ARTICLE IN PRESS

Journal of Econometrics xxx (xxxx) xxx

Contents lists available at ScienceDirect

Journal of Econometrics

journal homepage: www.elsevier.com/locate/jeconom

Standard errors for panel data models with unknown clusters

Jushan Bai [a], Sung Hoon Choi [b,*], Yuan Liao [b]

[a] Columbia University, 420 West 118th St. MC 3308, New York, NY 10027, USA

[b] Rutgers University, 75 Hamilton St., New Brunswick, NJ 08901, USA

ARTICLE INFO

Article history:
Received 4 October 2019
Received in revised form 18 May 2020
Accepted 3 August 2020
Available online xxxx

Keywords:
Panel data
Clustered standard errors
Thresholding
Cross-sectional correlation
Serial correlation
Heteroskedasticity

ABSTRACT

This paper develops a new standard-error estimator for linear panel data models. The proposed estimator is robust to heteroskedasticity, serial correlation, and cross-sectional correlation of unknown forms. The serial correlation is controlled by the Newey–West method. To control for cross-sectional correlations, we propose to use the thresholding method, without assuming the clusters to be known. We establish the consistency of the proposed estimator. Monte Carlo simulations show the method works well. An empirical application is considered.

1. Introduction

Consider a linear panel regression with fixed-effects:

$$y_{it} = x_{it}'\beta + \alpha_i + \mu_t + u_{it},$$

where α_i and μ_t are individual fixed-effects and time fixed effects; x_{it} is a $k \times 1$ vector of explanatory variables; u_{it} is an unobservable error component. The outcome variable y_{it} and fixed effects are scalars, and β is a $k \times 1$ vector.

This paper is about the standard error of the fixed-effect ordinary least squares (OLS). One of the commonly used standard errors for OLS in empirical research is the (White, 1980) heteroskedasticity robust standard error in the cross-sectional setting. In the presence of serial and cross-sectional correlations, the conventional panel standard errors may be biased. Newey and West (1987) introduced heteroskedasticity and autocorrelation consistent (HAC) covariance matrix estimator for time series, which allows serial correlations (also see Andrews, 1991; Newey and West, 1994). The cluster standard errors suggested by Arellano (1987) are often reported in studies of the panel model. This estimator is robust to heteroskedasticity in the cross-section and also arbitrary serial correlation, but it focuses on the large-N small-T scenario. The case of large-N large-T is then studied by Ahn and Moon (2014) and Hansen (2007), among many others, while either cross-sectional or serial independence is required. Hansen (2007) examined the covariance estimator when the time series dependence is left unrestricted. In addition, Vogelsang (2012) studied the asymptotic theory that is robust to heteroskedasticity, autocorrelation, and spatial correlation, which extended and generalized the asymptotic results of Hansen (2007) for the conventional cluster standard errors including time fixed effects. Stock and Watson (2008) suggested a bias-adjusted heteroskedasticity-robust variance matrix estimator that handles serial correlations under any sequences of N or T. Also, see Petersen (2009) who used a simulation study to examine different types of

* Corresponding author.
E-mail addresses: jb3064@columbia.edu (J. Bai), shchoi@economics.rutgers.edu (S.H. Choi), yuan.liao@rutgers.edu (Y. Liao).

https://doi.org/10.1016/j.jeconom.2020.08.006
0304-4076/

Please cite this article as: J. Bai, S.H. Choi and Y. Liao, Standard errors for panel data models with unknown clusters. Journal of Econometrics (2020), https://doi.org/10.1016/j.jeconom.2020.08.006.

standard errors, including the clustered, Fama–MacBeth, and the modified version of Newey–West standard errors for panel data. In general, on the other hand, the conventional cluster standard errors assume that individuals across clusters are independent. Also, the cluster structure should be known such as schools, villages, industries, or states. See Arellano (2003), Cameron and Miller (2015), and Greene (2003). However, the knowledge of clusters is not available in many applications.

In a recent interesting paper, Abadie et al. (2017) argue that clustering is an issue more of sampling design or experimental design. Clustered standard errors are not always necessary and researchers should be more thoughtful when applying them. One reason is that clustering may result in an unnecessarily wider confidence interval. Clustered standard errors are derived from the modeling perspective (model implied variance matrix) and are widely practiced, see, for example, Angrist and Pischke (2008), Cameron and Trivedi (2005), and Wooldridge (2003, 2010). In this paper, we continue to take the modeling perspective. Because of our use of thresholding method, the resulting confidence interval is not necessarily much wider, even if all cross-sectional units are allowed to be correlated. Furthermore, the proposed approach is also applicable when the knowledge of clustering is not available.

We provide a robust standard error that allows both serial and cross-sectional correlations. We do not impose parametric structures on the serial or cross-sectional correlations. We assume these correlations are weak and apply nonparametric methods to estimate the standard errors. To control for the autocorrelation in time series, we employ the Newey–West truncation. To control for the cross-sectional correlation, we assume sparsity for cross-section (i, j) pairs, potentially resulting from the presence of cross-sectional clusters, but the knowledge on clustering (the number of clusters and the size of each cluster) is not assumed. We then estimate them by applying the thresholding approach of Bickel and Levina (2008). We also show how to make use of information on clustering when available. In passing we point out that instead of robust standard errors, in a separate study, Bai et al. (2019) proposed a feasible GLS (FGLS) method to take into account heteroskedasticity and both serial and cross-sectional correlations. The FGLS is more efficient than OLS.

The regularization methods such as banding and thresholding employed in this paper have been used extensively in the recent machine learning literature for estimating high-dimensional parameters. Nonparametric machine learning techniques have been proved to be useful tools in econometric studies.

The rest of the paper is organized as follows. In Section 2, we describe the models and standard errors as well as the asymptotic results of OLS. Monte Carlo studies evaluating the finite sample performance of the estimators are presented in Section 3. Section 4 illustrates our methods in an application of US divorce law reform effects. Conclusions are provided in Section 5 and all proofs are given in the Appendix.

Throughout this paper, $\nu_{min}(A)$ and $\nu_{max}(A)$ denote the minimum and maximum eigenvalues of matrix A. We use $\|A\| = \sqrt{\nu_{max}(A'A)}$, $\|A\|_1 = max_i \sum_j |A_{ij}|$ and $\|A\|_F = \sqrt{tr(A'A)}$ as the operator norm, the ℓ_1-norm and the Frobenius norm of a matrix A, respectively. Note that if A is a vector, $\|A\|$ is the Euclidean norm, and $|a|$ is the absolute-value norm of a scalar a.

2. OLS and standard error estimation

We consider the following model:

$$y_{it} = x_{it}'\beta + u_{it}, \tag{2.1}$$

where β is a $k \times 1$ vector of unknown coefficients, x_{it} is a $k \times 1$ vector of regressors, and u_{it} represents the error term, often known as the idiosyncratic component. This formulation incorporates the standard fixed effects models as in Hansen (2007). For example, x_{it}, y_{it} and u_{it} can be interpreted as variables resulting from removing the nuisance parameters from the equation, such as first-differencing to remove the fixed effects. Indeed, it is straightforward to allow additive fixed effects by using the usual demean procedure.

For a fixed t, model (2.1) can be written as:

$$y_t = x_t\beta + u_t, \tag{2.2}$$

where $y_t = (y_{1t}, \dots, y_{Nt})'$ $(N \times 1)$, $x_t = (x_{1t}, \dots, x_{Nt})'$ $(N \times k)$, and $u_t = (u_{1t}, \dots, u_{Nt})'$ $(N \times 1)$. To economize notation, we define $y_i = (y_{i1}, \dots, y_{iT})'$ $(T \times 1)$, $x_i = (x_{i1}, \dots, x_{iT})'$ $(T \times k)$, and $u_i = (u_{i1}, \dots, u_{iT})'$ $(T \times 1)$. So when the vector y is indexed by t, it refers to an $N \times 1$ vector, and when y is indexed by i it refers to a $T \times 1$ vector. Similar meaning is applied to x and u. There is no confusion when context is clear.

The (pooled) ordinary least square (OLS) estimator of β from Eqs. (2.1) and (2.2) may be defined as

$$\widehat{\beta} = (\sum_{i=1}^{N}\sum_{t=1}^{T} x_{it}x_{it}')^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T} x_{it}y_{it} = (\sum_{t=1}^{T} x_t'x_t)^{-1}\sum_{t=1}^{T} x_t'y_t. \tag{2.3}$$

The variance of $\widehat{\beta}$ depends on both $V_X \equiv \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} x_{it}x_{it}'$, and particularly,

$$V \equiv Var(\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T} x_{it}u_{it})$$
$$= \frac{1}{NT}\sum_{t=1}^{T} Ex_t'u_tu_t'x_t + \frac{1}{NT}\sum_{h=1}^{T-1}\sum_{t=h+1}^{T}[Ex_t'u_tu_{t-h}'x_{t-h} + Ex_{t-h}'u_{t-h}u_t'x_t]. \tag{2.4}$$

The goal of this paper is to consistently estimate V in the presence of both serial and cross-sectional correlations in $\{u_{it}\}$.

There are two types of clustered standard errors suggested by Arellano (1987). The original individual clustered version is

$$\widehat{V}_{CX} = \frac{1}{NT}\sum_{i=1}^{N} x_i'\widehat{u}_i\widehat{u}_i'x_i,$$

with $\widehat{u}_i = y_i - x_i\widehat{\beta}$ are the OLS residuals, and this estimator allows for arbitrary serial dependence and heteroskedasticity within individuals. In addition, $\widehat{V}_{CX}$ assumes no cross-section correlation.

