



# Feasible generalized least squares for panel data with cross-sectional and serial correlations

Jushan Bai<sup>1</sup> · Sung Hoon Choi<sup>2</sup> · Yuan Liao<sup>2</sup> 

Received: 21 March 2020 / Accepted: 30 October 2020 / Published online: 9 November 2020  
© Springer-Verlag GmbH Germany, part of Springer Nature 2020

## Abstract

This paper considers generalized least squares (GLS) estimation for linear panel data models. By estimating the large error covariance matrix consistently, the proposed feasible GLS estimator is more efficient than the ordinary least squares in the presence of heteroskedasticity, serial and cross-sectional correlations. The covariance matrix used for the feasible GLS is estimated via the banding and thresholding method. We establish the limiting distribution of the proposed estimator. A Monte Carlo study is considered. The proposed method is applied to an empirical application.

**Keywords** Panel data · Efficiency · Thresholding · Banding · Cross-sectional correlation · Serial correlation · Heteroskedasticity

## 1 Introduction

Heteroskedasticity, cross-sectional and serial correlations are important problems in the error terms of panel regression models. There are two approaches to deal with these problems. The first approach is to use the ordinary least squares (OLS) estimator but with a robust standard error that is robust to heteroskedasticity and correlations, for example, White (1980); Newey and West (1987); Liang and Zeger (1986); Arel-

---

**Electronic supplementary material** The online version of this article (<https://doi.org/10.1007/s00181-020-01977-2>) contains supplementary material, which is available to authorized users.

---

✉ Yuan Liao  
yuan.liao@rutgers.edu

Jushan Bai  
jb3064@columbia.edu

Sung Hoon Choi  
shchoi@economics.rutgers.edu

<sup>1</sup> Columbia University, 420 West 118th St. MC 3308, New York, NY 10027, USA

<sup>2</sup> Rutgers University, 75 Hamilton St., New Brunswick, NJ 08901, USA

lano (1987); Driscoll and Kraay (1998); Hansen (2007a); Vogelsang (2012), among others. A widely used class of robust standard errors is clustered standard errors, for example, Petersen (2009), Wooldridge (2010) and Cameron and Miller (2015). Bai et al. (2019) proposed a robust standard error with unknown clusters. In an interesting paper by Abadie et al. (2017), they argued for caution in the application of clustered standard errors since they may give rise to conservative confidence intervals. The second approach is to use the generalized least squares estimator (GLS) that directly takes into account heteroskedasticity, and cross-sectional and serial correlations in the estimation. It is well known that GLS is more efficient than OLS.

This paper focuses on the second approach. For panel models, the underlying covariance matrix involves a large number of parameters. It is important to make GLS operational. We thus consider feasible generalized least squares (FGLS). Hansen (2007b) studied FGLS estimation that takes into account serial correlation and clustering problems in fixed effects panel and multilevel models. His approach requires the cluster structure to be known. This gives motivation to our paper. We assume the unknown cluster structure, and control heteroskedasticity, both serial and cross-sectional correlations by estimating the large error covariance matrix consistently. In cross-sectional setting, Romano and Wolf (2017) obtained asymptotically valid inference of the FGLS estimator, combined with heteroskedasticity-consistent standard errors without knowledge of the conditional heteroskedasticity functional form. Moreover, Miller and Startz (2018) adapted machine learning methods (i.e., support vector regression) to take into account the misspecified form of heteroskedasticity.

In this paper, we consider (i) balanced panel data, (ii) the case of large- $N$  large- $T$ , and (iii) both serial and cross-sectional correlations, but unknown structure of clusters. We introduce a modified FGLS estimator that eliminates the cross-sectional and serial correlation bias by proposing a high-dimensional error covariance matrix estimator. In addition, our proposed method is applicable when the knowledge of clusters is not available. Let  $u_t$  be an  $N \times 1$  vector of regressor noises, whose definition is to be clear later. Following the idea of Bai and Liao (2017), in this paper, the FGLS involves estimating an  $NT \times NT$  dimensional inverse covariance matrix  $\Omega^{-1}$ , where

$$\Omega = (Eu_t u_t')$$

where each block  $Eu_t u_t'$  is an  $N \times N$  autocovariance matrix. Here, parametric structures on the serial or cross-sectional correlations are not imposed. By assuming weak dependences, we apply nonparametric methods to estimate the covariance matrix. To address the estimation of serial autocorrelations, we employ the idea of Newey–West truncation. This method, in the FGLS setting, is equivalent to “banding”, previously proposed by Bickel and Levina (2008b) for estimating large covariance matrices. We apply it to banding out off-diagonal  $N \times N$  blocks that are far from the diagonal block. In addition, to control for the cross-sectional correlation, we assume that each of the  $N \times N$  block matrices are sparse, potentially resulting from the presence of cross-sectional correlations within clusters. We then estimate them by applying the thresholding approach of Bickel and Levina (2008a). We apply thresholding separately to the  $N \times N$  blocks, which are formed by time lags  $Eu_t u_{t-h}' : h = 0, 1, 2, \dots$ . This allows the cluster-membership to be potentially changing over-time. A contribution

of this paper is the theoretical justification for estimating the large error covariance matrix.

For the FGLS, it is crucial for the asymptotic analysis to prove that the effect of estimating  $\Omega$  is first-order negligible. In the usual low-dimensional settings that involve estimating optimal weight matrix, such as the optimal GMM estimations, it has been well known that consistency for the inverse covariance matrix estimator is sufficient for the first-order asymptotic theory, e.g., Hansen (1982), Newey (1990), Newey and McFadden (1994). However, it turns out that when the covariance matrix is of high-dimensions, not even the optimal convergence rate for estimating  $\Omega^{-1}$  is sufficient. In fact, proving the first-order equivalence between the FGLS and the infeasible GLS (that uses the true  $\Omega^{-1}$ ) is a very challenging problem under the large  $N$ , large  $T$  setting. We provide a new theoretical argument to achieve this goal.

The banding and thresholding methods, which we employ in this paper, are two useful regularization methods. In the recent machine learning literature, these methods have been extensively exploited for estimating high-dimensional parameters. Moreover, in the econometric literature, nonparametric machine learning techniques have been verified to be powerful tools: Bai and Ng (2017); Chernozhukov et al. (2016); Chernozhukov et al. (2017); Wager and Athey (2018), etc.

The rest of the paper is organized as follows. In Section 2, we describe the model and the large error covariance matrix estimator. Also, we introduce the implementation of FGLS estimator and its limiting distribution. In Sect. 3, we apply our methods to study the US divorce rate problem. Conclusions are provided in Sect. 4. All proofs and Monte Carlo studies are given in the online supplement.

Throughout this paper, let  $v_{\min}(A)$  and  $v_{\max}(A)$  denote the minimum and maximum eigenvalues of matrix  $A$ , respectively. Also, we use  $\|A\| = \sqrt{v_{\max}(A'A)}$ ,  $\|A\|_1 = \max_i \sum_j |A_{ij}|$  and  $\|A\|_F = \sqrt{\text{tr}(A'A)}$  as the operator norm,  $\ell_1$ -norm and the Frobenius norm of a matrix  $A$ , respectively. Note that if  $A$  is a vector,  $\|A\| = \|A\|_F$  is equal to the Euclidean norm.

