Gao (2007) and show that it is $(\ \ ^{-1})$ For the second term, we can apply the Davydov inequality directly to show that it is bounded from above by

$$\frac{2}{\ _k\ ^3} \sum_{i\in G_k^0} \left(8 \ \| \ _{it}\|_{4q} \| \ _1' \ _{is}\|_{4q} \sum_{s=1}^{T} \ _i(\ )^{(2q-1)/(2q)} \right)^2 = (\ \ ^{-1})$$

It follows that $\mathbb{B}_{1kNT} - \bar{\mathbb{B}}_{1kNT} = (\ \ ^{-1/2}) = \ _P(1)$

We now show (ii2), we first make the following decomposition:

$$\begin{aligned}
\bar{\mathbb{B}}_{1kNT} - \hat{\mathbb{B}}_{1kNT} &= \frac{1}{\hat{\ }_k^{1/2}\ \ ^{3/2}} \sum_{i\in\hat{G}_k} \sum_{s=1}^{T} \sum_{t=1}^{T} \ _{M_T}(\ \ ) \ _{is} \hat{\ }_{it} - \frac{1}{\ _k^{1/2}\ \ ^{3/2}} \sum_{i\in G_k^0} \sum_{s=1}^{T} \sum_{t=1}^{T} \mathbb{E}(\ _{it} \ _{is}) \\
&= \frac{1}{\hat{\ }_k^{1/2}\ \ ^{3/2}} \sum_{i\in G_k^0} \sum_{s=1}^{T} \sum_{t=1}^{T} \ _{M_T}(\ \ ) \ _{is} \hat{\ }_{it} - \frac{1}{\ _k^{1/2}\ \ ^{3/2}} \sum_{i\in G_k^0} \sum_{s=1}^{T} \sum_{t=1}^{T} \mathbb{E}(\ _{it} \ _{is}) + \ _P(1) \\
&= \frac{1}{\hat{\ }_k^{1/2}\ \ ^{3/2}} \sum_{i\in G_k^0} \sum_{s=1}^{T} \sum_{t=1}^{T} \ _{M_T}(\ \ ) \ _{is}(\hat{\ }_{it} - \ _{it}) \\
&\quad + \frac{1}{\hat{\ }_k^{1/2}\ \ ^{3/2}} \sum_{i\in G_k^0} \sum_{s=1}^{T} \sum_{t=1}^{T} \ _{M_T}(\ \ )[\ _{is} \ _{it} - \mathbb{E}(\ _{is} \ _{it})] \\
&\quad + \frac{\ _k^{-1/2} - \hat{\ }_k^{-1/2}}{\ \ ^{3/2}} \sum_{i\in G_k^0} \sum_{s=1}^{T} \sum_{t=1}^{T} \ _{M_T}(\ \ ) \mathbb{E}(\ _{is} \ _{it}) \\
&\quad + \frac{1}{\ _k^{1/2}\ \ ^{3/2}} \sum_{i\in G_k^0} \sum_{s=1}^{T} \sum_{t=1}^{T} [1 - \ _{M_T}(\ \ )] \mathbb{E}(\ _{is} \ _{it}) + \ _P(1) \\
&\equiv \hat{\ }_{kNT,1} + \hat{\ }_{kNT,2} + \hat{\ }_{kNT,3} + \hat{\ }_{kNT,4} + \ _P(1)
\end{aligned}$$

where the $\ _P(1)$ term arises due to the replacement of $\hat{\ }_k$ by $\ _k^0$ and this can be easily justified by using the uniform classification consistency result and arguments as used in the proof of Theorem 2.5. We prove (ii) by demonstrating that $\hat{\ }_{kNT,s} = \ _P(1)$ for $\ = 1\ 2\ 3$ and 4

We first study $\hat{\ }_{kNT,1}$ Noting that $\hat{\ }_{it} = \ _{it} - \ _{it}'\hat{\ }_{\hat{G}_k} - \hat{\ }_i = \ _{it} - \ _{it}'\hat{\ }_{\hat{G}_k} - \frac{1}{T}\sum_{t=1}^{T}(\ _{it} - \ _{it}'\hat{\ }_{\hat{G}_k})$ and $\ _{it} = \ _{it}'\ _k^0 + \ _i + \ _{it}$ for $\ \in \ _k^0$ we have that for $\ \in \ _k^0$

$$\hat{\ }_{it} - \ _{it} = \ _{it} - \ _{it}'\hat{\ }_{\hat{G}_k} - \frac{1}{\ }\sum_{t=1}^{T}(\ _{it} - \ _{it}'\hat{\ }_{\hat{G}_k}) - \ _{it} = \tilde{\ }_{it}'(\ _k^0 - \hat{\ }_{\hat{G}_k}) - \bar{\ }_{i\cdot}$$

where $\bar{\ }_{i\cdot} = \frac{1}{T}\sum_{t=1}^{T} \ _{it}$ Then

$$\begin{aligned}
\hat{\ }_{kNT,1} &= \frac{1}{\hat{\ }_k^{1/2}\ \ ^{3/2}} \sum_{i\in G_k^0} \sum_{s=1}^{T} \sum_{t=1}^{T} \ _{M_T}(\ \ ) \ _{is} \tilde{\ }_{it}'(\ _k^0 - \hat{\ }_{\hat{G}_k}) - \frac{1}{\hat{\ }_k^{1/2}\ \ ^{3/2}} \sum_{i\in G_k^0} \sum_{s=1}^{T} \sum_{t=1}^{T} \ _{M_T}(\ \ ) \ _{is} \bar{\ }_{i\cdot} \\
&\equiv \ _{kNT,1}(1) - \ _{kNT,1}(2) \quad \text{say.}
\end{aligned}$$

In view of the fact that $\hat{\ }_{\hat{G}_k} - {}^0_k = {}_P\left(({}_k)^{-1/2} + {}^{-1} \right)$ and $\hat{\ }_k = {}_k (1 + {}_P (1))$ we have

$$\begin{aligned}
\| {}_{kNT,1} (1) \| &= \frac{1}{\hat{\ }_k^{1/2} \ {}^{3/2}} \left\| \sum_{i \in G_k^0} \sum_{s=1}^T \sum_{t=1}^T {}_{M_T} (\ ) \ {}_{is} \tilde{\ }'_{it} ({}^0_k - \hat{\ }_{\hat{G}_k}) \right\| \\
&\leq \frac{{}_k^{1/2}}{\hat{\ }_k^{1/2}} \left\| {}^0_k - \hat{\ }_{\hat{G}_k} \right\| \frac{1}{{}_k \ {}^2} \sum_{i \in G_k^0} \sum_{|s-t| \leq M_T} \| {}_{is} \tilde{\ }'_{it} \| \\
&= {}_k^{1/2} \ {}^{1/2} \ {}_P \left(({}_k)^{-1/2} + {}^{-1} \right) {}_P ({}_T) \\
&= {}_P \left(1 + {}_k^{1/2} \ {}^{-1/2} \right) {}_P ({}_T) = {}_P (1)
\end{aligned}$$

where we use the fact that $\frac{1}{N_k T^2} \sum_{i \in G_k^0} \sum_{|s-t| \leq M_T} \| {}_{is} \tilde{\ }'_{it} \| = {}_P ({}_T)$ by moment calculation and the Markov inequality. Let $\bar{\ }_{kNT,1} (2) \equiv \frac{1}{N_k^{1/2} T^{3/2}} \sum_{i \in G_k^0} \sum_{s=1}^T \sum_{t=1}^T {}_{M_T} (\ ) \ ' {}_{is} \bar{\ }_i$. where is any $\times 1$ nonrandom vector such that $\| \ \| = 1$ Then by Assumptions D1(i), (iii) and (iv),

