Standard Errors for Panel Data Models with Unknown Clusters

Jushan Bai∗  Sung Hoon Choi†  Yuan Liao‡
Columbia University  Rutgers University  Rutgers University
October 17, 2019

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 form. Serial correlation is controlled by the Newey-West method. To control 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.

Keywords: Panel data, clustered standard errors, thresholding, cross-sectional correlation, serial correlation, heteroskedasticity

∗Address: 420 West 118th St. MC 3308, New York, NY 10027, USA. E-mail: jb3064@columbia.edu.
†Address: 75 Hamilton St., New Brunswick, NJ 08901, USA. E-mail: sc1711@economics.rutgers.edu.
‡Address: 75 Hamilton St., New Brunswick, NJ 08901, USA. Email: yuan.liao@rutgers.edu.
1 Introduction

Consider a linear panel regression with fixed-effects model defined by

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

where \( \alpha_i \) and \( \mu_t \) 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, \( y_{it} \) and fixed effects are scalars, and \( \beta \) is \( k \times 1 \) vector. To control the fixed-effects, let

\[ \tilde{y}_{it} = y_{it} - \bar{y}_i - \bar{y}_t + \bar{y}, \quad \tilde{x}_{it} = x_{it} - \bar{x}_i - \bar{x}_t + \bar{x}, \]

with \( \bar{y}_i = T^{-1} \sum_{t=1}^{T} y_{it}, \bar{y}_t = N^{-1} \sum_{i=1}^{N} y_{it}, \bar{y} = (NT)^{-1} \sum_{i=1}^{N} \sum_{t=1}^{T} y_{it}, \) and \( \bar{x}_i, \bar{x}_t \) and \( \bar{x} \) are defined similarly. Then the fixed-effects ordinary least squares (OLS) estimator of \( \beta \) is given by

\[ \hat{\beta} = \left( \sum_{i=1}^{N} \sum_{t=1}^{T} \tilde{x}_{it} \tilde{x}_{it}' \right)^{-1} \sum_{i=1}^{N} \sum_{t=1}^{T} \tilde{x}_{it} \tilde{y}_{it}. \]

One of the commonly used standard errors for \( \hat{\beta} \) in empirical and applied econometrics is the White (1980) heteroskedasticity robust standard error in the cross-sectional setting. In the presence of serial and cross-sectional correlations, the panel clustered standard errors may be biased. Newey and West (1987) 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 panel model because they are general forms of serial correlation under nonstationarity cases and robust to heteroskedasticity in the cross-section. This estimator focuses on the large-\( N \) small-\( T \) scenario. The case of large-\( N \) large-\( T \) is then studied by Ahn and Moon (2014), 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 an asymptotic theory that are robust to heteroskedasticity, autocorrelation and spatial correlation, which extended and generalized the asymptotic results by 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 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, Athey, Imbens, and Wooldridge (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 modelling 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 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, Choi, and Liao (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 methods we employ in this paper, banding and thresholding, are regularization methods, and 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 Appendix A.

Throughout this paper, $\nu_{\min}(A)$ and $\nu_{\max}(A)$ denote the minimum and maximum eigenvalues of matrix $A$. We use $||A|| = \sqrt{\nu_{\max}(A^TA)}$ and $||A||_1 = \max_i \sum_j |A_{ij}|$ as the operator norm and the $\ell_1$-norm of a matrix $A$, respectively.
2 OLS and Standard Error Estimation

We consider the following model:

\[ y_{it} = x'_{it}\beta + u_{it}, \quad (2.1) \]

where \( \beta \) 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:

\[ y_{t} = x_{t}\beta + u_{t}, \quad (2.2) \]

where \( y_{t} = (y_{1t}, ..., y_{Nt})' \) \((N \times 1)\), \( x_{t} = (x_{1t}, ..., x_{Nt})' \) \((N \times k)\), and \( u_{t} = (u_{1t}, ..., u_{Nt})' \) \((N \times 1)\). To economize notation, we define \( y_{i} = (y_{i1}, ..., y_{iT})' \) \((T \times 1)\), \( x_{i} = (x_{i1}, ..., x_{iT})' \) \((T \times k)\), and \( u_{i} = (u_{i1}, ..., u_{iT})' \) \((T \times 1)\). 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.

For a fixed \( i \), model (2.1) can be represented as

\[ y_{i} = x_{i}\beta + u_{i}. \]

The (pooled) ordinary least square (OLS) estimator of \( \beta \) from equations (2.1) and (2.2) may then be defined as

\[ \hat{\beta}_{OLS} = \left( \sum_{i=1}^{N} \sum_{t=1}^{T} x_{it}x_{it}' \right)^{-1} \sum_{i=1}^{N} \sum_{t=1}^{T} x_{it}y_{it} = \left( \sum_{t=1}^{T} x_{t}x_{t}' \right)^{-1} \sum_{t=1}^{T} x_{t}'y_{t}, \quad (2.3) \]