The time-clustered version, which allows for heteroskedasticity and arbitrary cross-sectional correlation, is

$$\widehat{V}_{CT} = \frac{1}{NT}\sum_{t=1}^{T} x_t'\widehat{u}_t\widehat{u}_t'x_t,$$

with $\widehat{u}_t = y_t - x_t\widehat{\beta}$. Here $\widehat{V}_{CT}$ assumes no serial correlation.

The above clustered standard errors are robust to either arbitrary serial correlation or arbitrary cross-sectional correlation, respectively. In practice, however, since the dependence assumption is unknown, an over-rejection problem may occur. Specifically, if there exist both serial and cross-sectional correlations, these estimators are not robust anymore, as our numerical evidence shows in Section 3 (e.g., Tables 3 and 5).

To control for the serial correlation, a simple modification of $\widehat{V}_{CT}$ using (Newey and West, 1987) is

$$\widehat{V}_{DK} = \frac{1}{NT}\sum_{t=1}^{T} x_t'\widehat{u}_t\widehat{u}_t'x_t + \frac{1}{NT}\sum_{h=1}^{L}\omega(h,L)\sum_{t=h+1}^{T}[x_t'\widehat{u}_t\widehat{u}_{t-h}'x_{t-h} + x_{t-h}'\widehat{u}_{t-h}\widehat{u}_t'x_t], \tag{2.5}$$

where $\omega(\cdot)$ is the kernel function and L is the bandwidth. This estimator is suggested by Driscoll and Kraay (1998). When N is large, however, (2.5) accumulates a large number of cross-sectional estimation noises.

More generally, let

$$V_{ij} \equiv \frac{1}{T}\sum_{t=1}^{T} Ex_{it}u_{it}u_{jt}x_{jt}' + \frac{1}{T}\sum_{h=1}^{T-1}\sum_{t=h+1}^{T}[Ex_{it}u_{it}u_{j,t-h}x_{j,t-h}' + Ex_{i,t-h}u_{i,t-h}u_{jt}x_{jt}'].$$

Then equation (2.4) can be written as

$$V = \frac{1}{N}\sum_{ij} V_{ij}.$$

Unlike time series observations, cross-sectional observations have no natural ordering. They can be arranged in different orders. That is why cross-sectional correlation is more difficult to control. The usual cluster standard error makes the following assumption: let $C_1, \ldots, C_G$ be disjoint subsets of $\{1, \ldots, N\}$, so that they are *known* clusters and that $V_{ij} = 0$ when i and j belong to different clusters. So V can be expressed as

$$V = \frac{1}{N}\sum_{g=1}^{G}\sum_{(i,j)\in C_g} V_{ij}.$$

See Liang and Zeger (1986). Suppose the cardinality of each C_g is small (this would be the case if the number of clusters G is large) or grows slowly with N, then we only need to estimate $\sum_{g=1}^{G}\sum_{(i,j)\in C_g} 1$ number of V_{ij}'s, greatly reducing the number of pair-wise covariances. But as commented in the literature, this requires the knowledge of $C_1, \ldots, C_G$, which in some applications, is not naturally available.

2.1. The estimator of V with unknown clusters

The key assumption we make is that conditionally on x_t, $\{u_{it}\}$ is weakly correlated across both t and i. Essentially, this means V_{ij} is zero or nearly so for most pairs of (i,j). There is a partition $\{(i,j) : i, j \le N\} = S_s \bigcup S_l$ so that

$$S_s = \{(i,j) : \|Ex_{it}u_{it}u_{j,t+h}x_{j,t+h}'\| = 0\ \forall h\},$$
$$S_l = \{(i,j) : \|Ex_{it}u_{it}u_{j,t+h}x_{j,t+h}'\| \neq 0\ \exists h\},$$

where the subscript "s" indicates "small", and "l" indicates "large". We assume that $(i,i) \in S_l$ for all $i \leq N$, and importantly, most pairs (i,j) belong to S_s. Yet, we do not need to know which elements belong to S_s or S_l. Then

$$V = \frac{1}{N} \sum_{(i,j)\in S_l} V_{ij}.$$

Furthermore, let $\omega(h,L) = 1 - h/(L+1)$ be the Bartlett kernel. Also see Andrews (1991) for other kernel functions. As suggested by Newey and West (1987), V_{ij} can be approximated by

$$V_{u,ij} \equiv \frac{1}{T}\sum_{t=1}^{T} Ex_{it}u_{it}u_{jt}x_{jt}' + \frac{1}{T}\sum_{h=1}^{L}\omega(h,L)\sum_{t=h+1}^{T}[Ex_{it}u_{it}u_{j,t-h}x_{j,t-h}' + Ex_{i,t-h}u_{i,t-h}u_{jt}x_{jt}'].$$

Then approximately,

$$V \approx \frac{1}{N}\sum_{(i,j)\in S_l} V_{u,ij}.$$

The above approximation plays the fundamental role of our standard error estimator. We estimate V_{ij} using Newey and West (1987), and estimate S_l using the cross-sectional thresholding.

To apply Newey and West (1987), we estimate $V_{u,ij}$ by

$$S_{u,ij} \equiv \frac{1}{T}\sum_{t=1}^{T} x_{it}\widehat{u}_{it}\widehat{u}_{jt}x_{jt}' + \frac{1}{T}\sum_{h=1}^{L}\omega(h,L)\sum_{t=h+1}^{T}[x_{it}\widehat{u}_{it}\widehat{u}_{j,t-h}x_{j,t-h}' + x_{i,t-h}\widehat{u}_{i,t-h}\widehat{u}_{jt}x_{jt}'],$$

where $\widehat{u}_{it} = y_{it} - x_{it}'\widehat{\beta}$. For a predetermined threshold value λ_{ij}, we approximate S_l by

$$\widehat{S}_l = \{(i,j) : \|S_{u,ij}\| > \lambda_{ij}\}.$$

Hence, a "matrix hard-thresholding" estimator of V is

$$\widehat{V}_{\text{Hard}} \equiv \frac{1}{N}\sum_{(i,j)\in\widehat{S}_l\cup\{i=j\}} S_{u,ij}.$$

As for the threshold value, we specify

$$\lambda_{ij} = M\,\omega_{NT}\sqrt{\|S_{u,ii}\|\|S_{u,jj}\|},\ \text{ where } \omega_{NT} = L\sqrt{\frac{\log(LN)}{T}}$$

for a constant $M > 0$. The converging sequence $\omega_{NT} \to 0$ is chosen to satisfy:

$$\max_{i,j\leq N}\|S_{u,ij} - V_{u,ij}\| = O_P(\omega_{NT}).$$

In practice, the thresholding constant, M, can be chosen through multifold cross-validation, which is discussed in the next subsection. In addition, we can obtain $\widehat{V}_{DK}$ from $\widehat{V}_{\text{Hard}}$ by setting $M = 0$.

We also recommend a "matrix soft-thresholding" estimator as follows:

$$\widehat{V}_{\text{Soft}} \equiv \frac{1}{N}\sum_{i,j}\widehat{S}_{u,ij},$$

where $\widehat{S}_{u,ij}$ is

$$\widehat{S}_{u,ij} = \begin{cases} S_{u,ij}, & \text{if } i=j,\\ A_{u,ij}, & \text{if } \|S_{u,ij}\| > \lambda_{ij},\ \text{ and } i\neq j,\\ 0, & \text{if } \|S_{u,ij}\| < \lambda_{ij},\ \text{ and } i\neq j,\end{cases}$$

where the (k,k')'s element of $A_{u,ij}$ is ($sgn(x)$ denotes the sign function)

$$A_{u,ij,kk'} = \begin{cases} sgn(S_{u,ij,kk'})[|S_{u,ij,kk'}| - \eta_{ij,kk'}]_+, & \text{if } |S_{u,ij,kk'}| > \eta_{ij,kk'},\\ 0, & \text{if } |S_{u,ij,kk'}| < \eta_{ij,kk'},\end{cases}$$

for the threshold value

$$\eta_{ij,kk'} = M\,\omega_{NT}\sqrt{|S_{u,ii,kk'}||S_{u,jj,kk'}|},\ \text{ where } \omega_{NT} = L\sqrt{\frac{\log(LN)}{T}}$$

for some constant $M > 0$.

Remark 2.1. The thresholding estimators for V do not assume known cluster information (the number of clusters and the membership of clusters). The method can also be modified to take into account the clustering information when available, and is particularly suitable when the number of clusters is small, and the size of each cluster is large. The modification is to apply the thresholding method within each cluster. 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). The thresholding avoids this problem, while allowing correlations of unknown form within each cluster.

2.2. Choice of tuning parameters

Our suggested estimators, $\widehat{V}_{\text{Hard}}$ and $\widehat{V}_{\text{Soft}}$, require the choice of tuning parameters L and M, which are the bandwidth and the threshold constant respectively. To choose the bandwidth L, we recommend using $L = 4(T/100)^{2/9}$ as Newey and West (1994) suggested.

In practice, M can be chosen through multifold cross-validation. After obtaining the estimated residuals $\widehat{u}_{it}$ by OLS, we split the data into two subsets, denoted by $\{\widehat{u}_{it}\}_{t\in J_1}$ and $\{\widehat{u}_{it}\}_{t\in J_2}$; let $T(J_1)$ and $T(J_2)$ be the sizes of J_1 and J_2, which are $T(J_1) + T(J_2) = T$ and $T(J_1) \asymp T$. As suggested by Bickel and Levina (2008), we can set $T(J_1) = T(1 - \log(T)^{-1})$ and $T(J_2) = T/\log(T)$; J_1 represents the training data set, and J_2 represents the validation data set.