$$\begin{aligned}
\left| \mathbb{E} \left[\bar{\ }_{kNT,1} (2) \right] \right| &\leq \frac{1}{{}_k^{1/2} \ {}^{5/2}} \sum_{i \in G_k^0} \sum_{s=1}^T \sum_{t=1}^T \sum_{r=1}^T {}_{M_T} (\ ) | \mathbb{E} (\ ' {}_{is} \ {}_{ir}) | \\
&\leq \frac{8}{{}_k^{1/2} \ {}^{5/2}} \sum_{i \in G_k^0} \sum_{s=1}^T \sum_{t=1}^T \sum_{r=1}^T {}_{M_T} (\ ) \| \ ' {}_{is} \|_{4q} \| {}_{ir} \|_{4q} \ {}_i (| \ - \ |)^{(2q-1)/(2q)} \\
&\leq \frac{{}_k^{1/2}}{{}^{3/2}} \left\{ \frac{1}{{}_k} \sum_{i \in G_k^0} {}_{\alpha,i}^{(2q-1)/(2q)} \right\} \left\{ \frac{1}{} \sum_{t,s,r: \ |s-t| \leq M_T} {}^{|r-s|(2q-1)/(2q)} \right\} \\
&= {}_k^{1/2} \ {}^{-3/2} \ (1) \ ({}_T) = \left({}_T \ {}_k^{1/2} \ {}^{-3/2} \right) = (1)
\end{aligned}$$

Similarly, by Assumptions D1(i)-(iv),

$$\begin{aligned}
\mathrm{Var} \left(\bar{\ }_{kNT,1} (2) \right) &= \frac{1}{{}_k \ {}^5} \sum_{i \in G_k^0} \mathrm{Var} \left(\sum_{s=1}^T \sum_{t=1}^T \sum_{r=1}^T {}_{M_T} (\ ) \ ' {}_{is} \ {}_{ir} \right) \\
&\leq \frac{1}{{}_k \ {}^5} \sum_{i \in G_k^0} \mathbb{E} \left[\left(\sum_{s=1}^T \sum_{t=1}^T \sum_{r=1}^T {}_{M_T} (\ ) \ ' {}_{is} \ {}_{ir} \right)^2 \right] \\
&= \frac{1}{{}_k \ {}^5} \sum_{i \in G_k^0} \sum_{1 \leq t_1, t_2, \ldots, t_6 \leq T} {}_{M_T} ({}_1 \ {}_2) \ {}_{M_T} ({}_4 \ {}_5) \mathbb{E} (\ ' {}_{it_2} \ {}_{it_3} \ ' {}_{it_5} \ {}_{it_6}) \\
&\leq \frac{1}{{}_k \ {}^5} \sum_{i \in G_k^0} \sum_{\substack{1 \leq t_1, t_2, \ldots, t_6 \leq T \\ |t_1 - t_2| \leq M_T, |t_4 - t_5| \leq M_T}} | \mathbb{E} (\ ' {}_{it_2} \ {}_{it_3} \ ' {}_{it_5} \ {}_{it_6}) | \\
&= \left({}_T^2 \ \right) = (1)
\end{aligned}$$

Consequently, $\bar{\ }_{kNT,1} (2) = {}_P (1)$ This, in conjunction with Corollary 2.3, implies that ${}_{kNT,1} (2) = {}_P (1)$ as is arbitrary. Thus we have shown that $\hat{\ }_{kNT,1} = {}_P (1)$

For $\hat{\ }_{kNT,2}$ note that $\hat{\ }_{kNT,2} = \bar{\ }_{kNT,2} \ {}_k^{1/2} \ \tilde{\ }_k^{-1/2} = \bar{\ }_{kNT,2} (1 + {}_P (1))$ where $\bar{\ }_{kNT,2} = \frac{1}{N_k^{1/2} T^{3/2}} \sum_{i \in G_k^0} \sum_{s=1}^T \sum_{t=1}^T {}_{M_T} (\ ) [\ {}_{is} \ {}_{it} - \mathbb{E} (\ {}_{is} \ {}_{it})]$ By construction $\mathbb{E} (\bar{\ }_{kNT,2}) = 0$ By Assumptions D1(ii)-(iii) and

Jensen inequality,

$$\begin{aligned}
\mathrm{Var}(\ '\ \bar{}_{kNT,2}) &= \frac{1}{k\ \ 3}\sum_{i\in G_k^0}\mathrm{Var}\left[\sum_{s=1}^{T}\sum_{t=1}^{T}{}_{M_T}(\ \ )\ '[\ {}_{is}\ {}_{it}-\mathbb{E}(\ {}_{is}\Delta\ {}_{it})]\right] \\
&\le \frac{1}{k\ \ 3}\sum_{i\in G_k^0}\sum_{s=1}^{T}\sum_{t=1}^{T}\sum_{r=1}^{T}\sum_{l=1}^{T}{}_{M_T}(\ \ )\ {}_{M_T}(\ \ )\,\mathbb{E}(\ '\ {}_{is}\ {}_{it}\ {}_{ir}\ {}_{il}\ ) \\
&\le \frac{1}{k\ \ 3}\sum_{i\in G_k^0}\sum_{|s-t|\le M_T}\sum_{|r-l|\le M_T}|\mathbb{E}(\ '\ {}_{is}\ {}_{it}\ {}_{ir}\ {}_{il}\ )| = (\ \tfrac{2}{T}\ ) = (1)
\end{aligned}$$

where the last equality follows from the fact that $\|\mathbb{E}(\ '\ {}_{is}\ {}_{it}\ {}_{ir}\ {}_{il}\ )\| \le \max_{i,s,t}\|\ {}_{is}\ {}_{it}\|_2^2 \le \max_{i,t}\|\ {}_{it}\|_4^2 \times \max_{i,t}\|\ {}_{it}\|_4^2\ \ \infty$ by Assumption D1(iii). Then $\bar{}_{kNT,2} = {}_P(1)$ by the Chebyshev inequality and thus $\hat{}_{kNT,2} = {}_P(1)$

By Corollary 2.3 and the Davydov inequality,

$$\begin{aligned}
\|\ \hat{}_{kNT,3}\| &= \frac{|\ {}_k^{-1}-\hat{}_k^{-1}|}{{}^{3/2}(\ {}_k^{-1/2}+\tilde{}_k^{-1/2})}\left\|\sum_{i\in G_k^0}\sum_{s=1}^{T}\sum_{t=1}^{T}{}_{M_T}(\ \ )\,\mathbb{E}(\ {}_{is}\ {}_{it})\right\| \\
&\le \frac{|\ \hat{}_k-\ {}_k|}{{}^{1/2}\ \hat{}_k(\ {}_k^{-1/2}+\hat{}_k^{-1/2})}\left\{\frac{1}{k}\sum_{i\in G_k^0}\sum_{|s-t|\le M_T}\|\mathbb{E}(\ {}_{is}\ {}_{it})\|\right\} \\
&= {}_P(\ {}_k^{-1/2}\ {}^{-1/2})\ (1) = {}_P(1)
\end{aligned}$$

By Assumptions D1(i)-(iv) and the Davydov inequality,

$$\begin{aligned}
\|\ \hat{}_{kNT,4}\| &= \frac{1}{{}_k^{1/2}\ {}^{3/2}}\sum_{i\in G_k^0}\sum_{s=1}^{T}\sum_{t=1}^{T}[1-\ {}_{M_T}(\ \ )]\,\mathbb{E}(\ {}_{is}\ {}_{it}) \\
&= \left\|\frac{1}{{}_k^{1/2}\ {}^{3/2}}\sum_{i\in G_k^0}\sum_{s=1}^{T}\sum_{t=1}^{T}[1-\ {}_{M_T}(\ \ )]\,\mathbb{E}(\ {}_{is}\ {}_{it})\right\| \\
&\le \frac{8}{{}_k^{1/2}\ {}^{3/2}}\sum_{i\in G_k^0}\sum_{|s-t|>M_T}{}_i(|\ -\ |)^{(2q-1)/(2q)}\|\ {}_{is}\|_{4q}\|\ {}_{it}\|_{4q} \\
&\le \ {}_k^{-1/2}\ {}^{1/2}\sum_{i\in G_k^0}{}_i(\ {}_T)^{(2q-1)/(2q)} = (1)
\end{aligned}$$