Let \( V_{X} \equiv \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} x_{it}x_{it}' \), then we can rewrite equation (2.3) as

\[ \hat{\beta}_{OLS} = \beta + V_{X}^{-1} \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} x_{it}u_{it} = \beta + V_{X}^{-1} \frac{1}{NT} \sum_{t=1}^{T} x_{t}'u_{t}. \]
The variance of \( \hat{\beta}_{OLS} \) depends on

\[
V \equiv \text{Var}(\frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \sum_{t=1}^{T} x_{it}u_{it})
\]

\[
= \frac{1}{NT} \sum_{t=1}^{T} E x_{it}'u_{it}u_{it}'x_{it} + \frac{1}{NT} \sum_{h=1}^{T-1} \sum_{t=h+1}^{T} [E x_{t}u_{t}u_{t-h}x_{t-h} + E x_{t-h}u_{t-h}u_{t}x_{t}].
\] (2.4)

The goal 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

\[
\hat{V}_{CX} = \frac{1}{NT} \sum_{i=1}^{N} x_{i}'\hat{u}_{i}\hat{u}_{i}'x_{i},
\]

with \( \hat{u}_{i} = y_{i} - x_{i}\hat{\beta}_{OLS} \) are the OLS residuals, and this estimator allows for arbitrary serial dependence and heteroskedasticity within individuals, where \( \hat{V}_{CX} \) assumes no cross-section correlation.

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

\[
\hat{V}_{CT} = \frac{1}{NT} \sum_{t=1}^{T} x_{i}'\hat{u}_{t}\hat{u}_{t}'x_{t},
\]

with \( \hat{u}_{t} = y_{t} - x_{t}\hat{\beta}_{OLS} \), where \( \hat{V}_{CT} \) assumes no serial correlation. To control for the serial correlation, a simple modification of \( \hat{V}_{CT} \) using Newey and West (1987) is

\[
\hat{V}_{DK} = \frac{1}{NT} \sum_{t=1}^{T} x_{i}'\hat{u}_{t}\hat{u}_{t}'x_{t} + \frac{1}{NT} \sum_{h=1}^{L} \sum_{t=h+1}^{T} [x_{t}x_{t-h}x_{t-h} + x_{t-h}u_{t-h}u_{t-h}x_{t}],
\] (2.6)

which is suggested by Driscoll and Kraay (1998). When \( N \) is large, however, (2.6) accumulates a large number of cross-sectional estimation noises.

More generally, let

\[
V_{ij} \equiv \frac{1}{T} \sum_{t=1}^{T} E x_{it}u_{it}u_{jt}x_{jt}' + \frac{1}{T} \sum_{h=1}^{T-1} \sum_{t=h+1}^{T} [E x_{it}u_{it}u_{jt-h}x_{jt-h} + E x_{i,t-h}u_{i,t-h}u_{jt}x_{jt}].
\]
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. The clustered variance estimator is

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

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 conditional 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 of pairs of \( (i,j) \). There is a partition \( \{(i,j): i, j \leq N\} = S_s \cup S_l \) 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. As suggested by Newey and West (1994), we can set \( L \) equal to \( 4(T/100)^{2/9} \). Also see Andrews (1991) for other popular kernel functions. Then as suggested by Newey and West (1987), \( V_{ij} \) can be approximated.
by

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

Then approximately,

\[ V \approx \frac{1}{N} \sum_{(i,j) \in S} V_{u,ij}. \]

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 “cross-sectional thresholding” method.

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

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

where \( \hat{u}_{it} = y_{it} - x'_{it}\hat{\beta}_{OLS} \). For a predetermined threshold value \( \lambda_{ij} \), we approximate \( S_{l} \) by

\[ \hat{S}_{l} = \{(i,j) : \|S_{u,ij}\| > \lambda_{ij}\}. \]

Hence, a “matrix hard-thresholding” estimator of \( V \) is

\[ \hat{V}_{Hard} \equiv \frac{1}{N} \sum_{(i,j) \in \hat{S}_{l} \cup \{i=j\}} S_{u,ij}. \]

As for the threshold value, we specify

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

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 next subsection. In addition, we can obtain \( \hat{V}_{DK} \) from \( \hat{V}_{Hard} \) by setting \( M = 0 \).

We also recommend a “matrix soft-thresholding” estimator as follow:

\[ \hat{V}_{Soft} \equiv \frac{1}{N} \sum_{i,j} \hat{S}_{u,ij}, \]
where $\hat{S}_{u,ij}$ is

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

where the $(k,k')$’s element of $A_{u,ij}$ is

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

for the threshold value

$$
\eta_{ij,kk'} = C \omega_{NT} \sqrt{|S_{u,ii',kk'}||S_{u,jj',kk'}|},
$$

where $\omega_{NT} = L \sqrt{\log(LN)}/T$ for some constant $C > 0$.

**Remark 2.1.** One advantage of the proposed method is that it does 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). Our method avoids this problem, while allowing correlations of unknown form within each cluster.