The procedure requires splitting the data multiple times, say P times. At the pth split, we denote by $\widehat{V}^p$ the sample covariance matrix based on the validation set, defined by

$$\widehat{V}^p = \frac{1}{N}\sum_{ij} S^p_{u,ij},$$

with $S^p_{u,ij}$ defined similarly to $S_{u,ij}$ using data on J_2. Let $\widehat{V}_s(M)$ be the thresholding estimator with threshold constant M using the entire sample. Then we choose the constant M^* by minimizing a cross-validation objective function

$$M^* = \arg\min_{0<M<M_0} \frac{1}{P}\sum_{p=1}^{P} \|\widehat{V}_s(M) - \widehat{V}^p\|_F^2, \quad s \in \{\text{Hard}, \text{Soft}\}$$

and the resulting estimator is $\widehat{V}_s(M^*)$. We use $L = 4(T/100)^{2/9}$ for both $\widehat{V}_s(M)$ and $\widehat{V}^p$ and find that setting $M_0 = 1$ works well. So the minimization is taken over $M \in (0, 1)$ through a grid search.

The above procedure modifies that of Bickel and Levina (2008) in two aspects. One is to use the entire sample when computing $\widehat{V}_s$ instead of J_1. Since $T(J_1)$ is close to T, this modification does not change the result much, but simplifies the computation. The second modification is to use a consecutive block for the validation set because of time series, so that the serial correlation is not perturbed. Hence in view of the time series nature, we first divide the data into $P = \log(T)$ blocks with block length $T/\log(T)$. Each J_2 is taken as one of the P blocks when computing $\widehat{V}^p$, similar to the K-fold cross-validation. We have conducted simulations of the cross-validation in the presence of both correlations, and the results show that this procedure performs well. For instance, the cross-validation tends to choose smaller M as the cross-sectional correlation becomes stronger. Due to the page limit, however, those are not reported in this paper.

2.3. Consistency

Below we present assumptions under which $\widehat{V}$ (either $\widehat{V}_{\text{Hard}}$ or $\widehat{V}_{\text{Soft}}$) consistently estimates V. We define

$$\alpha_{NT}(h) \equiv \sup_X \max_{t\le T}[\|E(u_t u'_{t-h}|X)\| + \|E(u_{t-h}u'_t|X)\|]$$

and

$$\rho_{ij,h} \equiv \sup_X \max_{t\le T}[|E(u_{it}u_{j,t-h}|X)|+|E(u_{i,t-h}u_{jt}|X)|],$$

where $X = \{x_{it}\}_{i\le N,t\le T}$. These coefficients give measures of autocovariances and cross-section covariances.

Assumption 2.1. (i) $E(u_t|x_t) = 0$.
(ii) Let $\nu_1 \le \cdots \le \nu_k$ be the eigenvalues of $(\frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T Ex_{it}x'_{it})$. Then there exist constants $c_1, c_2 > 0$ such that $c_1 < \nu_1 \le \cdots \le \nu_k < c_2$.

Assumption 2.2 (*Weak Serial and Cross-sectional Dependence*)**.** (i) $\sum_{h=0}^{\infty}\alpha_{NT}(h) \le C$ for some $C > 0$. In addition, there exist $\kappa \in (0, 1)$, $C > 0$ such that for all $T > 0$,

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

where $\mathcal{F}^0_{-\infty}$ and $\mathcal{F}^\infty_T$ denote the σ-algebras generated by $\{(x_t, u_t) : t \le 0\}$ and $\{(x_t, u_t) : t \ge T\}$ respectively.

(ii) For some $q \in [0, 1)$, $\omega_{NT}^{1-q}\max_{i\le N}\sum_{j=1}^N(\sum_{h=0}^L \rho_{ij,h})^q = o(1)$, where $\omega_{NT} \equiv L\sqrt{\frac{\log(LN)}{T}}$.

Assumption 2.2 (i) is the standard alpha-mixing condition, adapted to the large-N panel. Condition (ii) is new here. It requires weak cross-sectional correlations. It is similar to the "approximate sparse assumption" in Bickel and Levina (2008). Note that we actually allow the presence of many "small" but nonzero $\|Ex_{it}u_{it}u_{j,t+h}x'_{j,t+h}\|$. Clusters that have "large" $\|Ex_{it}u_{it}u_{j,t+h}x'_{j,t+h}\|$ are unknown to us. Hence the appealing feature of our method is that we allow for unknown clusters.

Essentially the assumption $\omega_{N,T}^{1-q}\max_{i\leq N}\sum_{j\leq N}(\sum_{h=0}^{L}\rho_{ij,h})^q = o(1)$ controls the order of elements in S_l. The following example presents a case of cross sectional weak correlations that satisfies condition (ii).

Example 2.1. Suppose uniformly for all $h = 0, \ldots, L$, $E(u_t u'_{t-h}|X)$ is an $N \times N$ block-diagonal matrix, where the size of each block is at most S_{NT}, which practically means that each cluster contains no more than S_{NT} individuals, assuming clusters are mutually uncorrelated. Then $\rho_{ij,h} = 0$ for (i,j) belong to different blocks. Within the same block, almost surely in X,

$$|E(u_{i,t}u_{j,t-h}|X)|+|E(u_{i,t-h}u_{jt}|X)| \leq \alpha_{NT}(h), \quad \sum_{h=0}^{\infty}\alpha_{NT}(h) < \infty$$

Then let $B(i)$ denote the block that i belongs to, whose size is at most S_{NT}.

$$\begin{aligned}\omega_{NT}^{1-q}\max_{i\leq N}\sum_{j=1}^{N}(\sum_{h=0}^{L}\rho_{ij,h})^q &= \omega_{NT}^{1-q}\max_{i\leq N}\sum_{j\in B(i)}(\sum_{h=0}^{L}\rho_{ij,h})^q \\ &\leq C\omega_{NT}^{1-q}S_{NT}(\sum_{h=0}^{\infty}\alpha_{NT}(h))^{q/c}\end{aligned}$$

for constants $c, C > 0$. The last term converges to zero so long as $\omega_{NT}^{1-q}S_{NT} \to 0$. This then requires either fixed or slowly growing cluster size S_{NT}.

Assumption 2.3. (i) For each fixed h, $\omega(h,L) \to 1$ as $L \to \infty$ and $\max_{h\leq L}|\omega(h,L)| \leq C$ for some $C > 0$.
(ii) Exponential tail: There exist $r_1, r_2 > 0$ and $b_1, b_2 > 0$, such that $r_1^{-1} + r_2^{-1} + \kappa^{-1} > 1$, and for any $s > 0, i \leq N$,

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

(iii) There is $c_1 > 0$, for all i, $\lambda_{\min}(\mathrm{var}(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}x_{it}u_{it})) > c_1$. Additionally, the eigenvalues of V and V_X are bounded away from both zero and infinity.

Condition (i) is well satisfied by various kernels for the HAC-type estimator. Condition (ii) ensures the Bernstein-type inequality for weakly dependent data. Note that it requires the underlying distributions to be thin-tailed. Allowing for heavy-tailed distributions is also an important issue. However, it would require a very different estimation method, and is out of the scope of this paper. Nevertheless, we have conducted simulation studies under heavy-tailed distributions (e.g., t-distribution with the degree of freedom 5). Indeed, the proposed estimator works well in this case, even though the theory requires thin-tailed distributions.[1]

We have the following main theorem and all proofs are contained in the Appendix.

Theorem 2.1. *Under Assumptions 2.1–2.3, as $N, T \to \infty$,*

$$\sqrt{NT}[V_X^{-1}\widehat{V}V_X^{-1}]^{-1/2}(\widehat{\beta} - \beta) \xrightarrow{d} \mathcal{N}(0, I).$$

Theorem 2.1 allows us to construct a $(1-\tau)\%$ confidence interval for $c'\beta$ for any given $c \in \mathbb{R}^k$. The standard error of $c'\hat{\beta}_{OLS}$ is

$$\left(\frac{1}{NT}c'(V_X^{-1}\widehat{V}V_X^{-1})c\right)^{1/2}$$

and the confidence interval for $c'\beta$ is $[c'\widehat{\beta} \pm Z_\tau\hat{\sigma}/\sqrt{NT}]$ where Z_τ is the $(1-\tau)\%$ quantile of standard normal distribution and $\hat{\sigma} = (c'(V_X^{-1}\widehat{V}V_X^{-1})c)^{1/2}$.

3. Monte Carlo experiments

3.1. DGP and methods

In this section we examine the finite sample performance of the robust standard errors using simulation study. The data generating process (DGP) used for the simulation is produced by the fixed effect linear regression model

$$y_{it} = \alpha_i + \mu_t + \beta_0 x_{it} + u_{it},$$

[1] The simulation results for the heavy-tailed distributions are available upon request from the authors.

where the true $\beta_0 = 1$. The DGP allows for serial and cross-sectional correlations in x_{it} as follows:

$$x_{it} = a_i \nu_{i+1,t} + \nu_{i,t} + b_i \nu_{i-1,t}, \quad \nu_{it} = \rho_X \nu_{i,t-1} + \epsilon_{it}, \quad \epsilon_{it} \sim N(0,1), \quad \nu_{i0} = 0,$$
$$\alpha_i \sim N(0, 0.5), \quad \mu_t \sim N(0, 0.5),$$

where the constants $\{a_i, b_i\}_{i=1}^N$ are i.i.d. Uniform$(0, \gamma_X)$, which introduce cross-sectional correlation. In addition, ν_{it} is modeled as $AR(1)$ process with the autoregressive parameter ρ_X. Throughout this simulation study, we set $\rho_X = 0.3$ and $\gamma_X = 1$.