This completes the proof of the proposition. ■

With the above result in hand, we can readily show that

$$\begin{aligned}
\sqrt{\ {}_k}(\ \hat{}_k^{(c)}-\ {}_k^0) &= \sqrt{\ {}_k}(\ \hat{}_k-\ {}_k^0)-\mathbb{H}_{kNT}^{-1}\mathbb{B}_{1kNT} + (\ {}_k\ \hat{}_k)^{1/2}\left[\mathbb{H}_{kNT}^{-1}\mathbb{B}_{1kNT}-\mathbb{H}_{kNT}^{-1}\hat{\mathbb{B}}_{1kNT}\right] \\
&\quad + \left[1-(\ {}_k\ \hat{}_k)^{1/2}\right]\mathbb{H}_{kNT}^{-1}\mathbb{B}_{1kNT} \\
&= \sqrt{\ {}_k}(\ \hat{}_k-\ {}_k^0)-\mathbb{H}_{kNT}^{-1}\mathbb{B}_{1kNT} + {}_P(1) + {}_P(\ {}_k^{-1})\ \left((\ K\ )^{1/2}\right) \\
&= \sqrt{\ {}_k}(\ \hat{}_k-\ {}_k^0)-\mathbb{H}_{kNT}^{-1}\mathbb{B}_{1kNT} + {}_P(1)
\end{aligned}$$

That is, $\sqrt{\ {}_k}(\ \hat{}_k^{(c)}-\ {}_k^0)$ has the desired limiting distribution centered on the origin.

Bias correction for the PGMM C-Lasso estimator in dynamic panel data models can be done analogously. For simplicity we focus on the case where $_{iNT} = \ _d$ for all Recall from Theorem 3.4 and the remark regarding Assumption B3(iii) (see (3.3) in particular) that

$$\square \overline{\quad}_k (\tilde{\ }_k - {}^0_k) - \bar{\ }^{-1}_k \ _{kNT} \xrightarrow{D} (0 \ {}^{-1}_k \ _k \ {}^{-1}_k) \text{ for } = 1 \qquad _0$$

where $\bar{\ }_k \equiv \frac{1}{N_k} \square_{i \in G^0_k} \bar{\ }'_{i,z\Delta x} \bar{\ }_{i,z\Delta x}$ and $_{kNT} = \frac{1}{N_k^{1/2} T^{3/2}} \square_{i \in G^0_k} \square^T_{s=1} \square^T_{t=1} \mathbb{E}(\Delta \ _{is} \ '_{is} \ _{it} \Delta \ _{it})$ Based on (3.3), in order to verify Assumption B3(iii) we also need to show

$$_{kNT} = \frac{1}{_k^{1/2} \ ^{1/2}} \square_{i \in G^0_k} \square^T_{t=1} \bar{\ }'_{i,z\Delta x} \ _{it} \Delta \ _{it} \xrightarrow{D} (0 \ _k) \text{ and} \tag{S1}$$

$$_{kNT} = \frac{1}{_k^{1/2} \ ^{3/2}} \square_{i \in G^0_k} \square^T_{s=1} \square^T_{t=1} \{[\Delta \ _{is} \ '_{is} - \mathbb{E}(\Delta \ _{is} \ '_{is})] \ _{it} \Delta \ _{it} - \mathbb{E}(\Delta \ _{is} \ '_{is} \ _{it} \Delta \ _{it})\} = \ _P(1) \tag{S2}$$

The first part is assured by a version of the CLT. Below we first propose an estimate of the bias $\bar{\ }^{-1}_k \ _{kNT}$ and then demonstrate (S2).

To correct the bias, we propose to obtain consistent estimates of $\bar{\ }_k$ and $_{kNT}$ respectively by

$$\tilde{\ }_k = \frac{1}{\tilde{\ }_k} \square_{i \in \tilde{G}_k} \tilde{\ }'_{i,z\Delta x} \tilde{\ }_{i,z\Delta x} \text{ and } \tilde{\ }_{kNT} = \frac{1}{\tilde{\ }_k^{1/2} \ ^{3/2}} \square_{i \in \tilde{G}_k} \square^T_{s=1} \square^T_{t=1} \ _{M_T}(\quad) \Delta \ _{is} \ '_{is} \ _{it} \Delta \tilde{\ }_{it}$$

where $\Delta \tilde{\ }_{it} = \Delta \ _{it} - \tilde{\ }'_{\tilde{G}_k} \Delta \ _{it}$ for all $\in \tilde{\ }_k$[3] $_{M_T}(\quad)$ is as defined above: $_{M_T}(\quad) = {}^0_{M_T}(| - |)$ and ${}^0_{M_T}(\ )$ denotes the Bartlett kernel: ${}^0_{M_T}(\ ) = (1 - | \ | \quad _T) \mathbf{1}\{| \ | \leq \quad _T\}$ Note that we also allow dynamic misspecification here. If one is sure that the model is dynamically correctly specified in the sense that $\mathbb{E}(\Delta \ _{it} | \mathcal{F}_{i,t-1}) = 0$ where $\mathcal{F}_{i,t-1} = (\Delta \ _{i,t-1} \ \Delta \ _{it-1} \ _{it}; \Delta \ _{i,t-2} \ \Delta \ _{it-2} \ _{i,t-1}; \ )$ one can use the one-sided kernel: $_{M_T}(\quad) = {}^1_{M_T}(\ - \ )$ where ${}^1_{M_T}(\ ) = (1 - \quad _T) \mathbf{1}\{0 \leq \quad \leq \quad _T\}$ The bias-corrected C-Lasso estimator of ${}^0_k$ would be

$$\tilde{\ }^{(c)}_k = \tilde{\ }_k - \frac{1}{\square \overline{\tilde{\ }_k}} \tilde{\ }^{-1}_k \tilde{\ }_{kNT}$$

Note that Theorem 3.4 indicates that there is no need to consider bias correction for the post Lasso estimator $\tilde{\ }_{\tilde{G}_k}$

Let $_i \equiv (\ _{i1} \qquad _{iT})'$ and $_i \equiv (\ _{i1} \qquad _{iT})'$ We add the following assumption.

ASSUMPTION D2. *(i) For each* $= 1$ $\{(\Delta \ _{it} \ _{it} \ \Delta \ _{it}) : \ = 1 \ 2 \quad \}$ *is strong mixing with mixing coefficients* $\{\ _i(\cdot)\}$. *In addition,* $_i(\ ) \leq \ _{\alpha,i} \ ^\tau$ *for some* $_{\alpha,i} \quad \infty$ *and* $\in (0 \ 1)$ *where* $\frac{1}{N_k} \square_{i \in G^0_k} \ ^{(2q-1)/(2q)}_{\alpha,i} = \ (1)$ *and* $\frac{1}{N_k} \square_{i \in G^0_k} \ ^{(q-1)/q}_{\alpha,i} = \ (1)$

(ii) $(\ _i \ _i)$ *are independent across* $\in \ ^0_k$ *where* $= 1 \qquad _0$

(iii) $\max_{i,t} \mathbb{E} \|\Delta \ _{it} \ '_{it}\|^{4q} \quad \infty$ *and* $\max_{i,t} \mathbb{E} \| \ _{it} \Delta \ _{it}\|^{4q} \quad \infty$ *for some* $\quad 1$

(iv) As $(\qquad) \to \infty \quad _T \to \infty \quad \frac{2}{T} \quad \to 0$ *and* $_k^{-1/2} \ ^{1/2} \square_{i \in G^0_k} \ _i(\ _T)^{(2q-1)/(2q)} \to 0$ *for each* $= 1 \qquad _0$

Assumptions D2(i)-(iv) parallel D1(i)-(iv). The major difference is that we do not need $\frac{2}{T} \ _k \ ^3 \to 0$ in D2(iv) but require $\quad 1$ in D2(iii).