## 2 Feasible generalized least squares

We consider a linear model<sup>1</sup>

$$y_{it} = x'_{it}\beta + u_{it}. \quad (1)$$

The model (1) can be stacked and represented in full matrix notation as

$$Y = X\beta + U, \quad (2)$$

where  $Y = (y'_1, \dots, y'_T)'$  is the  $NT \times 1$  vector of  $y_{it}$  with each  $y_t$  being an  $N \times 1$  vector;  $X = (x'_1, \dots, x'_T)'$  is the  $NT \times d$  matrix of  $x_{it}$  with each  $x_t$  being an  $N \times d$ ;  $U = (u'_1, \dots, u'_T)'$  is the  $NT \times 1$  vector of  $u_{it}$  with each  $u_t$  being an  $N \times 1$  vector.

<sup>1</sup> For technical simplicity, we focus on a simple model where there are no fixed effects. It is straightforward to allow additive fixed effects  $\alpha_i + \mu_t$  by applying the de-meaning first. The theories would be slightly more sophisticated, though such extensions are straightforward.

Let  $\Omega = (Eu_t u'_s)$  be an  $NT \times NT$  matrix, consisting of many blocks matrices. The  $(t, s)$ th block is an  $N \times N$  covariance matrix  $Eu_t u'_s$ . We consider the following (infeasible) GLS estimator of  $\beta$ :

$$\tilde{\beta}_{GLS}^{inf} = (X' \Omega^{-1} X)^{-1} X' \Omega^{-1} Y. \tag{3}$$

Note that  $\Omega$  is a high-dimensional conditional covariance matrix, which is very difficult to estimate. We aim to achieve the following: (i) obtain a “good” estimator of  $\Omega^{-1}$ , allowing an arbitrary form of weak dependence in  $u_{it}$ , and (ii) show that the effect of replacing  $\Omega^{-1}$  by  $\tilde{\Omega}^{-1}$  is asymptotically negligible.

We start with a population approximation for  $\Omega$  in order to gain the intuitions. Then, we suggest the estimator for  $\Omega$  that takes into account both serial and cross-sectional correlations.

### 2.1 Population approximation

We start with a “banding” approximation to control serial correlations. Recall that  $\Omega = (Eu_t u'_s)$ , where the  $(t, s)$  block is  $Eu_t u'_s$ . By assuming serial stationarity and strong mixing condition,  $Eu_t u'_s$  depends on  $(t, s)$  only through  $h = t - s$ . Specifically, with slight abuse of notation, we can write  $\Omega_{t,s} = \Omega_h = Eu_t u'_{t-h}$ . Note for  $i \neq j$ , it is possible that  $Eu_{it} u_{jt-t-h} \neq Eu_{it-t-h} u_{jt}$ , so  $\Omega_h$  is possibly non-symmetric for  $h > 0$ . On the other hand,  $\Omega$  is symmetric due to  $\Omega_{s,t} = \Omega'_{t,s}$ . The diagonal blocks are the same, and all equal  $\Omega_0 = Eu_t u'_t$ , while magnitudes of the elements of the off-diagonal blocks  $\Omega_h = Eu_t u'_{t-h}$  decay to zero as  $|h| \rightarrow \infty$  under the weak serial dependence assumption.

In the Newey–West spirit,  $\Omega$  can be approximated by  $\Omega^{NW} = (\Omega_{t,s}^{NW})$ , where each block can be written as  $\Omega_{t,s}^{NW} = \Omega_h^{NW}$  for  $h = t - s$ . Here,  $\Omega_h^{NW}$  is an  $N \times N$  block matrix, defined as:

$$\Omega_h^{NW} = \begin{cases} Eu_t u'_{t-h}, & \text{if } |h| \leq L \\ 0, & \text{if } |h| > L, \end{cases}$$

for some pre-determined  $L \rightarrow \infty$ . For instance, as suggested by Newey and West (1994), we can set  $L$  equal to  $4(T/100)^{(2/9)}$ . Note that  $\Omega_h^{NW} = \Omega_{-h}^{NW}$ . We regard  $\Omega^{NW} = (\Omega_h^{NW})$  as the “population banding approximation.”

Next, we focus on the  $N \times N$  block matrix  $\Omega_h = Eu_t u'_{t-h}$  to control cross-sectional correlations. Under the intuition that  $u_{it}$  is cross-sectional weakly dependent, we assume  $\Omega_h$  is a sparse matrix, that is,  $\Omega_{h,ij} = Eu_{it} u_{j,t-h}$  is “small” for “many” pairs  $(i, j)$ . Then,  $\Omega_h$  can be approximated by a sparse matrix  $\Omega_h^{BL} = (\Omega_{h,ij}^{BL})_{N \times N}$  (Bickel and Levina (2008a)), where

$$\Omega_{h,ij}^{BL} = \begin{cases} Eu_{it} u_{j,t-h}, & \text{if } |Eu_{it} u_{j,t-h}| > \tau_{ij} \\ 0, & \text{if } |Eu_{it} u_{j,t-h}| \leq \tau_{ij}, \end{cases}$$

for some pre-determined threshold  $\tau_{ij} \rightarrow 0$ . We regard  $\Omega_h^{BL}$  as the ‘‘population sparse approximation.’’

In summary, we approximate  $\Omega$  by an  $NT \times NT$  matrix  $(\tilde{\Omega}_{t,s}^{NT})$ , where each block  $\tilde{\Omega}_{t,s}^{NT}$  is an  $N \times N$  matrix, defined as: for  $h = t - s$ ,

$$\tilde{\Omega}_{t,s}^{NT} := \begin{cases} \Omega_h^{BL}, & \text{if } |h| \leq L \\ 0, & \text{if } |h| > L. \end{cases}$$

Therefore, we use ‘‘banding’’ to control the serial correlation, and ‘‘sparsity’’ to control the cross-sectional correlation.

## 2.2 Implementation of feasible GLS

### 2.2.1 The estimator of $\Omega$ and FGLS

Given the intuition of the population approximation, we construct the large covariance estimator as follows. First, we denote the OLS estimator of  $\beta$  by  $\hat{\beta}_{OLS}$  and the corresponding residuals by  $\hat{u}_{it} = y_{it} - x'_{it}\hat{\beta}_{OLS}$ .