We generate the error terms, u_{it}, in three different cases as follows:

$$\text{Case 1: } u_{it} = c_i m_{i+1,t} + m_{i,t} + d_i m_{i-1,t}, \quad m_{it} = \rho m_{i,t-1} + \varepsilon_{it}, \quad \varepsilon_{it} \sim N(0,1), \quad m_{i0} = 0,$$

$$\text{Case 2: } u_{it} = \psi \sum_{j=1}^{N} w_{ij} u_{it} + \eta_{it}, \quad \eta_{it} \sim N(0,1), \quad u_{i0} = 0,$$

$$\text{Case 3: } u_{it} = \sum_{k=1}^{r} \lambda_{ij} F_{tk} + e_{it}, \quad F_{tk} = \rho_F F_{t-1,k} + \xi_{tk}, \quad \lambda_{ik} = \rho_\lambda \lambda_{i-1,k} + \zeta_{tj},$$
$$e_{it} \sim N(0,1), \quad \xi_{it} \sim N(0,1), \quad \zeta_{it} \sim N(0,1).$$

The regressor is uncorrelated with the error term u_{it} each other. In Case 1, we generate the error term similar to x_{it}. The constants $\{c_i, d_i\}_{i=1}^N$ are i.i.d. Uniform$(0, \gamma)$, which introduce cross-sectional correlation, and heteroskedasticity when $\gamma > 0$. m_{it} is modeled as $AR(1)$ process with the autoregressive parameter ρ. Varying $\gamma > 0$ allows us to control for the strength of the cross-sectional correlation. Data are generated with four different structures of regressors and error terms: (a) no correlations ($\rho = 0, \gamma = 0$); (b) only serial correlation ($\rho = 0.5, \gamma = 0$); (c) only cross-sectional correlation ($\rho = 0, \gamma = 1$); and (d) both serial and cross-sectional correlations ($\rho = \{0.3, 0.9\}, \gamma = 1$). In Case 2, the error terms are modeled as a spatial autoregressive (SAR(1)) process. The matrix $W = (w_{ij})_{N\times N}$ is a rook type weight matrix whose diagonal elements are zero. Note that the rows of W are standardized, hence they sum to one. ψ is the scalar spatial autoregressive coefficient with $|\psi| < 1$. In this paper, we report the case of $\psi = 0.5$. Importantly, SAR(1) model does not produce the serial correlation on the error term. In Case 3, we consider an error factor structure. Both factors and factor loadings follow AR(1) processes, which introduce both serial and cross-sectional correlations. We set $r = 2$, and consider the cases of $\rho_\lambda = 0.3$ and $\rho_F = 0.9$.

In this simulation study, we examined t-statistics for testing the null hypothesis $H_0 : \beta_0 = 1$ against the alternative $H_1 : \beta_0 \neq 1$. In each simulation we compare the proposed estimator with that of other common five types of standard errors for $\widehat{\beta}$: the standard White estimator given by $\widehat{V}_{White} = \frac{1}{NT}\sum_{i=1}^N \sum_{t=1}^T \widetilde{x}_{it}\widetilde{x}_{it}'\widehat{u}_{it}^2$, where $\tilde{x}_{it}$ is demeaned version of regressor. Two types of clustered standard errors, $\widehat{V}_{CX}$ and $\widehat{V}_{CT}$, as defined in Section 2. In addition, we use two types of Newey and West HAC estimators for the panel version as follows:

$$\widehat{V}_{DK} = \frac{1}{NT}\sum_{t=1}^{T} \widetilde{x}_t'\widehat{u}_t\widehat{u}_t'\widetilde{x}_t + \frac{1}{NT}\sum_{h=1}^{L}\omega(h,L)\sum_{t=h+1}^{T}[\widetilde{x}_t'\widehat{u}_t\widehat{u}_{t-h}'\widetilde{x}_{t-h} + \widetilde{x}_{t-h}'\widehat{u}_{t-h}\widehat{u}_t'\widetilde{x}_t]$$

and

$$\widehat{V}_{HAC} = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\widetilde{x}_{it}\widetilde{x}_{it}'\widehat{u}_{it}^2 + \frac{1}{NT}\sum_{i=1}^{N}\sum_{h=1}^{L}\omega(h,L)\sum_{t=h+1}^{T}[\widetilde{x}_{it}\widehat{u}_{it}\widehat{u}_{i,t-h}\widetilde{x}_{i,t-h}' + \widetilde{x}_{i,t-h}\widehat{u}_{i,t-h}\widehat{u}_{it}\widetilde{x}_{it}'].$$

Note that $\widehat{V}_{HAC}$ assumes cross-sectional independence, while $\widehat{V}_{DK}$ allows arbitrary cross-sectional dependence. In addition, $\widehat{V}_{DK}$ and $\widehat{V}_{HAC}$ can be obtained from our proposed estimator with $M = 0$ and a large constant M, respectively.

Results are given for sample sizes $N = 50, 200$ and $T = 100, 200$. For each $\{N, T\}$ combination, we set $L = 3, 7, 11$ as the bandwidth for $\widehat{V}_{HAC}$, $\widehat{V}_{DK}$, and the proposed estimator, $\widehat{V}_{\text{Hard}}$. We also use Bartlett kernel for these three estimators. For the thresholding constant parameters of $\widehat{V}_{\text{Hard}}$, we set $M = 0.10, 0.15, 0.20, 0.25$ in all cases. The simulation is replicated for one thousand times for each case and the nominal significance level is 0.05. Simulation results are reported in Tables 1–5.

3.2. Results

Tables 1–5 present the simulation results, where each table corresponds to different cases. Each table reports null rejection probabilities for 5% level tests based on six different standard errors. As expected, a common feature in all tables is that when both N and T are small, all six estimators have rejection probabilities greater than 0.05. This might happen even when the errors are drawn from i.i.d. standard normal, and this problem becomes more noticeable in the presence of serial, cross-sectional, or both correlations. A number of interesting findings based on tables are summarized below.

Table 1
Null rejection probabilities, 5% level. Two-tailed test of H_0 : $\beta = 1$. Case 1: No cross-sectional correlation ($\gamma = 0$).

N	T	L\M	$\widehat{V}_{\text{Hard}}$				$\widehat{V}_{HAC}$	$\widehat{V}_{DK}$	$\widehat{V}_{CX}$	$\widehat{V}_{CT}$	$\widehat{V}_{W}$
			0.10	0.15	0.20	0.25					
			A. No serial correlation: $\rho = 0$								
50	100	3	.067	.065	.065	.067	.057	.068	.059	.058	.054
		7	.070	.066	.070	.062	.058	.073	.059	.058	.054
		11	.082	.071	.056	.055	.057	.088	.059	.058	.054
50	200	3	.054	.053	.053	.051	.046	.053	.057	.044	.047
		7	.055	.054	.056	.053	.047	.056	.057	.044	.047
		11	.056	.054	.051	.047	.047	.061	.057	.044	.047
200	100	3	.062	.065	.059	.060	.047	.065	.051	.055	.047
		7	.071	.066	.057	.050	.047	.075	.051	.055	.047
		11	.079	.065	.057	.047	.047	.091	.051	.055	.047
200	200	3	.051	.051	.053	.052	.048	.051	.051	.051	.048
		7	.057	.055	.055	.054	.047	.057	.051	.051	.048
		11	.058	.052	.049	.046	.047	.060	.051	.051	.048
			B. Serial correlation: $\rho = 0.5$								
50	100	3	.077	.078	.081	.078	.070	.078	.065	.104	.104
		7	.085	.084	.078	.076	.068	.083	.065	.104	.104
		11	.086	.082	.070	.066	.067	.091	.065	.104	.104
50	200	3	.070	.071	.075	.074	.069	.071	.067	.096	.100
		7	.073	.070	.068	.067	.063	.072	.067	.096	.100
		11	.077	.074	.065	.064	.061	.072	.067	.096	.100
200	100	3	.078	.080	.080	.077	.065	.080	.053	.103	.094
		7	.083	.078	.072	.061	.057	.082	.053	.103	.094
		11	.087	.071	.059	.058	.055	.105	.053	.103	.094
200	200	3	.057	.056	.055	.056	.052	.057	.045	.085	.082
		7	.059	.054	.053	.051	.048	.064	.045	.085	.082
		11	.064	.057	.054	.047	.047	.067	.045	.085	.082

Tables 1–3 show the results of Case 1. In Table 1, Panel A indicates that all the estimators perform well due to no correlation. Especially, White standard error estimators give rejection probabilities close to 0.05. In Panel B, when the serial correlation is introduced, the performances of $\widehat{V}_{CX}$ and $\widehat{V}_{HAC}$ are markedly better than others except for small sample size. In addition, our proposed estimator, $\widehat{V}_{\text{Hard}}$, also performs well when we use both larger threshold constant M and bandwidth L. Since there is only a serial correlation in the error term, these estimators take this correlation into account and perform well. As the size of bandwidth increases, the standard error estimated by $\widehat{V}_{HAC}$ increases to a level similar to the results of $\widehat{V}_{CX}$ and the tendency to over-reject diminishes. Since the Newey–West technique gives the weight, which is less than one, the estimated standard error may be underestimated. Hence, the traditional cluster standard error, $\widehat{V}_{CX}$, dominates the standard error of Newey–West panel version, $\widehat{V}_{HAC}$. Note that the unreported rejection probabilities of $\widehat{V}_{DK}$ exponentially increases as the bandwidth L increases.

Table 2 considers the case of cross-sectionally correlated errors and regressors. In Panel A, except the case of small sample size, $\widehat{V}_{CT}$ and $\widehat{V}_{DK}$ with small bandwidth L have rejection probabilities close to 0.05 in the first panel. Also, $\widehat{V}_{\text{Hard}}$ with small L and M performs well. Importantly, notice that the rejection rate of $\widehat{V}_{DK}$ and $\widehat{V}_{\text{Hard}}$ tend to over-reject substantially as the lag length L increases. In addition, as the cross-section size N increases, the over-rejection problem becomes worse, as we mentioned in Section 2. This tendency is easy to explain. Since $\widehat{V}_{DK}$ is an estimator based on a single time series and it is zero when full weight is given to the sample autocovariance, the bias in $\widehat{V}_{DK}$ initially falls but then increases as the lag length increases, while the variance of $\widehat{V}_{DK}$ is initially increasing but eventually becomes decreasing. Hence, $\widehat{V}_{DK}$ is biased downward substantially, and its t-statistics tends to over-reject when a large bandwidth is used. On the other hand, in the case of the small size of L and M, $\widehat{V}_{\text{Hard}}$ gives less bias on the estimated standard error.