[3]Observe that $\alpha_k - \alpha^0_k \quad O_P(\ N_k T \ ^{-1/2} \quad T^{-1})$ and $\alpha_{\tilde{G}_k} - \alpha^0_k \quad O_P(\ N_k T \ ^{-1/2})$. We recommend using the post-Lasso estimator $\alpha_{\tilde{G}_k}$.

Suppose that the conditions of Theorem 3.4 hold. Suppose Assumption D2 holds. Then

$\tilde{}^{-1}_{k}\,\tilde{}_{kNT} - \bar{}^{-1}_{k}\,{}_{kNT} = {}_P(1)$

[redacted]

$1 + {}_P(1)$ by Corollary 3.3, we have

$$
\begin{aligned}
\| {}_{kNT,1} \| &= \frac{1}{\tilde{}_k^{1/2}\ {}^{3/2}} \left\| \sum_{i \in G_k^0} \sum_{s=1}^{T} \sum_{t=1}^{T} {}_{M_T}(\ ) \Delta\ {}_{is}\ {}'_{is}\ {}_{it} (\Delta\ {}_{it})' (\ {}^0_k - \tilde{}_{\tilde{G}_k}) \right\| \\
&\leq (\tilde{}_k)^{1/2} \left\| {}^0_k - \tilde{}_{\tilde{G}_k} \right\| \frac{k}{\tilde{}_k} \frac{1}{k\ {}^2} \sum_{i \in G_k^0} \sum_{|s-t| \leq M_T} \| \Delta\ {}_{is}\ {}'_{is}\ {}_{it} (\Delta\ {}_{it})' \| \\
&= {}_P(1)\ {}_{kNT,1}
\end{aligned}
$$

where ${}_{kNT,1} = \frac{1}{N_k T^2} \sum_{i \in G_k^0} \sum_{|s-t| \leq M_T}^{T} \| \Delta\ {}_{is}\ {}'_{is}\ {}_{it} (\Delta\ {}_{it})' \|$ By the Markov inequality, ${}_{kNT,1} = {}_P(\ {}_T\ )$ It follows that $\| {}_{kNT,1} \| = {}_P(\ {}_T\ ) = {}_P(1)$ under Assumption D2(iv).

For ${}_{kNT,2}$ note that ${}_{kNT,2} = {}_{kNT,2}\ {}^{1/2}_k\ \tilde{}^{1/2}_k = {}_{kNT,2}(1 + {}_P(1))$ where

$$
{}_{kNT,2} = \frac{1}{{}^{1/2}_k\ {}^{3/2}} \sum_{i \in G_k^0} \sum_{s=1}^{T} \sum_{t=1}^{T} {}_{M_T}(\ ) [\Delta\ {}_{is}\ {}'_{is}\ {}_{it} \Delta\ {}_{it} - \mathbb{E}(\Delta\ {}_{is}\ {}'_{is}\ {}_{it} \Delta\ {}_{it})]
$$

Let be any $\times 1$ nonrandom vector such that $\| \ \| = 1$ Then $\mathbb{E}(\ {}'\ {}_{kNT,2}) = 0$ By Assumptions D2(ii)-(iv) and Jensen inequality,

$$
\begin{aligned}
&\mathrm{Var}(\ {}'\ {}_{kNT,2}) \\
&= \frac{1}{k\ {}^3} \sum_{i \in G_k^0} \mathrm{Var}\left[\sum_{s=1}^{T} \sum_{t=1}^{T} {}_{M_T}(\ )\ {}' \{\Delta\ {}_{is}\ {}'_{is}\ {}_{it} \Delta\ {}_{it} - \mathbb{E}(\Delta\ {}_{is}\ {}'_{is}\ {}_{it} \Delta\ {}_{it})\} \right] \\
&\leq \frac{1}{k\ {}^3} \sum_{i \in G_k^0} \sum_{s=1}^{T} \sum_{t=1}^{T} \sum_{r=1}^{T} \sum_{l=1}^{T} {}_{M_T}(\ )\ {}_{M_T}(\ )\ {}' \mathbb{E}[\Delta\ {}_{is}\ {}'_{is}\ {}_{it} \Delta\ {}_{it} \Delta\ {}_{il}\ {}'_{il}\ {}_{ir} \Delta\ {}_{ir}] \\
&\leq \frac{1}{k\ {}^3} \sum_{i \in G_k^0} \sum_{|s-t| \leq M_T} \sum_{|r-l| \leq M_T} \| \mathbb{E}[\ {}' \Delta\ {}_{is}\ {}'_{is}\ {}_{it} \Delta\ {}_{it} \Delta\ {}_{il}\ {}'_{il}\ {}_{ir} \Delta\ {}_{ir}\ ] \| \\
&= (\ {}^2_T\ ) = (1)
\end{aligned}
$$

where the last equality follows from the fact that $\| \mathbb{E}[\ {}' \Delta\ {}_{is}\ {}'_{is}\ {}_{it} \Delta\ {}_{it} \Delta\ {}_{il}\ {}'_{il}\ {}_{ir} \Delta\ {}_{ir}\ ] \| \leq \max_{i,s} \left(\mathbb{E} \| \Delta\ {}_{is}\ {}'_{is} \|^4 \right)^{1/2} \times \max_{i,t} \left(\mathbb{E} \| {}_{it} \Delta\ {}_{it} \|^4 \right)^{1/2} \infty$ by Assumption D2(iii). It follows that ${}_{kNT,2} = {}_P(1)$

By Corollary 3.3 and the Davydov inequality,

$$
\begin{aligned}
\| {}_{kNT,3} \| &= \frac{\left| {}^{-1}_k - \tilde{}^{-1}_k \right|}{{}^{3/2}(\ {}^{-1/2}_k + \tilde{}^{-1/2}_k)} \left\| \sum_{i \in G_k^0} \sum_{s=1}^{T} \sum_{t=1}^{T} {}_{M_T}(\ ) \mathbb{E}(\Delta\ {}_{is}\ {}'_{is}\ {}_{it} \Delta\ {}_{it}) \right\| \\
&\leq \frac{\left| \tilde{}_k - {}_k \right|}{{}^{1/2}\ \tilde{}_k(\ {}^{-1/2}_k + \tilde{}^{-1/2}_k)} \left\{ \frac{1}{k} \sum_{i \in G_k^0} \sum_{|s-t| \leq M_T} \| \mathbb{E}(\Delta\ {}_{is}\ {}'_{is}\ {}_{it} \Delta\ {}_{it}) \| \right\} \\
&= {}_P(\ {}^{-1/2}_k\ {}^{-1/2})\ (1) = {}_P(1)
\end{aligned}
$$

By Assumptions D2(i)-(iii) and the Davydov inequality,

$$
\begin{aligned}
\| \;\; _{kNT,4}\| &= \left\| \frac{1}{\;_k^{1/2}\;\;^{3/2}} \sum_{i\in G_k^0} \sum_{s=1}^{T} \sum_{t=1}^{T} [1- \;\;_{M_T}(\;\;)]\, \mathbb{E}(\Delta \;\;_{is}\;\;'_{is}\;\;_{it}\Delta \;\;_{it}) \right\| \\
&\leq \frac{8}{\;_k^{1/2}\;\;^{3/2}} \sum_{i\in G_k^0} \sum_{|s-t|>M_T} \;\;_i(|\;\;-\;\;|)^{(2q-1)/(2q)} \|\Delta \;\;_{is}\;\;'_{is}\|_{4q} \| \;\;_{it}\Delta \;\;_{it}\|_{4q} \\
&\leq \;\;_k^{-1/2}\;\;^{1/2} \sum_{i\in G_k^0} \;\;_i(\;\;_T)^{(2q-1)/(2q)} = \;\;(1)
\end{aligned}
$$