Now, we estimate the  $N \times N$  block matrix  $\Omega_h = Eu_tu'_{t-h}$ . To do so, let

$$\tilde{R}_{h,ij} = \begin{cases} \frac{1}{T} \sum_{t=h+1}^T \hat{u}_{it}\hat{u}_{j,t-h}, & \text{if } h \geq 0 \\ \frac{1}{T} \sum_{t=1}^{T+h} \hat{u}_{it}\hat{u}_{j,t-h}, & \text{if } h < 0 \end{cases}, \quad \text{and } \tilde{\sigma}_{h,ij} = \begin{cases} \tilde{R}_{h,ii}, & \text{if } i = j \\ s_{ij}(\tilde{R}_{h,ij}), & \text{if } i \neq j, \end{cases}$$

where  $s_{ij}(\cdot) : \mathbb{R} \rightarrow \mathbb{R}$  is a ‘‘soft-thresholding function’’ with an entry-dependent threshold  $\tau_{ij}$  such that

$$s_{ij}(z) = \text{sgn}(z)(|z| - \tau_{ij})_+,$$

where  $(x)_+ = x$  if  $x \geq 0$ , and zero otherwise. Here,  $\text{sgn}(\cdot)$  denotes the sign function, and other thresholding functions, e.g., hard thresholding, are possible. For the threshold value, we specify

$$\tau_{ij} = M\gamma_T \sqrt{|\tilde{R}_{0,ii}| |\tilde{R}_{0,jj}|},$$

for some pre-determined value  $M > 0$ , where  $\gamma_T = \sqrt{\frac{\log(LN)}{T}}$  is such that  $\max_{h \leq L} \max_{i,j \leq N} |\tilde{R}_{h,ij} - Eu_{it}u_{i,t-h}| = O_P(\gamma_T)$ . Note that here we use an entry-dependent threshold  $\tau_{ij}$ , which may vary across  $(i, j)$ . Then, define

$$\tilde{\Omega}_h = (\tilde{\sigma}_{h,ij})_{N \times N}. \tag{4}$$

Next, we define the  $(t, s)$ th block  $\widehat{\Omega}_{t,s}$  as an  $N \times N$  matrix: for  $h = t - s$ ,

$$\widehat{\Omega}_{t,s} = \begin{cases} \omega(|h|, L)\widetilde{\Omega}_h, & \text{if } |h| \leq L \\ 0, & \text{if } |h| > L. \end{cases}$$

Here,  $\omega(h, L)$  is the kernel function (see Andrews (1991) and Newey and West (1994)). We let  $\omega(h, L) = 1 - h/(L + 1)$  be the Bartlett kernel function, where  $L$  is the bandwidth. Our final estimator of  $\Omega$  is an  $NT \times NT$  matrix:

$$\widehat{\Omega} = (\widehat{\Omega}_{t,s}).$$

Here,  $\widehat{\Omega}$  is a nonparametric estimator, which does not require an assumed parametric structure on  $\Omega$ .

Finally, given  $\widehat{\Omega}$ , we propose the feasible GLS (FGLS) estimator of  $\beta$  as

$$\widehat{\beta}_{FGLS} = [X'\widehat{\Omega}^{-1}X]^{-1}X'\widehat{\Omega}^{-1}Y.$$

Note that the above defined FGLS estimator leaves two quantities to be specified to applied researchers: (i) the constant  $M > 0$  in the threshold value for  $\tau_{ij}$ , and (ii) the Newey–West bandwidth  $L$ . We discuss the choice of these two quantities in Sect. 2.2.2 below.

**Remark 2.1 (Universal thresholding)** We apply thresholding separately to the  $N \times N$  blocks,  $(\widetilde{\sigma}_{h,ij})_{N \times N}$ , which are estimated lagged blocks for  $Eu_tu_{t-h} : h = 0, 1, 2, \dots$ . This allows the cluster-membership to be potentially changing over-time, that is, the identities of zeros and nonzero elements of  $Eu_tu_{t-h}$  can change over  $h$ . If it is known that the cluster-membership (i.e., identities of nonzero elements) is time-invariant, then one would set  $\widetilde{\sigma}_{h,ij} = 0$  if  $\max_{h \leq L} |\widetilde{R}_{h,ij}| \leq \tau_{ij}$  for  $i \neq j$ . This potentially would increase the finite sample accuracy of identifying the cluster-membership.

## 2.2.2 Choice of tuning parameters

Our suggested covariance matrix estimator,  $\widehat{\Omega}$ , requires the choice of tuning parameters  $L$  and  $M$ , which are the bandwidth and the threshold constant, respectively. We write  $\widehat{\Omega}(M, L) = \widehat{\Omega}$ , where the covariance estimator depends on  $M$  and  $L$ . First, to choose the bandwidth  $L$ , we suggest using  $L^* = 4(T/100)^{2/9}$ , which is proposed by Newey and West (1994). For a small size of  $T$ , we also recommend  $L \leq 3$ .

As for the choice of the thresholding constant  $M$ , our recommended rule-of-thumb choice is any constant that is on the interval  $[0.5, 2]$ . Based on our simulations in extensive studies with various values for  $N$  and  $T$ , we find that  $M = 1.8$  is a universally good choice.

Alternatively,  $M$  can also be chosen through multifold cross-validation. To discuss this procedure, let us randomly split the data  $P$  times. We divide the data into  $P = \log(T)$  blocks  $J_1, \dots, J_P$  with block length  $T/\log(T)$  and take one of the  $P$  blocks as the validation set. At the  $p$ th split, we denote by  $\widetilde{\Omega}_0^p$  the sample covariance matrix

based on the validation set, defined by  $\tilde{\Omega}_0^p = |J_p|^{-1} \sum_{t \in J_p} \hat{u}_t \hat{u}_t'$ . Let  $\tilde{\Omega}_0^{S,p}(M)$  be the thresholding estimator with threshold constant  $M$  using the training data set  $\{\hat{u}_t\}_{t \notin J_p}$ . Finally, we choose the constant  $M^*$  by minimizing the cross-validation objective function

$$M^* = \arg \min_{c < M < \bar{C}} \frac{1}{P} \sum_{j=1}^P \|\tilde{\Omega}_0^{S,p}(M) - \tilde{\Omega}_0^p\|_F^2,$$

where  $\bar{C}$  is a large constant such that  $\tilde{\Omega}_0^S(\bar{C})$  is a diagonal matrix and can be fixed as, e.g.,  $\bar{C} = 3$ ;  $c$  is a constant that guarantees the positive definiteness of  $\hat{\Omega}(M, L)$  for  $M > c$ : for each fixed  $L$ ,

$$c = \inf[M > 0 : \lambda_{\min}\{\hat{\Omega}(C, L)\} > 0, \forall C > M].$$

Here,  $\tilde{\Omega}_0^S(M)$  is the soft-thresholded estimator as defined in the Eq. (4). Then, the resulting estimator of  $\Omega$  is  $\hat{\Omega}(M^*, L^*)$ . To determine this value, one can plot  $\lambda_{\min}\{\hat{\Omega}(C, L)\}$  as a function of  $C$ , fixing  $L = L^*$  and visually determine  $c$ .

In summary, Table 1 summarizes the recommended quantities for implementing the proposed FGLS estimator.

### 2.2.3 Incorporating known clusters

Note that an advantage of the method proposed in this paper is that it does not assume known cluster information (i.e., the number of clusters and the membership of clusters). On the other hand, when clustering information is available, this method can be modified to take into account that information and is particularly suitable when the number of clusters is small, and the size of each cluster is large.