Panel B of Table 2 allows the serial correlation as well as the cross-sectional correlation. Not surprisingly, all estimators except $\widehat{V}_{\text{Hard}}$ and $\widehat{V}_{DK}$ tend to over-reject substantially. In the small sample, these two estimators get worse than the case of the first panel. In the large sample, however, rejection probabilities of $\widehat{V}_{\text{Hard}}$ and $\widehat{V}_{DK}$ are close to 0.05. Importantly, $\widehat{V}_{\text{Hard}}$ outperforms $\widehat{V}_{DK}$ by choosing M properly. Unreported results of $\widehat{V}_{DK}$ with larger bandwidth, L, show much larger rejection probabilities than that of $\widehat{V}_{\text{Hard}}$. This indicates that we can obtain unbiased standard error estimator and appropriate rejection rates using our proposed estimators, $\widehat{V}_{\text{Hard}}$. Table 3 is the result of strong serial correlation with the cross-sectional dependence. When the serial correlation gets stronger, such as $\rho = 0.9$, all estimators tend to over-reject exponentially in small samples. However, $\widehat{V}_{\text{Hard}}$ and $\widehat{V}_{DK}$ outperform other estimators as the dimensionality increases.

Table 4 considers the error with SAR(1) structure, which does not require the serial correlation on the error term. Similar to the results reported in the first panel of Table 2, $\widehat{V}_{CT}$ gives rejection probabilities close to 0.05. $\widehat{V}_{DK}$ and $\widehat{V}_{\text{Hard}}$

Table 2
Null rejection probabilities, 5% level. Two-tailed test of $H_0 : \beta = 1$. Case 1: Cross-sectional correlation ($\gamma = 1$).

N	T	L\M	$\widehat{V}_{\text{Hard}}$				$\widehat{V}_{HAC}$	$\widehat{V}_{DK}$	$\widehat{V}_{CX}$	$\widehat{V}_{CT}$	$\widehat{V}_{W}$
			0.10	0.15	0.20	0.25					
			A. No serial correlation: $\rho = 0$								
50	100	3	.054	.054	.055	.055	.142	.055	.152	.054	.140
		7	.066	.064	.063	.069	.141	.068	.152	.054	.140
		11	.078	.077	.082	.109	.145	.079	.152	.054	.140
50	200	3	.046	.049	.046	.047	.133	.049	.145	.043	.132
		7	.054	.055	.060	.060	.134	.052	.145	.043	.132
		11	.059	.062	.063	.073	.135	.058	.145	.043	.132
200	100	3	.060	.060	.060	.064	.148	.058	.150	.053	.148
		7	.069	.073	.075	.080	.149	.067	.150	.053	.148
		11	.086	.085	.096	.126	.151	.084	.150	.053	.148
200	200	3	.050	.051	.051	.050	.121	.050	.128	.049	.121
		7	.057	.058	.057	.057	.122	.058	.128	.049	.121
		11	.063	.062	.064	.079	.123	.062	.128	.049	.121
			B. Serial correlation: $\rho = 0.3$								
50	100	3	.070	.069	.069	.067	.150	.069	.155	.074	.176
		7	.074	.075	.073	.078	.150	.077	.155	.074	.176
		11	.083	.079	.093	.108	.150	.085	.155	.074	.176
50	200	3	.058	.058	.058	.058	.150	.058	.146	.069	.171
		7	.055	.060	.062	.061	.142	.056	.146	.069	.171
		11	.059	.063	.071	.082	.142	.060	.146	.069	.171
200	100	3	.078	.076	.077	.072	.162	.080	.157	.091	.185
		7	.083	.087	.083	.084	.160	.086	.157	.091	.185
		11	.097	.089	.103	.133	.159	.101	.157	.091	.185
200	200	3	.055	.055	.054	.056	.132	.056	.133	.068	.157
		7	.053	.051	.056	.057	.130	.057	.133	.068	.157
		11	.057	.059	.065	.078	.130	.061	.133	.068	.157

Table 3
Null rejection probabilities, 5% level. Two-tailed test of $H_0 : \beta = 1$. Case 1: Both strong serial and cross-sectional correlations ($\rho = 0.9, \gamma = 1$).

N	T	L\M	$\widehat{V}_{\text{Hard}}$				$\widehat{V}_{HAC}$	$\widehat{V}_{DK}$	$\widehat{V}_{CX}$	$\widehat{V}_{CT}$	$\widehat{V}_{W}$
			0.10	0.15	0.20	0.25					
50	100	3	.096	.097	.098	.098	.180	.098	.168	.145	.271
		7	.101	.101	.100	.101	.173	.102	.168	.145	.271
		11	.113	.101	.110	.128	.174	.113	.168	.145	.271
50	200	3	.091	.092	.092	.092	.194	.092	.180	.157	.286
		7	.088	.087	.090	.093	.185	.087	.180	.157	.286
		11	.092	.092	.099	.114	.182	.092	.180	.157	.286
200	100	3	.089	.087	.087	.086	.193	.089	.146	.144	.256
		7	.092	.089	.095	.097	.178	.097	.146	.144	.256
		11	.100	.096	.109	.128	.173	.109	.146	.144	.256
200	200	3	.069	.069	.069	.067	.146	.068	.125	.121	.226
		7	.066	.068	.069	.067	.136	.069	.125	.121	.226
		11	.072	.071	.076	.087	.133	.072	.125	.121	.226

with small bandwidth L also perform well. Moreover, $\widehat{V}_{\text{Hard}}$ with proper thresholding constant M gives less bias than $\widehat{V}_{DK}$ on the estimated standard error.

Finally, Table 5 presents the results of the error factor structure. Similar to the results of Table 3, all estimators except $\widehat{V}_{\text{Hard}}$ and $\widehat{V}_{DK}$ tend to over-reject. Rejection probabilities of $\widehat{V}_{\text{Hard}}$ and $\widehat{V}_{DK}$ are relatively close to 0.05 when the sample size is large.

4. Empirical study: Effects of divorce law reforms

In this section, we re-examine the empirical work of the association between divorce law reforms and divorce rates using our proposed OLS standard error. There are many empirical studies on the effects of divorce law reforms on divorce rates. Friedberg (1998) found that state law reforms significantly increased divorce rates with controls for state and year fixed effects. Wolfers (2006) investigated the question of whether law reform continues to have an impact on the divorce

Table 4
Null rejection probabilities, 5% level. Two-tailed test of $H_0 : \beta = 1$. Case 2: Errors with Spatial AR(1) structure ($\psi = 0.5$).

N	T	L\M	$\widehat{V}_{\mathrm{Hard}}$				$\widehat{V}_{HAC}$	$\widehat{V}_{DK}$	$\widehat{V}_{CX}$	$\widehat{V}_{CT}$	$\widehat{V}_{W}$
			0.10	0.15	0.20	0.25					
50	100	3	.061	.060	.062	.057	.124	.059	.143	.053	.125
		7	.068	.068	.073	.077	.125	.067	.143	.053	.125
		11	.086	.083	.089	.110	.128	.085	.143	.053	.125
50	200	3	.046	.046	.047	.047	.113	.046	.130	.043	.111
		7	.052	.053	.053	.059	.114	.048	.130	.043	.111
		11	.052	.059	.065	.083	.112	.061	.130	.043	.111
200	100	3	.061	.057	.058	.057	.123	.062	.120	.051	.122
		7	.070	.071	.070	.081	.122	.068	.120	.051	.122
		11	.082	.080	.101	.117	.121	.088	.120	.051	.122
200	200	3	.055	.055	.053	.050	.124	.056	.125	.049	.123
		7	.064	.061	.062	.064	.123	.064	.125	.049	.123
		11	.069	.070	.083	.099	.123	.068	.125	.049	.123

Table 5
Null rejection probabilities, 5% level. Two-tailed test of $H_0 : \beta = 1$. Case 3: Errors with factor structure ($\rho_F = 0.9$, $\rho_\lambda = 0.3$).

N	T	L\M	$\widehat{V}_{\mathrm{Hard}}$				$\widehat{V}_{HAC}$	$\widehat{V}_{DK}$	$\widehat{V}_{CX}$	$\widehat{V}_{CT}$	$\widehat{V}_{W}$
			0.10	0.15	0.20	0.25					
50	100	3	.081	.080	.078	.081	.129	.078	.102	.130	.202
		7	.091	.084	.079	.082	.115	.093	.102	.130	.202
		11	.103	.086	.094	.086	.115	.108	.102	.130	.202
50	200	3	.083	.082	.081	.081	.117	.080	.095	.132	.183
		7	.066	.066	.065	.073	.107	.065	.095	.132	.183
		11	.072	.076	.080	.084	.104	.072	.095	.132	.183
200	100	3	.076	.074	.076	.074	.109	.076	.086	.121	.167
		7	.077	.080	.077	.074	.105	.083	.086	.121	.167
		11	.081	.076	.084	.094	.103	.096	.086	.121	.167
200	200	3	.072	.072	.073	.069	.115	.071	.090	.126	.184
		7	.070	.067	.071	.074	.106	.068	.090	.126	.184
		11	.074	.073	.073	.075	.104	.072	.090	.126	.184

rate by including dummy variables for the first two years after the reforms, 3–4 years, 5–6 years, and so on. Specifically, he studied the following fixed effect panel data model

$$y_{it} = \alpha_i + \mu_t + \sum_{k=1}^{8} \beta_k X_{it,k} + \delta_i t + u_{it}, \tag{4.1}$$

where y_{it} is the divorce rate for state i and year t; α_i and μ_t are the state and year fixed effects; $X_{it,k}$ is a binary regressor that representing the treatment effect $2k$ years after the reform; $\delta_i t$ a linear time trend. Wolfers (2006) suggested that there might be two sides of the same treatment yield this phenomenon: a number of divorces gradually shifted after the earlier dissolution of bad matches, after the reform.