This completes the proof of the proposition. ■

With the above result in hand, we can readily show that

$$
\begin{aligned}
\sqrt{\;\;_k}(\tilde{\;}_k^{(c)} - \;\;_k^0) &= \sqrt{\;\;_k}(\tilde{\;}_k - \;\;_k^0) - \;\;_k^{-1} \;\;_{kNT} + (\;\;_k\;\;\tilde{\;}_k)^{1/2} \;\;_k^{-1} \;\;_{kNT} - \tilde{\;}_k^{-1} \tilde{\;}_{kNT} \\
&\quad + \left[1 - (\;\;_k\;\;\tilde{\;}_k)^{1/2}\right] \;\;_k^{-1} \;\;_{kNT} \\
&= \sqrt{\;\;_k}(\tilde{\;}_k - \;\;_k^0) - \;\;_k^{-1} \;\;_{kNT} + \;\;_P(1) + \;\;_P(\;\;_k^{-1}) \;\; ((\;\;_K\;\;)^{1/2}) \\
&= \sqrt{\;\;_k}(\tilde{\;}_k - \;\;_k^0) - \;\;_k^{-1} \;\;_{kNT} + \;\;_P(1)
\end{aligned}
$$

That is, $\sqrt{\;\;_k}(\tilde{\;}_k^{(c)} - \;\;_k^0)$ has the desired limiting distribution centered on the origin.

Now, we demonstrate (S2). Let $\;\;_{is} = \Delta \;\;_{is}\;\;'_{is} - \mathbb{E}(\Delta \;\;_{is}\;\;'_{is})$ and $\;\;_{it} = \;\;_{it}\Delta \;\;_{it}$ Noting that $\mathbb{E}(\;\;_{is}) = 0$ and $\mathbb{E}(\;\;_{it}) = 0$ we have

$$
\begin{aligned}
\;\;_{kNT} &= \frac{1}{\;_k^{1/2}\;\;^{3/2}} \sum_{i\in G_k^0} \sum_{s=1}^{T} \sum_{t=1}^{T} [\;\;_{is}\;\;_{it} - \mathbb{E}(\;\;_{is}\;\;_{it})] \\
&= \frac{1}{\;_k^{1/2}\;\;^{3/2}} \sum_{i\in G_k^0} \sum_{t=1}^{T} [\;\;_{it}\;\;_{it} - \mathbb{E}(\;\;_{it}\;\;_{it})] + \frac{1}{\;_k^{1/2}\;\;^{3/2}} \sum_{i\in G_k^0} \sum_{1\leq s<t\leq T} [\;\;_{is}\;\;_{it} - \mathbb{E}(\;\;_{is}\;\;_{it})] \\
&\quad + \frac{1}{\;_k^{1/2}\;\;^{3/2}} \sum_{i\in G_k^0} \sum_{1\leq t<s\leq T} [\;\;_{is}\;\;_{it} - \mathbb{E}(\;\;_{is}\;\;_{it})] \\
&\equiv \;\;_{kNT,1} + \;\;_{kNT,2} + \;\;_{kNT,3} \text{ say.}
\end{aligned}
$$

It is trivial to show that $\;\;_{kNT,1} = \;\;_P(\;\;^{-1})$ by the Chebyshev and Davydov inequalities. For $\;\;_{kNT,2}$ we have $\mathbb{E}(\;\;_{kNT,2}) = 0$ by construction, and by Assumption D2(ii) and Jensen inequality

$$
\mathbb{E}(\;\;_{kNT,2}^2) = \frac{1}{\;_k\;\;^3} \sum_{i\in G_k^0} \mathrm{Var}\left(\sum_{1\leq t_1<t_2\leq T} [\;\;_{it_1}\;\;_{it_2} \right. \qquad 00
$$

kNT,a

and obtain the updated estimate $(\hat{\boldsymbol{\beta}}^{(r,1)}\ \hat{}^{(r)}_{1})$ of $(\boldsymbol{\beta}\ \ _1)$ Next we choose $(\boldsymbol{\beta}\ \ _2)$ to minimize

$$ {}^{(r,2,K)}_{1NT,\lambda_1}(\boldsymbol{\beta}\ \ _2) = \ _{1,NT}(\boldsymbol{\beta}) + \frac{1}{\ } \sum_{i=1}^{N} \| \ _i - \ _2 \| \left\| \hat{}^{(r,1)}_{i} - \hat{}^{(r)}_{1} \right\| \Pi_{k\neq 1,2}^{K} \left\| \hat{}^{(r-1)}_{i} - \hat{}^{(r-1)}_{k} \right\| $$

and obtain the updated estimate $(\hat{\boldsymbol{\beta}}^{(r,2)}\ \hat{}^{(r)}_{2})$ of $(\boldsymbol{\beta}\ \ _2)$ Repeat this procedure until we choose $(\boldsymbol{\beta}\ \ _K)$ to minimize

$$ {}^{(r,K,K)}_{1NT,\lambda_1}(\boldsymbol{\beta}\ \ _K) = \ _{1,NT}(\boldsymbol{\beta}) + \frac{1}{\ } \sum_{i=1}^{N} \| \ _i - \ _K \| \Pi_{k=1}^{K-1} \left\| \hat{}^{(r,k)}_{i} - \hat{}^{(r)}_{k} \right\| $$

and obtain the updated estimate $(\hat{\boldsymbol{\beta}}^{(r,K)}\ \hat{}^{(r)}_{K})$ of $(\boldsymbol{\beta}\ \ _K)$ Let $\hat{\boldsymbol{\alpha}}^{(r)} = (\hat{}^{(r)}_{1}\ \ \hat{}^{(r)}_{K})$ and $\hat{\mathcal{Q}}^{(r,K)}_{1NT} = \sum_{k=1}^{K} {}^{(r,k,K)}_{1NT,\lambda_1}(\hat{\boldsymbol{\beta}}^{(r,k)}\ \hat{\boldsymbol{\alpha}}^{(r)}_{k})$. Update the iteration index from to $+1$.

Repeat Step 2 until a convergence criterion is achieved, e.g., when

$$ \hat{\mathcal{Q}}^{(r-1,K)}_{1NT} - \hat{\mathcal{Q}}^{(r,K)}_{1NT} \quad _{\text{tol}} \quad \text{and} \quad \frac{\sum_{k=1}^{K} \left\| \hat{}^{(r)}_{k} - \hat{}^{(r-1)}_{k} \right\|^2}{\sum_{k=1}^{K} \left\| \hat{}^{(r-1)}_{k} \right\|^2 + 0\ 0001} \quad _{\text{tol}} $$

where $_{\text{tol}}$ is some prescribed tolerance level (e.g., 0.0001). Define the final iterative estimate of $\boldsymbol{\alpha}$ as $\hat{\boldsymbol{\alpha}} = (\hat{}^{(R)}_{1}\ \ \hat{}^{(R)}_{K})$ for a sufficiently large such that the convergence criterion is met. Intuitively, individual is classified to group $\hat{}_k$ if $\hat{}^{(R,k)}_{i} = \hat{}_k$; otherwise, $\hat{}_i$ is assigned to be the ${}^{(R)}_{k}$ that is closest to some $\hat{}^{(R,l)}_{i}$, $=1$. In either case, we can write the individual estimate as $\hat{}_i = \hat{}^{(R)}_{k^*}$, where $^* = \text{argmin}_{k\in\{1,\ldots,K\}} \| \hat{}^{(R,l^*(k))}_{i} - \hat{}^{(R)}_{k} \|$ and $^*(\ ) = \text{argmin}_{l\in\{1,\ldots,K\}} \| \hat{}^{(R,l)}_{i} - \hat{}^{(R)}_{k} \|$.