For example, let  $C_1, \dots, C_G$  be disjoint subsets of  $\{1, \dots, N\}$ , so that they are *known* clusters and that  $u_{it}$  and  $u_{js}$  are uncorrelated if  $i$  and  $j$  belong to different clusters for any  $(t, s)$ . Then, naturally we can re-arranged the  $N \times N$  matrix  $\Omega_h = E u_t u_{t-h}'$  so that it can be decomposed into  $G$  disjoint blocks on the diagonal and off-diagonal blocks are zeros:

$$\Omega_h = \begin{pmatrix} \Omega_{h,1} & & & \\ & \ddots & & \\ & & \ddots & \\ & & & \Omega_{h,G} \end{pmatrix}.$$

It is assumed that  $G$  is small while the size of each diagonal block matrix is large. Within the  $g$ th ( $g \leq G$ ) diagonal block matrix, say  $\Omega_{h,g}$ , we apply thresholding to further reduce the dimensionality. So we estimate  $\Omega_{h,g}$  by  $\tilde{\Omega}_{h,g} = (\tilde{\sigma}_{h,g,ij})$ , where

$$\tilde{\sigma}_{h,g,ij} = \begin{cases} \tilde{R}_{h,ii}, & \text{if } i = j, \text{ and } i, j \in C_g \\ s_{ij}(\tilde{R}_{h,ij}), & \text{if } i \neq j, \text{ and } i, j \in C_g. \end{cases}$$

**Table 1** Recommended choices for implementations

Quantities	$s_{ij}(z)$	$\omega( h , L)$	$\tau_{ij}$	$\gamma T$	$L$
Choice	$\text{sgn}(z)( z  - \tau_{ij}) +$	$1 - \frac{ h }{L+1}$	$M\gamma T \sqrt{ \tilde{R}_{0,ii}   \tilde{R}_{0,jj} }$	$\sqrt{\frac{\log(LN)}{T}}$	$4(T/100)^{2/9}$
Quantities	$M$ (rule-of-thumb)	$P$	$\bar{C}$	$c$	
Choice	1.8	$\log T$	3	visually by plotting $\lambda_{\min} \{\hat{\Omega}(C, L)\}$	

Here,  $P$ ,  $\bar{C}$  and  $c$  are required constants for the choice of  $M$  based on cross-validations



Putting these estimated diagonal blocks together, we obtain  $\tilde{\Omega}_h$ , the estimated  $\Omega_h$ .

The within-cluster thresholding then allows unknown correlations within each cluster. In contrast, the conventional clustered standard errors lose a lot of degrees of freedom when the size of cluster is too large (because each cluster is effectively treated as a “single observation”), resulting in conservative confidence intervals. See Cameron and Miller (2015) for more discussions.

Moreover, when the number of clusters is large, and the size of each cluster is small, then this is the usual setting of cluster standard errors. One does not need to apply thresholding, as the known clusters naturally form small diagonal blocks on  $\Omega_h$ . Because the size of these blocks are small, sufficient degrees of freedom is kept and it is then straightforward to estimate  $\Omega_h$ .

### 2.3 The effect of $\hat{\Omega}^{-1} - \Omega^{-1}$

A key step of proving the asymptotic property for  $\hat{\beta}_{FGLS}$  is to show that it is asymptotically equivalent to  $\tilde{\beta}_{GLS}^{inf}$ , that is:

$$\frac{1}{\sqrt{NT}} X'(\hat{\Omega}^{-1} - \Omega^{-1})U = o_P(1). \tag{5}$$

In the usual low-dimensional settings that involve estimating optimal weight matrix, such as the optimal GMM estimations, it has been well known that consistency for the inverse covariance matrix estimator is sufficient for the first-order asymptotic theory, e.g., Hansen (1982), Newey (1990), Newey and McFadden (1994). It turns out, when the covariance matrix is of high-dimensions, not even the optimal convergence rate of  $\|\hat{\Omega} - \Omega\|$  is sufficient. In fact, proving equation (5) is a very challenging problem. In the general case when both cross-sectional and serial correlations are present, our strategy is to use a careful expansion for  $\frac{1}{\sqrt{NT}} X'(\hat{\Omega}^{-1} - \Omega^{-1})U$ . We shall proceed in two steps:

Step 1: Show that  $\frac{1}{\sqrt{NT}} X'(\hat{\Omega}^{-1} - \Omega^{-1})U = \frac{1}{\sqrt{NT}} W'(\hat{\Omega} - \Omega)\varepsilon + o_P(1)$ , where  $W = \Omega^{-1}X$ , and  $\varepsilon = \Omega^{-1}U$ .

Step 2: Show that  $\frac{1}{\sqrt{NT}} W'(\hat{\Omega} - \Omega)\varepsilon = o_P(1)$ .

Now, we suppose  $\omega(h, L) = 1$ ,  $\Omega \approx \Omega^{NW}$  and let  $A_{bh} = \{(i, j) : |Eu_{it}u_{j,t-h}| \neq 0\}$ ,  $A_{sh} = \{(i, j) : |Eu_{it}u_{j,t-h}| = 0\}$ . As for Step 2, we shall show,

$$\begin{aligned} \frac{1}{\sqrt{NT}} W'(\hat{\Omega} - \Omega)\varepsilon &\approx \frac{1}{\sqrt{NT}} \sum_{|h| \leq L} \sum_{i, j \in A_{bh}} \sum_{t=h+1}^T w_{it} \varepsilon_{j,t-h} \frac{1}{T} \\ &\quad \sum_{s=h+1}^T (u_{is}u_{j,s-h} - Eu_{it}u_{j,t-h}). \end{aligned} \tag{6}$$

Here,  $w_{it}$  is defined such that, we can write  $W = (w'_1, \dots, w'_T)'$  with  $w_t$  being an  $N \times d$  matrix of  $w_{it}$ ;  $\varepsilon_{it}$  is defined similarly. While proving (6) to be  $o_P(1)$ , in the

presence of both serial and cross-sectional correlations, is very technically challenging. We thus directly assume it is  $o_P(1)$  as a high-level condition (see Assumption 2.4 in Sect. 2.4 below). To appreciate the need of this high-level condition, let us consider a simple example as follows.