Both Friedberg (1998) and Wolfers (2006) estimated OLS regressions using state population weight for each year. In addition, they estimated standard errors under the assumption that errors are homoskedastic, serially and cross-sectionally uncorrelated. However, ignoring these correlations might lead to bias in the standard error estimators. We re-estimated the model of Wolfers (2006) using proposed OLS standard error estimators.

The same data as in Wolfers (2006) are used, but we exclude Indiana, New Mexico, and Louisiana due to missing observations around divorce law reforms. As a result, we obtain a balanced panel data contain the divorce rates, state-level reform years, and binary regressors from 1956 to 1988 over 48 states. We fit models both with and without linear time trend, and also calculate our standard errors, as well as OLS, White, cluster, and HAC standard errors. We set lag choices $L = 3$ for HAC and our standard errors as suggested by Newey and West (1994) ($L = 4(T/100)^{2/9}$). The threshold values M chosen by the cross-validation method is $M = 0.2$ for the model without state-specific linear trends, and $M = 0.1$ with state-specific linear trends. These M values are relatively small, implying the existence of cross-sectional correlations. The estimated $\beta_1, \ldots, \beta_8$ with and without linear time trend and their different types of standard errors are presented respectively in Table 6. Note that robust standard errors are not necessarily larger than the usual OLS standard errors, as shown in columns corresponding to se_{CT}, se_{DK} and se_{Hard}.

In Table 6, OLS estimates with and without linear time trend are similar to each other. These estimates are also closely comparable to the results obtained in Wolfers (2006). The OLS estimates indicate that divorce rates rose soon

Table 6
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. OLS estimates and standard errors (using state population weights).

Effects:	$\hat{\beta}_{OLS}$	se_{OLS}	se_W	se_{CX}	se_{CT}	se_{HAC}	se_{DK}	se_{Hard}
Panel A: Without state-specific linear time trends								
1–2 years	.256	.086*	.140	.189	.139	.172	.155	.148
3–4 years	.209	.086*	.081*	.159	.075*	.114	.104*	.089*
5–6 years	.126	.086	.073	.168	.064*	.105	.088	.069
7–8 years	.105	.086	.070	.165	.059	.100	.065	.040*
9–10 years	−.122	.085	.060*	.161	.041*	.088	.058*	.054*
11–12 years	−.344	.085*	.071*	.173*	.043*	.101*	.056*	.075*
13–14 years	−.496	.085*	.074*	.188*	.050*	.110*	.054*	.062*
15+ years	−.508	.081*	.089*	.223*	.048*	.139*	.061*	.077*
Panel B: With state-specific linear time trends								
1–2 years	.286	.064*	.152	.206	.143*	.185	.145*	.140*
3–4 years	.254	.071*	.099*	.171	.102*	.140	.134	.126*
5–6 years	.186	.079*	.102	.206	.110	.145	.148	.143
7–8 years	.177	.086*	.109	.230	.120	.153	.155	.146
9–10 years	−.037	.093	.111	.241	.120	.156	.164	.154
11–12 years	−.247	.100*	.128	.268	.141	.179	.196	.183
13–14 years	−.386	.108*	.137*	.296	.164*	.193*	.218	.209
15+ years	−.414	.120*	.158*	.337	.186*	.221	.251	.243

Note: se_{OLS} and se_W refer to OLS and White standard errors respectively; se_{CX} and se_{CT} are clustered standard errors suggested by Arellano (1987); se_{HAC} and se_{DK} are two types of Newey–West HAC estimator as explained in the text; se_{Hard} is our standard error. Bartlett kernel with lag length $L = 3$ is used for se_{HAC}, se_{DK} and se_{Hard}. The threshold value for se_{Hard} by the cross-validation is $M = 0.2$ (for the first panel) and $M = 0.1$ (for the second panel).
*Standard errors indicate significance at 5% level using $N(0, 1)$ critical values.

after the law reform. However, within a decade, divorce rates had fallen over time. Most of the coefficient estimates are statistically significant at the 5% level using usual OLS standard errors. According to the cluster standard errors, however, the only significant estimates are 11–15+ after the reform in the model without linear time trend. We use our method of correcting standard error estimates for heteroskedasticity, serial correlation, and also cross-sectional correlation. In the model without linear trend, the estimates for 3–4 and 7–15+ are significant. On the other hand, the estimates for 1–4 are significant when linear trend is added. Our estimated standard errors are close to those of se_{CT} and se_{DK}, which allow arbitrary cross-section correlations. The result indicates non-negligible cross-sectional correlations. The result is also consistent with Kim and Oka (2014), who used the interactive fixed effects approach. The latter approach is suitable for models with strong cross-sectional correlations.

5. Conclusion

This paper studies the standard error problem for the OLS estimator in linear panel models, and proposes a new standard-error estimator that is robust to heteroskedasticity, serial and cross-sectional correlations when clusters are unknown. Simulated experiments demonstrate the robustness of the new standard-error estimator to various correlation structures.

Acknowledgments

We are grateful to the Editor and two anonymous referees for their insightful comments, which helped generate a substantially improved paper.

Appendix. Proofs

Throughout the proof, max_i, max_t, max_h, max_{ij}, max_{it}, $\sum_i$, $\sum_t$, and $\sum_{ij}$ denote $max_{i\leq N}$, $max_{t\leq T}$, $max_{h\leq L}$, $max_{i,j}$, $max_{i,t}$, $\sum_{i=1}^N$, $\sum_{t=1}^T$, and $\sum_{i=1}^N \sum_{j=1}^N$ respectively.

A.1. Technical lemmas

First let

$$V_L = \frac{1}{NT} \sum_t Ex_t' u_t u_t' x_t + \frac{1}{NT} \sum_{h=1}^{L} \omega(h, L) \sum_{t=h+1}^{T} [Ex_t' u_t u_{t-h}' x_{t-h} + Ex_{t-h}' u_{t-h} u_t' x_t].$$

We need following lemmas to prove the main results.

Lemma A.1. *(i)* $\|V - V_L\| \leq C\sum_{h=L}^{T-1}\alpha_{NT}(h) + C\sum_{h=1}^{L}(1-\omega(h,L))\alpha_{NT}(h)$.

(ii) $\max_i |V_{u,ii} - \mathrm{var}(\frac{1}{\sqrt{T}}\sum_{t=1}^{T} x_{it}u_{it})| = o(1)$.

(iii) $\min_i \lambda_{\min}(V_{u,ii}) > c$.

Proof. (i) First note that

$$\begin{aligned}\|Ex_t'u_tu_{t-h}'x_{t-h} + Ex_{t-h}'u_{t-h}u_t'x_t\| &\leq E\|x_t\|\|E(u_tu_{t-h}'|X)\|\|x_{t-h}\| + E\|x_{t-h}\|\|E(u_{t-h}u_t'|X)\|\|x_t\| \\ &\leq \alpha_{NT}(h)E\|x_t\|\|x_{t-h}\| \leq CN\alpha_{NT}(h).\end{aligned}$$

Hence for some $C, c > 0$,

$$\begin{aligned}\|V - V_L\| &\leq \|\frac{1}{NT}\sum_{h=L+1}^{T-1}\sum_{t=h+1}^{T}[Ex_t'u_tu_{t-h}'x_{t-h} + Ex_{t-h}'u_{t-h}u_t'x_t]\| \\ &\quad + \|\frac{1}{NT}\sum_{h=1}^{L}(1-\omega(h,L))\sum_{t=h+1}^{T}[Ex_t'u_tu_{t-h}'x_{t-h} + Ex_{t-h}'u_{t-h}u_t'x_t]\| \\ &\leq C\frac{1}{T}\sum_{h=L+1}^{T-1}\sum_{t=h+1}^{T}\alpha_{NT}(h^c) + C\frac{1}{T}\sum_{h=1}^{L}(1-\omega(h,L))\sum_{t=h+1}^{T}\alpha_{NT}(h^c) \\ &\leq C\sum_{h>L}\alpha_{NT}(h^c) + C\sum_{h=1}^{L}|1-\omega(h,L)|\alpha_{NT}(h^c) = o(1).\end{aligned}$$

The second term of the last equation goes to zero due to Assumption 2.2(iii) and the dominated convergence theorem, noting that $|1-\omega(h,L)|\alpha_{NT}(h^c) \leq C\alpha_{NT}(h^c)$ and $\alpha_{NT}(h^c)$ is summable over h.

(ii) The proof for $\max_i |V_{u,ii} - \mathrm{var}(\frac{1}{\sqrt{T}}\sum_{t=1}^{T} x_{it}u_{it})| = o(1)$ follows from the same argument.