The optimization of ${}^{(r,k,K)}_{1NT,\lambda_1}(\boldsymbol{\beta}\ \ _k)$ is conducted on the $(\ \ +\ )$-dimensional parameter space for $(\boldsymbol{\beta}\ \ _k)$. When is non-trivial, this is a high-dimensional optimization problem. Obviously, in the penalty term $_1 \cdots \ _N$ and $_k$ are jointly convex, given $\Pi_{l=k+1}^{K} \left\| \hat{}^{(r-1)}_{i} - \hat{}^{(r-1)}_{l} \right\|$ and $\Pi_{l=1}^{k-1} \left\| \hat{}^{(r)}_{i} - \hat{}^{(r)}_{l} \right\|$ for each $=$ 1 . If $_{1,NT}(\boldsymbol{\beta})$ is convex in $\boldsymbol{\beta}$, then ${}^{(r,k,K)}_{1NT,\lambda_1}(\boldsymbol{\beta}\ \ _k)$ as the summation of $_{1,NT}(\boldsymbol{\beta})$ and the penalty, is also convex in $(\boldsymbol{\beta}\ \ _k)$. Convexity can substantially reduce the computational burden of high-dimensional optimization.

A convex $_{1,NT}(\boldsymbol{\beta})$ is common in panel data models. Convexity apparently holds in the linear models in Examples 1 and 2. It also holds in the nonlinear models in Example 3 with $(\cdot)$ as the standard logistic or normal CDF, and in Example 4 after re-parameterizing the original parameter $(\ _i\ \ _i\ \ ^2_\varepsilon)$ into $(\ _{1i} = \ _i\ \ ^2_\varepsilon\ \ _{2i} = \ _i\ \ ^2_\varepsilon\ \ _3 = 1\ \ ^2_\varepsilon)$. We utilize the convexity throughout our numerical works.

Given the convexity in each substep ( ), the proposed algorithm consists of a sequence of convex problems implemented in an iterative manner. In particular, the only difference between the standard Lasso and a single substep of PPL is that Lasso shrinks the coefficients to a known center (zero), while the center of PPL is determined in the convex programming. Thus a PPL iteration has the same computational complexity as Lasso, which is $(\ \ ^3\ )$ in our context of panel linear regression (Efron, Hastie and Johnstone, 2004, p.443). The computational cost of a single iteration is minimal.

Since the additive-multiplicative penalty is not jointly convex in all the parameters $(\boldsymbol{\beta}\ \boldsymbol{\alpha})$, we can take advantage of convexity in each substep for $(\boldsymbol{\beta}\ \ _k)$ but not simultaneously for $(\boldsymbol{\beta}\ \boldsymbol{\alpha})$. As a consequence of

In this section we carry out two more simulation exercises, one using PGMM to estimate a static panel model with endogenous regressors as in DGP 4 below, and the other using PLS to estimate the linear panel AR(1) in DGP2.

(Linear static panel with endogeneity.) We maintain the linear panel structure model with two explanatory variables as in DGP 1, but the first regressors is endogenous as it is generated from the following underlying reduced-form equation: $_{it1} = 0\ 2\ {}^{0}_{i} + 0\ 5\ _{it1} + 0\ 5\ _{it2} + 0\ 5\ _{it}$ where $_{it1}$ and $_{it2}$ are two excluded instrumental variables, and the reduced-form error $_{it}$ and the structural-equation idiosyncratic shock $_{it}$ follow a bivariate normal distribution:

$$\begin{pmatrix} _{it} \\ _{it} \end{pmatrix} \sim \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix} \begin{pmatrix} 1 & 0\ 3 \\ 0\ 3 & 1 \end{pmatrix} \right)$$

The second regressor $_{it2}$ is exogenous, and $(\ _{it2}\ \ _{it1}\ \ _{it2}) \sim$ IID $(0\ \ _{3})$ is independent of $(\ _{it}\ \ _{it})$. All variables are independent across and . The econometrician observes $(\ _{it}\ \ _{it1}\ \ _{it2}\ \ _{it1}\ \ _{it2})$. The true coefficients of the three groups are (0 2 1 8), (1 1) and (1 8 0 2), respectively.

We report the statistics in Tables S1 and S2, which correspond to Tables 1 and 2, respectively, in the main text. The choice of tuning parameters are exactly the same as described in Section 4. When we compare PLS estimation with PGMM in DGP 2, we find that the PLS works better in determining the correct number of groups and in classifying the individual units. The 95% coverage probabilities are comparable to those of PGMM when $= 50$, but are lower than PGMM when is small. Similar to PPL in DGP 3, the lower coverage probabilities is caused by the bias. The analytical bias correction removes the bias asymptotically, but the effect is limited when is small, as is shown in the oracle. The post-Lasso has larger coverage probability than the oracle, as the estimated standard deviation is inflated by a few misclassified units.

Table S1: Frequency of selecting $= 1$ 5 groups when $_{0} = 3$

N	T	DGP 4					DGP 2 (PLS)				
		1	2		4	5	1	2		4	5
100	15	0	0.022		0.076	0	0	0.106		0	0
100	25	0	0		0.028	0.006	0	0		0	0
100	50	0	0		0.004	0	0	0		0	0
200	15	0	0		0.058	0.002	0	0		0	0
200	25	0	0		0.046	0.004	0	0		0	0
200	50	0	0		0.006	0	0	0		0	0

Table S3 reports the RMSE and bias of $\boldsymbol{\alpha}_1$ from post-Lasso and C-Lasso under the true $_{0}$ and the IC-determined $\hat{}$ (or $\tilde{}$ for PGMM). These estimates are bias corrected whenever necessary in the DGPs. For example, the RMSE of PPL under $_{0}$ is calculated as $\left(\frac{1}{S} \sum_{s=1}^{S} \sum_{k=1}^{K_0} \frac{\hat{N}_k^{(s)}}{N} \left(\hat{}^{(s)}_{k,1} (\ _0\ \ _1) - {}^{0}_{k,1} \right)^2 \right)^{1/2}$

where and

Table S2: Classification and point estimation of $\boldsymbol{\alpha}_1$ in additional simulations

	N	T	% of correct classification	Post-Lasso RMSE	Post-Lasso Bias	Post-Lasso Coverage	Oracle RMSE	Oracle Bias	Oracle Coverage
DGP 4	100	15	0.8287	0.1583	0.0462	0.7850	0.0806	0.0018	0.9344
	100	25	0.9281	0.0883	0.0195	0.8880	0.0617	0.0009	0.9380
	100	50	0.9885	0.0517	0.0075	0.9406	0.0437	-0.0012	0.9422
	200	15	0.8378	0.1155	0.0484	0.7860	0.0577	-0.0016	0.9454
	200	25	0.9320	0.0643	0.0199	0.8742	0.0436	0.0001	0.9506
	200	50	0.9881	0.0364	0.0074	0.9356	0.0311	-0.0005	0.9450
DGP 2	100	15	0.8907	0.0413	0.0061	0.9148	0.0352	0.0041	0.8524
(PLS)	100	25	0.9511	0.0261	0.0041	0.9710	0.0241	0.0028	0.9076
	100	50	0.9908	0.0160	0.0015	0.9908	0.0156	0.0013	0.9334
	200	15	0.8949	0.0294	0.0064	0.9154	0.0253	0.0052	0.8576
	200	25	0.9520	0.0188	0.0037	0.9714	0.0178	0.0036	0.8808
	200	50	0.9912	0.0113	0.0017	0.9934	0.0111	0.0015	0.9282

of post-Lasso, although C-Lasso appears to have larger RMSE in the PGMM estimation of DGP 2, where it does not enjoy the oracle property.