*A simple example* To illustrate the key technical issue, consider a simple and ideal case where  $u_{it}$  is known, and independent across both  $i$  and  $t$ , but with cross-sectional heteroskedasticity. In this case, the covariance matrix of the  $NT \times 1$  vector  $U$  is a diagonal matrix, with diagonal elements  $\sigma_i^2 = Eu_{it}^2$ :

$$\Omega = \begin{pmatrix} D & & \\ & \ddots & \\ & & D \end{pmatrix}, \text{ where } D = \begin{pmatrix} \sigma_1^2 & & \\ & \ddots & \\ & & \sigma_N^2 \end{pmatrix}.$$

Then, a natural estimator for  $\Omega$  is

$$\hat{\Omega} = \begin{pmatrix} \hat{D} & & \\ & \ddots & \\ & & \hat{D} \end{pmatrix}, \text{ where } \hat{D} = \begin{pmatrix} \hat{\sigma}_1^2 & & \\ & \ddots & \\ & & \hat{\sigma}_N^2 \end{pmatrix},$$

and  $\hat{\sigma}_i^2 = \frac{1}{T} \sum_{t=1}^T u_{it}^2$ , because  $u_{it}$  is known. Then, the GLS becomes:

$$\left( \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T x_{it} x'_{it} \hat{\sigma}_i^{-2} \right)^{-1} \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T x_{it} y_{it} \hat{\sigma}_i^{-2}.$$

A key step is to prove that the effect of estimating  $D$  is asymptotically negligible:

$$\frac{1}{\sqrt{NT}} \sum_{i=1}^N \sum_{t=1}^T x_{it} u_{it} (\hat{\sigma}_i^{-2} - \sigma_i^{-2}) = o_P(1). \tag{7}$$

It can be shown that the problem reduces to proving:

$$A \equiv \frac{1}{\sqrt{NT}} \sum_{i=1}^N \sum_{t=1}^T x_{it} u_{it} \sigma_i^{-2} \left( \frac{1}{T} \sum_{s=1}^T (u_{is}^2 - Eu_{is}^2) \right) \sigma_i^{-2} = o_P(1). \tag{8}$$

Under the simplified conditions of this example ( $u_{it}$  is independent across both  $i$  and  $t$ ), it is straightforward to calculate  $\text{var}(A)$  and show that it converges to zero as  $N, T \rightarrow \infty$  regardless of whether  $N < T$  or not.

As for  $EA$ , straightforward calculations yield

$$EA = \frac{\sqrt{NT}}{T} \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T E(x_{it} E(u_{it}^3 | x_{it})) \sigma_i^{-4}.$$

Generally, if  $u_{it}|x_{it}$  is non-Gaussian and asymmetric,  $E(u_{it}^3|x_{it}) \neq 0$ . Hence, we require  $N/T \rightarrow 0$  to have  $EA \rightarrow 0$ . Hence, to allow for non-Gaussian and asymmetric conditional distributions, in the GLS setting it turns out  $N = o(T)$  is required.

We shall not explicitly impose  $N = o(T)$  in this paper as a formal assumption, but instead impose Assumption 2.4. On one hand, when the distribution of  $u_{it}$  is symmetric, we *do not* require  $N = o(T)$  because as is shown in the above example,  $E(u_{it}^3|x_{it}) = 0$  is sufficient for  $EA \rightarrow 0$  and is satisfied by symmetric distributions. On the other hand, when  $u_{it}$  is non-symmetric, Assumption 2.4 then implicitly requires  $N = o(T)$ . Note that  $N = o(T)$  is a strong assumption in many microeconomic applications for panel data models. But as illustrated in the above simple example, if  $u_{it}|x_{it}$  is not symmetric, it is required for feasible GLS even if  $\Omega$  is diagonal. One possible approach to weakening this assumption is to remove the higher order bias from  $\widehat{\Omega}$ . Higher order debiasing is a complicated procedure in the presence of general weak dependences. This is left for future research.

### 2.4 Asymptotic results of FGLS

We impose the following conditions, regulating the sparsity and serial weak dependence.

- Assumption 2.1** (i)  $\{u_t, x_t\}_{t \geq 1}$  is strictly stationary. In addition, each  $u_t$  has zero mean vector, and  $\{u_t\}_{t \geq 1}$  and  $\{x_t\}_{t \geq 1}$  are independent.
- (ii) There are constants  $c_1, c_2 > 0$  such that  $\lambda_{\min}(\Omega_h) > c_1$  and  $\|\Omega_h\|_1 < c_2$  for each fixed  $h$ .
- (iii) Exponential tail: There exist  $r_1, r_2 > 0$  and  $b_1, b_2 > 0$ , and for any  $s > 0, i \leq N$  and  $l \leq d$ ,

$$P(|u_{it}| > s) \leq \exp(-(s/b_1)^{r_1}), \quad P(|x_{it,l}| > s) \leq \exp(-(s/b_2)^{r_2}).$$

- (iv) Strong mixing: There exist  $\kappa \in (0, 1)$  such that  $r_1^{-1} + r_2^{-1} + \kappa^{-1} > 1$ , and  $C > 0$  such that for all  $T > 0$ ,

$$\sup_{A \in \mathcal{F}_{-\infty}^0, B \in \mathcal{F}_T^\infty} |P(A)P(B) - P(AB)| < \exp(-CT^\kappa),$$

where  $\mathcal{F}_{-\infty}^0$  and  $\mathcal{F}_T^\infty$  denote the  $\sigma$ -algebras generated by  $\{(x_t, u_t) : t \leq 0\}$  and  $\{(x_t, u_t) : t \geq T\}$ , respectively.

Condition (ii) requires that  $\Omega_h$  be well conditioned. Condition (iii) ensures the Bernstein-type inequality for weakly dependent data, which requires the underlying distributions to be thin-tailed. Condition (iv) is the standard  $\alpha$ -mixing condition, adapted to the large- $N$  panel. In addition, we impose the following regularity conditions.

- Assumption 2.2** (i) There exists a constant  $C > 0$  such that for all  $i \leq N$  and  $t \leq T$ ,  $E\|x_{it}\|^4 < C$  and  $Eu_{it}^4 < C$ .
- (ii) Define  $\xi_T(L) = \max_{t \leq T} \sum_{|h| > L} \|Eu_t u'_{t-h}\|$ . Then,  $\xi_T(L) \rightarrow 0$ .

(iii) Define  $f_T(L) = \max_{t \leq T} \sum_{|h| \leq L} \|Eu_t u'_{t-h}(1 - \omega(|h|, L))\|$ . Then  $f_T(L) \rightarrow 0$ .

Assumption 2.2 allows us to prove the convergence rate of the covariance matrix estimator. Condition (ii) is an extension of the standard weak serial dependence condition to the high-dimensional case in panel data literature. It allows us to employ banding or Newey–West truncation procedure. Condition (iii) is well satisfied by various kernel functions for the HAC-type estimator. For the Bartlett kernel, for example,

$$\max_{t \leq T} \sum_{|h| \leq L} \|Eu_t u'_{t-h}(1 - \omega(|h|, L))\| \leq \frac{1}{L} \max_{t \leq T} \sum_{|h|=0}^{\infty} \|Eu_t u'_{t-h}\| |h|$$

converges to zero as  $L \rightarrow \infty$  as long as  $\max_{t \leq T} \sum_{|h|=0}^{\infty} \|Eu_t u'_{t-h}\| |h| < \infty$ .