(iii) The result follows from (ii) and the assumption that $\min_i \lambda_{\min}(\mathrm{var}(\frac{1}{\sqrt{T}}\sum_{t=1}^{T} x_{it}u_{it})) > c$. □

Lemma A.2. *Suppose* $\log N = o(T)$. *For* $f(t,h,L) = \omega(h,L)1\{t > h\}$,

$$\max_h \max_{i,j} \|\frac{1}{T}\sum_{t=1}^{T} x_{it}u_{it}u_{j,t-h}x_{j,t-h}'f(t,h,L) - Ex_{it}u_{it}u_{j,t-h}x_{j,t-h}'f(t,h,L)\| = O_P(\sqrt{\frac{\log(LN)}{T}}).$$

Proof. The left hand side can be written as $\max_h \max_{ij} \|\frac{1}{T}\sum_t Z_{h,ij,t}\|$, where $Z_{h,ij,t} = f(t,h,L)(x_{it}\varepsilon_{it}\varepsilon_{j,t-h}x_{j,t-h}' - Ex_{it}\varepsilon_{it}\varepsilon_{j,t-h}x_{j,t-h}')$. For convenience, assume that $\dim(Z_{h,ij,t}) = 1$ and there is no serial correlation. Set $\alpha_n = \sqrt{\frac{\log(LN)}{T}}$ and $c^2 = 2C$ for $c, C > 0$. Then, by using Bernstein inequality and exponential tail conditions (e.g., Merlevède et al., 2011), and that $f(t,h,L)$ is bounded,

$$P(\max_{h\leq L}\max_{ij}|\frac{1}{T}\sum_{t=1}^{T} Z_{h,ij,t}| > c\alpha_n) \leq LN^2 \max_{h\leq L}\max_{ij} P(|\frac{1}{T}\sum_{t=1}^{T} Z_{h,ij,t}| > c\alpha_n) \to 0.$$

Then $\max_h \max_{ij} \|\frac{1}{T}\sum_t Z_{h,ij,t}\| = O_P(\sqrt{\frac{\log(LN^2)\max_{i,j,h}\frac{1}{T}\sum_{t=1}^{T}\mathrm{var}(Z_{h,ij,t})}{T}}) = O_P(\sqrt{\frac{\log(LN)}{T}})$. □

Lemma A.3. *Suppose* $\log N = o(T)$. *For* $f(t,h,L) = \omega(h,L)1\{t > h\}$,

$$\max_h \max_{i,j} \|\frac{1}{T}\sum_{t=1}^{T} x_{it}\widehat{u}_{it}\widehat{u}_{j,t-h}x_{j,t-h}'f(t,h,L) - x_{it}u_{it}u_{j,t-h}x_{j,t-h}'f(t,h,L)\| = O_P(\frac{1}{T}\sqrt{\frac{\log(LN)}{N}}).$$

Proof. The left hand side is bounded by $a_1 + a_2 + a_3$, where

$$\begin{aligned}a_1 &= \max_h \max_{i,j} \|\frac{1}{T}\sum_{t=1}^{T} x_{it}(\widehat{u}_{it} - u_{it})(\widehat{u}_{j,t-h} - u_{j,t-h})x_{j,t-h}'f(t,h,L)\| \\ a_2 &= \max_h \max_{i,j} \|\frac{1}{T}\sum_{t=1}^{T} x_{it}u_{it}(\widehat{u}_{j,t-h} - u_{j,t-h})x_{j,t-h}'f(t,h,L)\| \\ a_3 &= \max_h \max_{i,j} \|\frac{1}{T}\sum_{t=1}^{T} x_{it}(\widehat{u}_{it} - u_{it})u_{j,t-h}x_{j,t-h}'f(t,h,L)\|.\end{aligned}$$

For simplicity, let us assume $\dim(x_{it}) = 1$. Then

$$a_1 \leq \|\widehat{\beta}-\beta\|^2 \max_h \max_{i,j} \|\frac{1}{T}\sum_{t=1}^{T} x_{it}x_{it}x_{j,t-h}x_{j,t-h}f(t,h,L)\|$$
$$\leq O_P(\frac{1}{NT})\max_i \frac{1}{T}\sum_{t=1}^{T}\|x_{it}\|^4 = O_P(\frac{1}{NT}).$$

By using Bernstein's inequality in Merlevède et al. (2011) for weakly dependent data and exponential tail conditions, and that $f(t,h,L)$ is bounded,

$$a_2 \leq \|\widehat{\beta}-\beta\| \max_h \max_{i,j} \|\frac{1}{T}\sum_{t=1}^{T} x_{it}u_{it}x_{j,t-h}x_{j,t-h}f(t,h,L)\|$$
$$\leq O_P(\frac{1}{\sqrt{NT}})O_P(\sqrt{\frac{\log(LN)}{T}}) = O_P(\frac{1}{T}\sqrt{\frac{\log(LN)}{N}}).$$

a_3 is bounded using the same argument. Together,

$$\max_h \max_{i,j} \|\frac{1}{T}\sum_{t=1}^{T} x_{it}\widehat{u}_{it}\widehat{u}_{j,t-h}x'_{j,t-h}f(t,h,L) - x_{it}u_{it}u_{j,t-h}x'_{j,t-h}f(t,h,L)\| = O_P(\frac{1}{T}\sqrt{\frac{\log(LN)}{N}}). \quad \square$$

A.2. Proof of Theorem 2.1 .

It suffices to prove $\|\widehat{V}-V\| = o_P(1)$. By Lemma A.1, we have

$$\|\widehat{V}-V\| \leq \|\widehat{V}-V_L\| + C\sum_{h>L}\alpha_{NT}(h) + C\sum_{h=1}^{L}(1-\omega(h,L))\alpha_{NT}(h).$$

The remaining proof is that of $\|\widehat{V}-V_L\| = o_P(1)$, given below.

Main proof of the convergence of $\|\widehat{V}-V_L\|$

Note that $V_L = \frac{1}{N}\sum_{ij}V_{u,ij}$, $\widehat{V} = \frac{1}{N}\sum_{ij}\widehat{S}_{u,ij}$. Hence

$$\|\widehat{V}-V_L\| \leq \frac{1}{N}\sum_{\widehat{S}_{u,ij}=0}\|V_{u,ij}-\widehat{S}_{u,ij}\| + \frac{1}{N}\sum_{\widehat{S}_{u,ij}\neq 0}\|V_{u,ij}-\widehat{S}_{u,ij}\|.$$

Note that $\|S_{u,ij}-V_{u,ij}\| < \frac{1}{2}\lambda_{ij}$ for $\forall(i,j)$ and $C_1 > 0$

$$\begin{aligned}\|S_{u,ii}\| &\geq \|V_{u,ii}\| - \|S_{u,ii}-V_{u,ii}\| \\ &\geq \|V_{u,ii}\| - \max_{ij}\|S_{u,ii}-V_{u,ii}\| \\ &\geq \|V_{u,ii}\| - C\omega_{NT} > C_1.\end{aligned}$$

From Assumption 2.3, $\|V_{u,ii}\| > c_1 > 0$, then, $\lambda_{ij} = M\omega_{NT}\sqrt{\|S_{u,ii}\|\|S_{u,jj}\|} > c_1 M\omega_{NT} > 2c_1\omega_{NT}$. Then, $\frac{\lambda_{ij}}{2} > c\omega_{NT} \geq \max_{ij}\|S_{u,ij}-V_{u,ij}\|$. Therefore, $\|S_{u,ij}-V_{u,ij}\| < \frac{1}{2}\lambda_{ij}$ for $\forall(i,j)$ Recall $\rho_{ij,h} = \sup_X \max_t |E(u_{it}u_{j,t-h}|X)| + |E(u_{i,t-h}u_{jt}|X)|$. Then,

$$\|V_{u,ij}\| \leq \|\frac{1}{T}\sum_t Ex_{it}u_{it}u_{jt}x'_{jt} + \frac{1}{T}\sum_{h=1}^{L}\omega(h,L)\sum_{t=h+1}^{T}[Ex_{it}u_{it}u_{j,t-h}x'_{j,t-h} + Ex_{i,t-h}u_{i,t-h}u_{jt}x'_{jt}]\|$$
$$\leq C\rho_{ij,0}/2 + C\frac{1}{T}\sum_{h=1}^{L}\omega(h,L)\sum_{t=h+1}^{T}\rho_{ij,h} \leq C\sum_{h=0}^{L}\rho_{ij,h}.$$

Hence, on the event $\max_{ij}\|S_{u,ij}-V_{u,ij}\| \leq C\omega_{NT}$,

$$\begin{aligned}\frac{1}{N}\sum_{\widehat{S}_{u,ij}=0}\|V_{u,ij}-\widehat{S}_{u,ij}\| &\leq \frac{1}{N}\sum_{\widehat{S}_{u,ij}=0}\|V_{u,ij}\| \leq \frac{1}{N}\sum_{ij}\|V_{u,ij}\|1\{\|S_{u,ij}\| < \lambda_{ij}\} \\ &= \frac{1}{N}\sum_{ij}\|V_{u,ij}\|1\{\|V_{u,ij}\| < \|S_{u,ij}\| + \|S_{u,ij}-V_{u,ij}\|, \|S_{u,ij}\| < \lambda_{ij}\} \\ &\leq \frac{1}{N}\sum_{ij}\|V_{u,ij}\|\frac{(1.5\lambda_{ij})^{1-q}}{\|V_{u,ij}\|^{1-q}}1\{\|V_{u,ij}\| < 1.5\lambda_{ij}\}\end{aligned}$$

$$\leq \frac{1}{N}\sum_{ij}\|V_{u,ij}\|^q(1.5\lambda_{ij})^{1-q} \leq C\omega_{NT}^{1-q}\frac{1}{N}\sum_{ij}\|V_{u,ij}\|^q$$
$$\leq C\omega_{NT}^{1-q}\max_i\sum_j(\sum_{h=0}^{L}\rho_{ij,h})^q.$$