When $\neq {}_0$, we generalize the definition of the set of true group-specific parameters. For ${}_0$, we shrink $\boldsymbol{\alpha}_1^0 = ({}^0_{1,1} \quad {}^0_{K_0,1})$ into a -element subset $\boldsymbol{\alpha}_1^0(\ )$. For ${}_0$, we augment $\boldsymbol{\alpha}_1^0$ by adding $- {}_0$ elements choosing from ${}^0_{k,1} \quad {}^0_{K_0,1}$ so that the resulting $\boldsymbol{\alpha}_1^0(\ )$ contains $\boldsymbol{\alpha}_1^0$. Elements are eliminated or concatenated in each replication to fit $\hat{\boldsymbol{\alpha}}(\hat{\ }^{(s)} \quad {}_1)$. In this scenario, the RMSE is calculated as $\left(\frac{1}{S} \sum_{s=1}^{S} \sum_{k=1}^{\hat{K}^{(s)}} \frac{\hat{N}_k^{(s)}}{N} \left(\hat{\ }_{k,1}(\hat{\ }^{(s)} \quad {}_1) - {}^0_{k,1}(\hat{\ }^{(s)}) \right)^2 \right)^{1/2}$ According to the simulation, the effect of not knowing ${}_0$ is noticeable when $= 15$ in the linear models and $= 25$ in the nonlinear model, but it does not necessarily enlarges the RMSE, for the estimator under ${}_0$ is also noisy when is small. The discrepancy of the RMSE and bias between ${}_0$ and $\hat{\ }$ (or $\tilde{\ }$) quickly vanishes when grows.

fi

All data are downloaded from the World Bank.[4] We extract all countries with all the variables in (5.1) available. Using the time span 1995–2010, we were able to construct a balanced panel of 57 countries. We remove one outlier Bulgaria, whose 1997 economic collapse produced hyperinflation in the CPI that significantly distorted the overall mean and the standard deviation. In total we collect 56 countries. The summary statistics are shown in Table S4.

In the implementation, we scale-normalize all the variables for each individual unit to guarantee that the coefficients are comparable. Moreover, in PGMM we use $\Delta \ {}_{t-2}$ and a constant as two excluded IVs. Although the constant is uncorrelated with the endogenous variable, adding it here stabilizes the post-Lasso estimation in finite samples.

Table S5 displays the group membership. The country names in bold are the 47 coincidences of PLS and PGMM classification out of the total 56 countries.

[4] http://data.worldbank.org/data-catalog/world-development-indicators.

Table S3: Estimation of $\boldsymbol{\alpha}_1$ by post-lasso and C-Lasso under $\ _0$ and $\hat{\ }$ or $\tilde{\ }$

			Post-Lasso				C-Lasso				Oracle	
			K	K_0	K	K	K	K_0	K	K		
	N	T	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias
DGP 1	100	15	0.0596	0.0108	0.0829	0.0092	0.0619	0.0133	0.0839	0.0120	0.0463	0.0012
	100	25	0.0385	0.0019	0.0385	0.0019	0.0396	0.0040	0.0396	0.0040	0.0353	0.0001
	100	50	0.0249	0.0000	0.0249	0.0000	0.0255	0.0011	0.0255	0.0011	0.0245	-0.0002
	200	15	0.0434	0.0079	0.1373	0.0081	0.0457	0.0107	0.1353	0.0114	0.0324	-0.0013
	200	25	0.0273	0.0015	0.0273	0.0015	0.0280	0.0040	0.0280	0.0040	0.0250	-0.0006
	200	50	0.0174	-0.0001	0.0174	-0.0001	0.0181	0.0011	0.0181	0.0011	0.0171	-0.0002
DGP 2	100	15	0.0848	-0.0090	0.0787	-0.0016	0.1311	-0.0372	0.1188	-0.0250	0.0502	-0.0037
(PGMM)	100	25	0.0556	-0.0055	0.0561	-0.0051	0.1042	-0.0267	0.1045	-0.0255	0.0351	0.0011
	100	50	0.0278	-0.0012	0.0278	-0.0012	0.0418	-0.0130	0.0418	-0.0130	0.0242	-0.0010
	200	15	0.0712	-0.0141	0.0743	-0.0145	0.1491	-0.0399	0.1483	-0.0383	0.0352	-0.0017
	200	25	0.0333	-0.0051	0.0333	-0.0051	0.0932	-0.0284	0.0932	-0.0284	0.0252	-0.0006
	200	50	0.0193	-0.0014	0.0193	-0.0014	0.0277	-0.0134	0.0277	-0.0134	0.0164	0.0000
DGP 3	100	25	0.1722	0.0587	0.1516	0.0727	0.2154	0.0615	0.1641	0.0688	0.1077	0.0114
	100	50	0.0853	0.0379	0.0878	0.0383	0.1178	0.0487	0.1191	0.0489	0.0752	0.0090
	200	25	0.1342	0.0483	0.1401	0.0649	0.1826	0.0487	0.1441	0.0573	0.0821	0.0116
	200	50	0.0632	0.0264	0.0632	0.0264	0.0948	0.0372	0.0948	0.0372	0.0573	0.0121
DGP 4	100	15	0.1691	0.0487	0.1803	0.0376	0.2148	0.1087	0.2102	0.0941	0.0806	0.0018
	100	25	0.0724	0.0189	0.1217	0.0207	0.0882	0.0523	0.1323	0.0539	0.0617	0.0009
	100	50	0.0450	0.0031	0.0645	0.0042	0.0532	0.0204	0.0707	0.0215	0.0437	-0.0012
	200	15	0.1271	0.0512	0.1348	0.0466	0.1777	0.1128	0.1793	0.1074	0.0577	-0.0016
	200	25	0.0513	0.0153	0.1392	0.0235	0.0720	0.0498	0.1485	0.0577	0.0436	0.0001
	200	50	0.0314	0.0036	0.0549	0.0049	0.0399	0.0221	0.0602	0.0234	0.0311	-0.0005
DGP 2	100	15	0.0482	0.0081	0.0487	0.0065	0.0747	0.0297	0.0715	0.0254	0.0352	0.0041
(PLS)	100	25	0.0263	0.0043	0.0263	0.0043	0.0418	0.0189	0.0418	0.0189	0.0241	0.0028
	100	50	0.0160	0.0016	0.0160	0.0016	0.0218	0.0085	0.0218	0.0085	0.0156	0.0013
	200	15	0.0295	0.0064	0.0295	0.0064	0.0567	0.0293	0.0567	0.0293	0.0253	0.0052
	200	25	0.0188	0.0037	0.0188	0.0037	0.0307	0.0174	0.0307	0.0174	0.0178	0.0036
	200	50	0.0113	0.0017	0.0113	0.0017	0.0171	0.0084	0.0171	0.0084	0.0111	0.0015

Table S4: Summary statistics for the savings dataset

	mean	median	s.e.	min	max
Savings rate	22.099	20.790	8.833	-3.207	53.434
Inflation rate	7.724	4.853	15.342	-3.846	293.679
Real interest rate	7.422	5.927	10.062	-63.761	93.915
Per capita GDP growth rate	2.855	2.971	3.865	-17.545	14.060

Table S5: Estimated group membership

PLS	PGMM
Group 1: (31 countries) , , , , , , , , , Guatemala, , , , , , , , , , , , , , Mongolia, , , , , , , ,	Group 1: (36 countries) , , , , , , , , , Egypt, , , India, , , , , , Kenya, , , , Malta, , , , , , , , , Swaziland, Switzerland, , , United Kingdom
Group 2: (25 countries) , , , , Egypt, , , India, Kenya, , , Malta, , , , , , Swaziland, Switzerland, , , , United Kingdom, ,	Group 2: (20 countries) , , , , Guatemala, , , , , Mongolia, , , , , , , , , ,

The replication data of Fearon and Laitin (2003) can be downloaded from Fearon's personal web page.[5] The data span from 1945–1998,[6] but the panel is highly unbalanced. Following Collier and Hoeffler (2004), Djankov and Reynal-Querol (2010) and Blattman and Miguel (2010), we choose 1960 as the staring year to generate a balanced panel of = 38, as many countries' civil war incidence are always 0 or 1 between 1960 and 1998.