In this paper, we assume  $\Omega_h$  to be a sparse matrix for each  $h$  and impose similar conditions as those in Bickel and Levina (2008a) and Fan et al. (2013): write  $\Omega_h = (\Omega_{h,ij})_{N \times N}$ , where  $\Omega_{h,ij} = Eu_{it} u_{j,t-h}$ . For some  $q \in [0, 1)$ , we define

$$m_N = \max_{|h| \leq L} \max_{i \leq N} \sum_{j=1}^N |\Omega_{h,ij}|^q,$$

as a measurement of the sparsity. We would require that  $m_N$  should be either fixed or grow slowly as  $N \rightarrow \infty$ . In particular, when  $q = 0$ ,  $m_N = \max_{|h| \leq L} \max_{i \leq N} \sum_{j=1}^N 1(\Omega_{h,ij} \neq 0)$ , which corresponds to the exact sparsity case.

Let

$$\gamma_T = \sqrt{\log(LN)/T}.$$

**Assumption 2.3** For any  $NT \times NT$  matrix  $M$ , we denote  $(M)_{ts,ij}$  as the  $(i, j)$ th element of the  $(t, s)$ th block of the matrix  $M$ .

- (i)  $\sum_{|h| > L} \|\Omega_h\|_1 = O(L^{-\alpha})$ , for a constant  $\alpha > 0$ .
- (ii)  $\max_{i \leq N, t \leq T} \sum_{s=1}^T \sum_{j=1}^N |(\Omega^{-1})_{ts,ij}| = O(1)$ .
- (iii) There is  $q \in [0, 1)$  such that  $Lm_N \gamma_T^{1-q} = o(1)$  holds. In addition,

$$\sqrt{T} L^2 m_N^2 \gamma_T^{3-2q} = o(1), \quad \text{and} \quad \sqrt{NT} L^3 m_N^3 \gamma_T^{3-3q} = o(1).$$

- (iv)  $\sqrt{NT} (\xi_T(L) + f_T(L))^3 = o(1)$  and  $L^{-\alpha} T \sqrt{NT} m_N \gamma_T^{1-q} = o(1)$ .

Conditions (i)-(ii) require the weak cross-sectional correlations. Condition (iii) is about the sparsity assumptions on the growth of  $m_N$ , associated with  $q$  and the speed of  $L$ .

**Remark 2.2** To understand Assumption 2.3, consider a simple case where  $Eu_{it} u_{j,t-h}$  is nonzero for only finitely many pairs  $i \neq j$ . This corresponds to  $q = 0$  and  $m_N = O(1)$ . Then condition (iii) requires

$$\sqrt{N} L^3 \log^{3/2}(LT) = o(T).$$

In practice, the bandwidth  $L$  and  $\log(LT)$  both grow very slowly compared to  $N$  and  $T$ . So essentially this condition requires  $N = o(T^2)$ . In addition, condition (iv) assumes that the autocorrelations should decay sufficiently fast as  $L \rightarrow \infty$ . Suppose both  $\zeta_T(L)$  and  $f_T(L)$  decay in a polynomial rate of  $L$  (e.g., with order  $L^{-c_0}$ ), then this condition requires that the order of the polynomial,  $c_0$ , should be sufficiently large.

All the above conditions, we show in the appendix the convergence of  $\|\widehat{\Omega} - \Omega\|$ . It then leads to the following proposition.

**Proposition 2.1** *Under Assumption 2.1-2.2, for  $q \in [0, 1)$  and  $\alpha > 0$  such that Assumption 2.3 holds,*

$$\begin{aligned} \sqrt{NT}(\widehat{\beta}_{FGLS} - \beta) &= \Gamma^{-1} \left( \frac{1}{\sqrt{NT}} X' \Omega^{-1} U \right) \\ &\quad + \Gamma^{-1} \left( \frac{1}{\sqrt{NT}} X' \Omega^{-1} (\widehat{\Omega} - \Omega) \Omega^{-1} U \right) + o_P(1), \end{aligned}$$

where  $\Gamma = E(X' \Omega^{-1} X / NT)$ .

As we see from the above proposition, the effect  $\widehat{\Omega} - \Omega$  also appears as an “weighted average” in the second term on the right-hand-side of the expansion. The negligibility of this term relies on the following high-level condition. We define  $W = \Omega^{-1} X$  and  $\varepsilon = \Omega^{-1} U$ . Then,  $W = (w'_1, \dots, w'_T)'$  with  $w_t$  being an  $N \times d$  matrix of  $w_{it}$ , and  $\varepsilon_{it}$  is defined similarly.

**Assumption 2.4** Let  $A_{bh} = \{(i, j) : |Eu_{it}u_{j,t-h}| \neq 0\}$ . Then,

$$\left\| \frac{1}{\sqrt{NT}} \sum_{h=0}^L \sum_{i,j \in A_{bh}} \mathbb{G}_{T,ij}^1(h) \mathbb{G}_{T,ij}^2(h) \right\| = o_P(1), \tag{9}$$

where  $\mathbb{G}_{T,ij}^1(h) = \frac{1}{\sqrt{T}} \sum_{t=h+1}^T (u_{it}u_{j,t-h} - Eu_{it}u_{j,t-h})$  and  $\mathbb{G}_{T,ij}^2(h) = \frac{1}{\sqrt{T}} \sum_{t=h+1}^T w_{it}\varepsilon_{j,t-h}$ .

**Remark 2.3** While it is difficult to verify the above high-level condition in the presence of either serial dependence or cross-sectional dependence or both, the intuition can be understood in the simple i.i.d. case. Suppose  $u_{it}$  is independent across both  $i$  and  $t$ . Then, we can set  $L = 0$  and this condition becomes

$$A \equiv \frac{1}{\sqrt{NT}} \sum_{i=1}^N \frac{1}{T} \sum_{t=1}^T (u_{it}^2 - Eu_{it}^2) \sum_{s=1}^T x_{is}u_{is}\sigma_i^{-4} = o_P(1),$$

which is (8) as we discussed in Sect. 2.3. As discussed therein, it is straightforward to see that  $\text{var}(A) = o(1)$ , proving  $EA = o(1)$  requires either  $u_{it}$  has a symmetric distribution so that  $Eu_{it}^3 = 0$ , or  $N = o(T)$  for asymmetric distributions. Similar conditions were required for high-dimensional GLS problems, for instance, by Bai and Liao (2017) in panel data with interactive effect estimations.

Then, we have the following limiting distribution.

**Theorem 2.1** *Suppose  $\text{var}(U|X) = \text{var}(U) = \Omega$ . Under the Assumptions 2.1-2.4, for  $q \in [0, 1)$  and  $\alpha > 0$  such that Assumption 2.3 holds, as  $N, T \rightarrow \infty$ ,*

$$\sqrt{NT}(\widehat{\beta}_{FGLS} - \beta) \xrightarrow{d} \mathcal{N}(0, \Gamma^{-1}),$$

where  $\Gamma = \lim_{NT} E(X'\Omega^{-1}X/NT)$ , assumed to exist. The consistent estimator of  $\Gamma$  is  $\widehat{\Gamma} = X'\widehat{\Omega}^{-1}X/NT$ .

The asymptotic variance of the FGLS estimator is  $\text{Avar}(\widehat{\beta}_{FGLS}) = \Gamma^{-1}/NT$ , and an estimator of it is  $(X'\widehat{\Omega}^{-1}X)^{-1}$ . Asymptotic standard errors can be obtained in the usual fashion from the asymptotic variance estimates.