On the other hand, on the event $\max_{ij}\|S_{u,ij}-V_{u,ij}\|\leq C\omega_{NT}$,

$$\frac{1}{N}\sum_{\widehat{S}_{u,ij}\neq 0}\|V_{u,ij}-\widehat{S}_{u,ij}\| \leq \frac{1}{N}\sum_{\widehat{S}_{u,ij}\neq 0}\|V_{u,ij}-S_{u,ij}\| + \frac{1}{N}\sum_{\widehat{S}_{u,ij}\neq 0}\|S_{u,ij}-\widehat{S}_{u,ij}\|$$
$$\leq \frac{1}{N}\sum_{\widehat{S}_{u,ij}\neq 0}0.5\lambda_{ij} + \frac{1}{N}\sum_{\widehat{S}_{u,ij}\neq 0}\lambda_{ij} \leq \frac{1}{N}\sum_{ij}1.5\lambda_{ij}1\{\|S_{u,ij}\|>\lambda_{ij}\}$$
$$= \frac{1}{N}\sum_{ij}1.5\lambda_{ij}1\{\|V_{u,ij}\|>\|S_{u,ij}\|-\|S_{u,ij}-V_{u,ij}\|, \|S_{u,ij}\|>\lambda_{ij}\}$$
$$\leq \frac{1}{N}\sum_{ij}1.5\lambda_{ij}\frac{\|V_{u,ij}\|^q}{(0.5\lambda_{ij})^q}1\{\|V_{u,ij}\|>0.5\lambda_{ij}\}$$
$$\leq \frac{1}{N}\sum_{ij}C\lambda_{ij}^{1-q}\|V_{u,ij}\|^q \leq \frac{1}{N}\sum_{ij}\|V_{u,ij}\|^q C\omega_{NT}^{1-q}$$
$$\leq C\omega_{NT}^{1-q}\max_i\sum_j(\sum_{h=0}^{L}\rho_{ij,h})^q.$$

Hence $\|\widehat{V}-V_L\|\leq C\omega_{NT}^{1-q}\max_i\sum_j(\sum_{h=0}^{L}\rho_{ij,h})^q$. Therefore, we have

$$\|\widehat{V}-V\| \leq O_P(\omega_{NT}^{1-q}\max_i\sum_j(\sum_{h=0}^{L}\rho_{ij,h})^q) + C\sum_{h=L}^{T-1}\alpha_{NT}(h) + C\sum_{h=1}^{L}(1-\omega(h,L))\alpha_{NT}(h).$$

Remaining proofs: $\max_{ij}\|S_{u,ij}-V_{u,ij}\| = O_P(\omega_{NT})$, where $\omega_{NT} = L\sqrt{\frac{\log(LN)}{T}}$] Recall

$$S_{u,ij} \equiv \frac{1}{T}\sum_t x_{it}\widehat{u}_{it}\widehat{u}_{jt}x_{jt}' + \frac{1}{T}\sum_{h=1}^{L}\omega(h,L)\sum_{t=h+1}^{T}[x_{it}\widehat{u}_{it}\widehat{u}_{j,t-h}x_{j,t-h}' + x_{i,t-h}\widehat{u}_{i,t-h}\widehat{u}_{jt}x_{jt}'],$$
$$V_{u,ij} \equiv \frac{1}{T}\sum_t Ex_{it}u_{it}u_{jt}x_{jt}' + \frac{1}{T}\sum_{h=1}^{L}\omega(h,L)\sum_{t=h+1}^{T}[Ex_{it}u_{it}u_{j,t-h}x_{j,t-h}' + Ex_{i,t-h}u_{i,t-h}u_{jt}x_{jt}'].$$

Let

$$M_{u,ij} \equiv \frac{1}{T}\sum_t x_{it}u_{it}u_{jt}x_{jt}' + \frac{1}{T}\sum_{h=1}^{L}\omega(h,L)\sum_{t=h+1}^{T}[x_{it}u_{it}u_{j,t-h}x_{j,t-h}' + x_{i,t-h}u_{i,t-h}u_{jt}x_{jt}'].$$

We first bound $\max_{ij}\|M_{u,ij}-V_{u,ij}\|$, then bound $\max_{ij}\|S_{u,ij}-M_{u,ij}\|$.

Proof of $\max_{ij}\|M_{u,ij}-V_{u,ij}\| = O_P(L\sqrt{\frac{\log(LN)}{T}})$

Given Lemma A.2, we have

$$\max_{ij}\|M_{u,ij}-V_{u,ij}\| \leq O_P(\sqrt{\frac{\log(LN)}{T}})$$
$$+ 2\max_{ij}\|\frac{1}{T}\sum_{h=1}^{L}\sum_{t=1}^{T}[x_{it}u_{it}u_{j,t-h}x_{j,t-h}'f(t,h,L) - Ex_{it}u_{it}u_{j,t-h}x_{j,t-h}'f(t,h,L)]\|$$
$$\leq O_P(\sqrt{\frac{\log(LN)}{T}}) + LO_P(\sqrt{\frac{\log(LN)}{T}}) = O_P(L\sqrt{\frac{\log(LN)}{T}}).$$

Prove of $\max_{ij}\|M_{u,ij}-S_{u,ij}\| = O_P(\frac{L}{T}\sqrt{\frac{\log(LN)}{N}})$

Given Lemma A.3, we have

$$\begin{aligned}\max_{ij} \|M_{u,ij} - S_{u,ij}\| &\leq O_P(\frac{1}{T}\sqrt{\frac{\log(LN)}{N}}) \\ &+ 2\max_{ij} \|\frac{1}{T}\sum_{h=1}^{L}\sum_{t=1}^{T}[x_{it}\widehat{u}_{it}\widehat{u}_{j,t-h}x'_{j,t-h}f(t,h,L) - x_{it}u_{it}u_{j,t-h}x'_{j,t-h}f(t,h,L)]\| \\ &\leq O_P(\frac{1}{T}\sqrt{\frac{\log(LN)}{N}}) + LO_P(\frac{1}{T}\sqrt{\frac{\log(LN)}{N}}) = O_P(\frac{L}{T}\sqrt{\frac{\log(LN)}{N}}).\end{aligned}$$

Together,

$$\max_{ij} \|V_{u,ij} - S_{u,ij}\| = O_P(L\sqrt{\frac{\log(LN)}{T}}) + O_P(\frac{L}{T}\sqrt{\frac{\log(LN)}{N}}) = O_P(L\sqrt{\frac{\log(LN)}{T}}). \quad \square$$

References

Abadie, A., Athey, S., Imbens, G.W., Wooldridge, J., 2017. When Should You Adjust Standard Errors for Clustering? National Bureau of Economic Research Working Paper No. 24003.

Ahn, S.C., Moon, H.R., 2014. Large-N and large-T properties of panel data estimators and the hausman test. In: Festschrift in Honor of Peter Schmidt. Springer, pp. 219–258.

Andrews, D.W., 1991. Heteroskedasticity and autocorrelation consistent covariance matrix estimation. Econometrica 59 (3), 817–858.

Angrist, J.D., Pischke, J.-S., 2008. Mostly Harmless Econometrics: An Empiricist's Companion. Princeton university press.

Arellano, M., 1987. Computing robust standard errors for within-groups estimators. Oxford Bull. Econ. Stat. 49 (4), 431–434.

Arellano, M., 2003. Panel Data Econometrics. Oxford university press.

Bai, J., Choi, S.H., Liao, Y., 2019. Feasible Generalized Least Squares for Panel Data Models with Cross-sectional and Serial Correlations. Working paper.

Bickel, P.J., Levina, E., 2008. Covariance regularization by thresholding. Ann. Statist. 36 (6), 2577–2604.

Cameron, A.C., Miller, D.L., 2015. A practitioner's guide to cluster-robust inference. J. Hum. Resour. 50 (2), 317–372.

Cameron, A.C., Trivedi, P.K., 2005. Microeconometrics: Methods and Applications. Cambridge university press.

Driscoll, J.C., Kraay, A.C., 1998. Consistent covariance matrix estimation with spatially dependent panel data. Rev. Econ. Stat. 80 (4), 549–560.

Friedberg, L., 1998. Did unilateral divorce raise divorce rates? Evidence from panel data. Amer. Econ. Rev. 88 (3), 608–627.

Greene, W.H., 2003. Econometric analysis. Pearson Education, Upper Saddle River, NJ.

Hansen, C.B., 2007. Asymptotic properties of a robust variance matrix estimator for panel data when T is large. J. Econometrics 141 (2), 597–620.

Kim, D., Oka, T., 2014. Divorce law reforms and divorce rates in the USA: An interactive fixed-effects approach. J. Appl. Econometrics 29 (2), 231–245.

Liang, K.-Y., Zeger, S.L., 1986. Longitudinal data analysis using generalized linear models. Biometrika 73 (1), 13–22.

Merlevède, F., Peligrad, M., Rio, E., 2011. A bernstein type inequality and moderate deviations for weakly dependent sequences. Probab. Theory Related Fields 151 (3–4), 435–474.

Newey, W.K., West, K.D., 1987. A simple, positive semi-definite, heteroskedasticity and autocorrelationconsistent covariance matrix. Econometrica 55, 703–708.

Newey, W.K., West, K.D., 1994. Automatic lag selection in covariance matrix estimation. Rev. Econom. Stud. 61 (4), 631–653.

Petersen, M.A., 2009. Estimating standard errors in finance panel data sets: Comparing approaches. Rev. Financ. Stud. 22 (1), 435–480.

Stock, J.H., Watson, M.W., 2008. Heteroskedasticity-robust standard errors for fixed effects panel data regression. Econometrica 76 (1), 155–174.

Vogelsang, T.J., 2012. Heteroskedasticity, autocorrelation, and spatial correlation robust inference in linear panel models with fixed-effects. J. Econometrics 166 (2), 303–319.

White, H., 1980. A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. Econometrica 817–838.

Wolfers, J., 2006. Did unilateral divorce laws raise divorce rates? A reconciliation and new results. Amer. Econ. Rev. 96 (5), 1802–1820.

Wooldridge, J.M., 2003. Cluster-sample methods in applied econometrics. Amer. Econ. Rev. 93 (2), 133–138.

Wooldridge, J.M., 2010. Econometric Analysis of Cross Section and Panel Data. MIT press.