In the regression, the dependent variable is the civil war incidence, and the explanatory variables are the lagged civil war incidence, the one-period difference of log GDP per capita and the one-period difference of log population. Moreover, in views of the natural scaling of the binary variable, we keep the original dependent variable and the lagged dependent variable. For the other two continuously distributed variables, we follow the practice as in the savings rate application to scale-normalize each time series by the individual sample standard deviation. To ensure that the estimated coefficients are comparable, we further multiply these two scale-normalized variables by the overall standard deviation of the lagged dependent variable so that all the explanatory regressors are of the same scale. Furthermore, the Probit regressions for the individual time series are unstable in those countries with only 1 or 2 incidences. Therefore the C-Lasso initial values are set as the pooled FE Probit coefficient estimates.

The summary statistics are displayed in Table S6. Membership is reported under "high-occurrence" and

[5] `https://www.stanford.edu/group/fearon-research/cgi-bin/wordpress/wp-content/uploads/2013/10/apsr03repdata.zip`

[6] The original data end at 1999, but no population information is provided for any country in the last year.

Table S6: Summary statistics for the civil war dataset

	mean	median	s.e.	min	max
Civil war incidence	0.352	0	0.478	0	1
GDP per capita growth	0.020	0.024	0.040	-0.811	0.306
Population growth	0.012	0.015	0.076	-0.507	0.661

"low-occurrence" groups with results as follows.

(23 countries): Guatemala, Peru, Argentina, Mali, Senegal, Chad, Congo (Dem.), Congo (Rep.), Somalia, Morocco, Sudan, Turkey, Iraq, Lebanon, Afghanistan, China, Pakistan, Sri Lanka, Nepal, Cambodia, Laos, Philippines, Indonesia

(15 countries): Haiti, Dominican, El Salvador, Nicaragua, UK, Yugoslavia, Cyprus, Russia, Liberia, Nigeria, Central African Republic, Ethiopia, South Africa, Iran, Jordan

In this section, we use the data provided by Bonhomme and Manresa (2015) to revisit the link between income growth and democracy across countries. Following BM's Equation (22), we specify a linear dynamic model, where the dependent variable is a country's democracy index (measured by Freedom House indicator between 0 (the lowest) and 1 (the highest)), and the explanatory variables are the first-order lagged democracy index and the income (measured by the logarithm of GDP per capita).

The dataset contains a balanced panel of 84 countries and 8 periods at a five year interval over 1965–2000. We use PLS to estimate the model in this short panel. Many developed countries, such as the United States or United Kingdom, kept their democracy index at the highest level throughout the time. Due to the lack of within-group variation in these countries, we scale normalize each variable by its pooled standard deviation. This standardization makes sure that the parameter $_{it-1}$ can still be interpreted as the autoregressive coefficient, and the magnitude is comparable with the income coefficient.

Table S7: Summary statistics for the democracy dataset

	mean	median	s.e.	min	max
Democracy index	0.5760	0.6667	0.3712	0	1
GDP per capita (in logarithm)	8.2981	8.3039	1.0685	6.0937	10.4450

Following practice in the simulation, the IC with $_{1NT} = \frac{2}{3}(\quad)^{-1/2}$ picks out $= 3$ and $_{\lambda_1} = 1\ 20$ in all combinations of $= 1 \quad 5$ and $_{\lambda_1}$ in a geometrically increasing sequence of 10 points in (0 2 $\quad$ 2). Under $= 3$ and $_{\lambda_1} = 1\ 20$, C-Lasso categorizes the 84 countries into the following groups:

(30 countries): Belgium, Bolivia, Brazil, Canada, Dominican, Ecuador, El Salvador, Finland, Guatemala, Guinea, Iceland, Indonesia, Italy, Japan, Jordan, Luxembourg, Mali, Morocco, Nepal, Panama, Peru, Philippines, Portugal, Romania, South Africa, Thailand, Turkey, United Kingdom, Uruguay, Venezuela

(36 countries): Algeria, Argentina, Australia, Austria, Barbados, Burkina Faso, Burundi, Cameroon, Chile, China, Colombia, Costa Rica, Cote d'Ivoire, Denmark, Egypt, France, Gabon, Ghana, Greece, India, Iran, Israel, Jamaica, Kenya, Malawi, Malaysia, Mexico, Nigeria, Norway, Paraguay, Rwanda, Spain, Sweden, Togo, Trinidad and Tobago, United States

(18 countries): Benin, Chad, Congo (Rep.), Honduras, Ireland, Korea (Rep.), Madagascar, Netherlands, New Zealand, Nicaragua, Niger, Sri Lanka, Switzerland, Syrian, Tanzania, Tunisia, Uganda, Zambia

The post-Lasso and pooled FE estimates are shown in Table S8. We focus on the coefficient for income. The common FE coefficient is positive and significant. The positive effect is echoed by Groups 1 and 2, but contrasts with Group 3, which consists mainly of low-income and low-democracy nations combined with a few selected OECD countries. OECD countries such as Ireland, Netherlands, New Zealand and Switzerland maintained their democracy index at 1 throughout the sample period. The lack of variation in the dependent variable makes them uninformative about the income coefficient.

Table S8: PLS estimation results

	Pooled FE		PLS					
			Group 1		Group 2		Group 3	
	coef.	s.e.	coef.	s.e.	coef.	s.e.	coef.	s.e.
Lagged democracy	. ***	0.0491	. ***	0.0643	.	0.0733	− .	0.0521
Income	. ***	0.0489	. ***	0.0930	. ***	0.0448	− . ***	0.0860

Note: ***1% significant, ** 5% significant, * 10% significant

BLATTMAN C. AND E. MIGUEL (2010): "Civil Wars," *Journal of Economic Literature* 48, 3-57.

BONHOMME, S., AND E. MANRESA (2015): "Grouped Patterns of Heterogeneity in Panel Data," *Econometrica* 83, 1147-1184.

COLLIER, P. AND A. HOEFFLER (2004): "Greed and Grievance in Civil Wars," *Oxford Economic Papers* 56, 563-595.

DJANKOV, S. AND M. REYNAL-QUEROL (2010): "Poverty and Civil War: Revisiting the Evidence," *Review of Economics and Statistics* 92, 1035-1041.

EFRON, B., T. HASTIE, I. JOHNSTONE and R. TIBSHIRANI (2004), "Least Angle Regression," *Annals of Statistics* 32, 407-499.

FEARON, J.D. AND D. D. LAITIN (2003): "Ethnicity, Insurgency, and Civil War," *The American Political Science Review* 97, 75-90.

HAHN, J., AND G. KUERSTEINER (2011): "Bias Reduction for Dynamic Nonlinear Panel Models with Fixed Effects," *Econometric Theory* 27, 1152-1191.

HAHN, J., AND W. NEWEY (2004): "Jackknife and Analytical Bias Reduction for Nonlinear Panel Models," *Econometrica* 72, 1295-1319.

HALL, P. AND C.C. HEYDE (1980): *Martingale Limit Theory and Its Applications.* Academic Press, New York.

MERLEVÈDE, F., M. PEILGRAD, M., AND E. RIO (2009): "Bernstein Inequality and Moderate Deviations under Strong Mixing Conditions," *IMS collections: High Dimensional Probability V.*, 273-292.

WHITE, H. (2001): *Asymptotic Theory for Econometricians.* Emerald, UK.