### 3 Empirical study: Effects of divorce law reforms on divorce rates

In the literature, the cause of the sharp increase in the US divorce rate in the 1960-1970s is an important research question. During 1970s, more than half of states in the US liberalized the divorce system, and the effects of reforms on divorce rates have been investigated by many such as Allen (1992) and Peters (1986). With controls for state and year fixed effects, Friedberg (1998) suggested that state law reforms significantly increased divorce rates. Also, she assumed that unilateral divorce laws affected divorce rates permanently. However, divorce rates from 1975 have been subsequently decreasing according to empirical evidence. Therefore, the question of whether law reforms also affect the divorce rate decrease has arisen. Wolfers (2006) revisited this question by using a treatment effect panel data model and identified only temporal effects of reforms on divorce rates. In particular, he used dummy variables for the first two years after the reforms, 3-4 years, 5-6 years, and so on. More specifically, the following fixed effect panel data model was considered:

$$y_{it} = \alpha_i + \mu_t + \sum_{k=1}^8 \beta_k X_{it,k} + \delta_i t + u_{it}, \quad (10)$$

where  $y_{it}$  is the divorce rate for state  $i$  and year  $t$ ,  $\alpha_i$  a state fixed effect,  $\mu_t$  a time fixed effects, and  $\delta_i t$  a linear time trend with unknown coefficient  $\delta_i$ .  $X_{it}$  is a binary regressor which denotes the treatment effect  $2k$  years after the reform. Wolfers (2006) suggested that “the divorce rate rose sharply following the adoption of unilateral divorce laws, but this rise was reversed within about a decade.” He also concluded that “15 years after reform the divorce rate is lower as a result of the adoption of unilateral divorce, although it is hard to draw any strong conclusions about long-run effects.

Both Friedberg (1998) and Wolfers (2006) used a weighted model by multiplying all variables by the square root of state population. In addition, they used ordinary OLS standard error, which does not take into account heteroskedasticity, serial and cross-sectional correlations. However, standard errors might be biased when one disregards these correlations. Therefore, we re-estimated the model of Wolfers (2006) using the

proposed FGLS method and OLS with the heteroskedastic standard errors of White (1980), the clustered standard error of Arellano (1987), and the robust standard error of Bai et al. (2019).

The same dataset as in Wolfers (2006) is used, which includes the divorce rate, state-level reform years, binary regressors, and state population. Due to missing observations around divorce law reforms, we exclude Indiana, New Mexico and Louisiana. As a result, we obtain balanced panel data from 1956 to 1988 for 48 states. We fit the models both with and without linear time trend and use OLS and FGLS in each model to estimate  $\beta$ .

The choice of the tuning parameters for implementing the FGLS follows the guidance provided in Table 1. Specifically, we set bandwidth  $L = 3$  as proposed by Newey and West (1994) ( $L = 4(T/100)^{2/9}$ ). The thresholding values are chosen by the cross-validation method as discussed in Sect. 2.2.2, more specifically,  $M = 1.8$  and  $M = 1.9$  for the model with and without linear time trends, respectively. The Bartlett kernel is used in the OLS robust standard error and FGLS estimation.

Model (10) is in fact more complicated than the model we formally studied in this paper, due to the inclusion of linear time trends and fixed effects. While theoretical studies of models with trends might be challenging in the high-dimensional GLS setting, it is straightforward to implement it in the same FGLS framework by applying a projection transformation to eliminate the time trend. Specifically, let  $\ell = (1, 2, \dots, T)'$  and  $P_\ell = I_T - \ell(\ell'\ell)^{-1}\ell'$ . We can define  $\tilde{Y}_i = P_\ell(y_{i1}, \dots, y_{iT})'$ , and  $\tilde{X}_i = P_\ell(X_{i1}, \dots, X_{iT})'$ , and define  $\tilde{y}_{it}$  and  $\tilde{X}_{it}$  accordingly from  $\tilde{y}_{it}$  and  $\tilde{X}_{it}$  that further remove the fixed effects.

The estimated  $\beta_1, \dots, \beta_8$  with and without linear time trend and standard errors are summarized in Table 2 below. The OLS and FGLS estimates in both models are similar to each other. The results show that divorce rates rose soon after the law reform. However, within a decade, divorce rates had fallen over time. Interestingly, FGLS confirms the negative effects of the law reforms on the divorce rates, specifically, 11-15+ years after the reform in the model with state-specific linear time trends, and 9-15+ years after the reform in the model without state-specific linear time trends. In addition, the FGLS estimates for 1-6 and 1-4 years are positive and statistically significant in the models with and without linear time trends, respectively. For OLS, the coefficient estimates for 3-4 and 7-15+ are significant in the model without linear time trends based on  $se_{BCL}$ . In contrast, the OLS estimates are statistically significant only for 1-4 years when a linear time trend is added. According to the clustered standard error,  $se_{CX}$ , note that only 11-15+ are statistically significant in the model without trends.

According to OLS and FGLS estimation results with and without a linear time trend, we make the following conclusion: in the first 8 years, the overall trend of divorce rate is increasing, but the law reform reduces the divorce rate after 3-4 years. However, 8 years after the reform, we observe that the law reform has a negative effect on divorce rate. Finally, we also note that there is a noticeable difference between the magnitudes of the OLS and FGLS estimates. While the difference of the magnitudes may be due to the small sample problem, the magnitude of these estimates are mostly within the 95% confidence interval of the other estimators. For instance,  $\hat{\beta}_{FGLS} = 0.133$  for the effect

**Table 2** Empirical application: effects of divorce law reform with state and year fixed effects: US state level data annual from 1956 to 1988, dependent variable is divorce rate per 1000 persons per year

Effects:	$\widehat{\beta}_{OLS}$	$se_W$	$se_{CX}$	$se_{BCL}$	$\widehat{\beta}_{FGLS}$	$se_{FGLS}$
<i>Panel A: Without state-specific linear time trends</i>						
1–2 years	0.256	0.140	0.189	0.148	0.133	0.046*
3–4 years	0.209	0.081*	0.159	0.089*	0.165	0.056*
5–6 years	0.126	0.073	0.168	0.069	0.100	0.059
7–8 years	0.105	0.070	0.165	0.040*	0.026	0.061
9–10 years	–0.122	0.060*	0.161	0.054*	–0.129	0.061*
11–12 years	–0.344	0.071*	0.173*	0.075*	–0.253	0.062*
13–14 years	–0.496	0.074*	0.188*	0.062*	–0.324	0.063*
15+years	–0.508	0.089*	0.223*	0.077*	–0.325	0.067*
<i>Panel B: With state-specific linear time trends</i>						
1–2 years	0.286	0.152	0.206	0.140*	0.171	0.044*
3–4 years	0.254	0.099*	0.171	0.126*	0.220	0.058*
5–6 years	0.186	0.102	0.206	0.143	0.175	0.067*
7–8 years	0.177	0.109	0.230	0.146	0.097	0.075
9–10 years	–0.037	0.111	0.241	0.154	–0.073	0.082
11–12 years	–0.247	0.128	0.268	0.183	–0.240	0.089*
13–14 years	–0.386	0.137*	0.295	0.209	–0.329	0.098*
15+years	–0.414	0.158*	0.337	0.243	–0.382	0.108*

*Note* Standard errors with asterisks indicate significance at 5% level using  $N(0, 1)$  critical values. For OLS standard errors,  $se_W$  and  $se_{CX}$  refer to the heteroskedastic standard errors by White (1980) and the clustered standard errors by Arellano (1987), respectively;  $se_{BCL}$  is the robust standard error suggested by Bai et al. (2019). The threshold values for FGLS by the cross-validation are  $M = 1.9$  and  $M = 1.8$  for Panel A and B, respectively. OLS and FGLS estimates and standard errors (using state population weights)

of 1–2 years, which is within the 95% confidence interval constructed using  $\widehat{\beta}_{OLS}$ ; the latter is  $[-0.0184, 0.5304]$ . For another example,  $\widehat{\beta}_{FGLS} = 0.165$  for the effect of 3–4 years, which is within the 95% confidence interval constructed using  $\widehat{\beta}_{OLS}$ ; the latter is  $[0.050, 0.367]$ . We note that these confidence intervals are relatively large, as a consequence of the relative small sample size in this study.

Overall, the results of FGLS estimates are consistent with Wolfers (2006). The FGLS confirms that the law reforms significantly contribute to the subsequent decrease in the divorce rates, more specifically, 9–15 years after the reform in the model without linear time trends, and 11–15 years after in the model with linear time trends. Though Wolfers (2006) de-emphasized the negative coefficient at the end of the periods, as these are not robust to inclusion of state-specific quadratic trends, which we did not employ in this paper, nevertheless, we still interpret the economic insight of these results as the consequence of two sides of the same treatment, the law reforms: after earlier dissolution of bad matches after law reforms, marital relations were gradually affected and changed.



## 4 Conclusions

This paper considers generalized least squares (GLS) estimation for linear panel data models. By estimating the large error covariance matrix consistently, the proposed feasible GLS estimator is more efficient than the ordinary least squares (OLS) in the presence of heteroskedasticity, serial and cross-sectional correlations. The covariance matrix used for the feasible GLS is estimated via the banding and thresholding method. We establish the limiting distribution of the proposed estimator. A Monte Carlo study is considered. The proposed method is applied to an empirical application.

## References

- Abadie A, Athey S, Imbens GW, Wooldridge J (2017) When should you adjust standard errors for clustering?. National Bureau of Economic Research Working Paper No 24003
- Allen DW (1992) Marriage and divorce: comment. *Am Econ Rev* 82(3):679–685
- Andrews DW (1991) Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econ J Econ Soc* 59(3):817–858
- Arellano M (1987) Computing Robust standard errors for within-groups estimators. *Oxford Bull Econ Stat* 49(4):431–434
- Bai J, Choi SH, Liao Y (2019) Standard Errors for Panel Data Models with Unknown Clusters. [arXiv:1910.07406](https://arxiv.org/abs/1910.07406)
- Bai J, Liao Y (2017) Inferences in panel data with interactive effects using large covariance matrices. *J Econ* 200(1):59–78
- Bai J, Ng S (2017) Principal components and regularized estimation of factor models. [arXiv:1708.08137](https://arxiv.org/abs/1708.08137)
- Bickel PJ, Levina E (2008a) Covariance regularization by thresholding. *Ann Stat* 36(6):2577–2604
- Bickel PJ, Levina E (2008b) Regularized estimation of large covariance matrices. *Ann Stat* 36(1):199–227
- Cameron AC, Miller DL (2015) A practitioner's guide to cluster-robust inference. *J Human Resour* 50(2):317–372
- Chernozhukov V, Chetverikov D, Demirer M, Duflo E, Hansen C, Newey W (2017) Double/debiased/neyman machine learning of treatment effects. *Am Econ Rev* 107(5):261–65
- Chernozhukov V, Chetverikov D, Demirer M, Duflo E, Hansen C, Newey WK (2016) Double machine learning for treatment and causal parameters. Discussion paper, cemmap working paper, Centre for Microdata Methods and Practice
- Driscoll JC, Kraay AC (1998) Consistent covariance matrix estimation with spatially dependent panel data. *Rev Econ Stat* 80(4):549–560
- Fan J, Liao Y, Mincheva M (2013) Large covariance estimation by thresholding principal orthogonal complements. *J R Stat Soc Ser B* 75(4):603–680
- Friedberg L (1998) Did unilateral divorce raise divorce rates? Evidence from panel data. *Am Econ Rev* 88(3):608–627
- Hansen CB (2007a) Asymptotic properties of a robust variance matrix estimator for panel data when T is large. *J Econ* 141(2):597–620
- Hansen CB (2007b) Generalized least squares inference in panel and multilevel models with serial correlation and fixed effects. *J Econ* 140(2):670–694
- Hansen LP (1982) Large sample properties of generalized method of moments estimators. *Econ J Econ Soc*, pp 1029–1054
- Liang K-Y, Zeger SL (1986) Longitudinal data analysis using generalized linear models. *Biometrika* 73(1):13–22
- Miller S, Startz R (2018) Feasible generalized least squares using machine learning. Available at SSRN 2966194
- Newey WK (1990) Efficient instrumental variables estimation of nonlinear models. *Econ J Econ Soc*, pp 809–837
- Newey WK, McFadden D (1994) Large sample estimation and hypothesis testing. *Handbook Econ* 4:2111–2245

- Newey WK, West KD (1987) A simple, positive semi-definite, heteroskedasticity and autocorrelation-consistent covariance matrix. *Econ J Econ Soc* 55:703–708
- Newey WK, West KD (1994) Automatic lag selection in covariance matrix estimation. *Rev Econ Stud* 61(4):631–653
- Peters HE (1986) Marriage and divorce: Informational constraints and private contracting. *Am Econ Rev* 76(3):437–454
- Petersen MA (2009) Estimating standard errors in finance panel data sets: Comparing approaches. *Rev Financ Stud* 22(1):435–480
- Romano JP, Wolf M (2017) Resurrecting weighted least squares. *J Econ* 197(1):1–19
- Vogelsang TJ (2012) Heteroskedasticity, autocorrelation, and spatial correlation robust inference in linear panel models with fixed-effects. *J Econ* 166(2):303–319
- Wager S, Athey S (2018) Estimation and inference of heterogeneous treatment effects using random forests. *J Am Stat Assoc* 113(523):1228–1242
- White H (1980) A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econ J Econ Soc*, pp 817–838
- Wolfers J (2006) Did unilateral divorce laws raise divorce rates? A reconciliation and new results. *Am Econ Rev* 96(5):1802–1820
- Wooldridge JM (2010) *Econometric analysis of cross section and panel data*. MIT press, Cambridge

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.