### 2.2 Choice of tuning parameters

Our suggested estimators, $\hat{V}_{Hard}$ and $\hat{V}_{Soft}$, 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, which is similar to Bickel and Levina (2008) and Fan, Liao, and Mincheva (2013). After obtaining the estimated residuals $\hat{u}_t$ by OLS, we split the data into two subsets, denoted by $\{\hat{u}_t\}_{t \in J_1}$ and $\{\hat{u}_t\}_{t \in J_2}$; 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 $T(J_1) \propto T$. 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 $p$th split,
we denote by $\hat{V}^p$ the sample covariance matrix based on the validation set, defined by

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

with $S^p_{u,ij} = T(J_2)^{-1} \sum_{t=1}^{T(J_2)} x_{it} \tilde{u}_{it} \tilde{u}_{jt} x'_{jt} + T(J_2)^{-1} \sum_{h=1}^{L} \omega(h, L) \sum_{t=h+1}^{T(J_2)} [x_{it} \tilde{u}_{it} \tilde{u}_{jt-h} x'_{jt-h} + x_{i,t-h} \tilde{u}_{i,t-h} \tilde{u}_{jt} x'_{jt}].$ Let $\hat{V}_{Hard}(M)$ be the hard thresholding estimator with threshold constant $M$ using the entire sample. Then we choose the constant $M^*$ by minimizing a cross-validation objective function

$$
M^* = \arg \min_{0 < M < M_0} \frac{1}{P} \sum_{p=1}^{P} \| \hat{V}_{Hard}(M) - \hat{V}^p \|_F^2,
$$

and the resulting estimator is $\hat{V}_{Hard}(M^*)$. We find that setting $M_0 = 1$ works well. So the minimization is taken over $M \in (0, 1)$ through a grid search. For the case of “soft-thresholding” estimator, $\hat{V}_{Soft}$, we choose $M$ in the same way.

The above procedure modifies that of Bickel and Levina (2008) and Fan, Liao, and Mincheva (2013) in two aspects. One is to use the entire sample when computing $\hat{V}_{Hard}(M)$ instead of $T(J_1)$. 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 $T(J_2)$ is taken as one of the $P$ blocks when computing $\hat{V}^p$, similar to the K-fold cross validation.

### 2.3 Consistency of the new estimator

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

$$
\alpha_{NT}(h) \equiv \sup_{X} \max_{t \leq T} \| \mathbb{E}(u_{t} u'_{t-h} | X) \| + \| \mathbb{E}(u_{t-h} u'_{t} | X) \|
$$

and

$$
\rho_{ij,h} \equiv \sup_{X} \max_{t \leq T} \| \mathbb{E}(u_{it} u_{jt-h} | X) \| + \| \mathbb{E}(u_{i,t-h} u_{jt} | X) \|.
$$

These coefficients give measures of autocovariances and cross-section covariances.

**Assumption 2.1.**

(i) $\mathbb{E}(u_t | x_t) = 0$.

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

8
Assumption 2.2. (weak serial and cross-sectional dependence).

(i) \( \sum_{h=1}^{\infty} \alpha_{NT}(h) \leq C \) for some \( C > 0 \).

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

(iii) For each fixed \( h \), \( \omega(h, L) \rightarrow 1 \) as \( L \rightarrow \infty \) and \( \max_{h \leq L} |\omega(h, L)| \leq C \) for some \( C > 0 \).

Assumption 2.2 (i) is an extension of the standard weak serial dependence condition to the case of high-dimensional panel data literature. Assumption 2.2 (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 \( \| \sum_{j=1}^{L} \rho_{ij,h} \| \). However the clusters that \( (i, j) \) has “small” \( \| Ex_{it}u_{it}u_{jt+h}x'_{jt+h} \| \) is unknown to us. Hence the appealing feature of our method is that we allow for unknown clusters. Essentially the assumption \( \omega_{NT}^{-q} \max_{i \leq N} \sum_{j=0}^{L} (\sum_{h=0}^{N} \rho_{ij,h})^q = o(1) \) controls the order of elements in \( S_t \).

Assumption 2.3. There is \( c_1 > 0 \), for all \( i \), \( \| V_{u,it} \| > c_1 \). A sufficient condition of this assumption is that \( \| \frac{1}{T} \sum_{t=1}^{T} \sum_{s=1}^{T} Ex_{it}u_{it}x'_{is} \| > 2c_1 \).

Assumption 2.4. (i) Exponential tail: There exist \( r_1, r_2 > 0 \) and \( b_1, b_2 > 0 \), such that for any \( s > 0, i \leq N \) and \( j \leq r, P(|u_{it}| > s) \leq \exp(-(s/b_1)^{r_1}) \), and \( P(|X_t| > s) \leq \exp(-(s/b_2)^{r_2}) \).

(ii) Strong mixing: There exist \( r_3 > 0, C > 0 \) such that for all \( T > 0, r_1^{-1} + r_2^{-1} + r_3^{-1} < 1 \)

\[
\sup_{A \in \mathcal{F}_t, B \in \mathcal{F}_\infty} |P(A)P(B) - P(AB)| < \exp(-CT^{r_3}),
\]

where \( \mathcal{F}_-^{\infty} \) 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.

We have the following main theorem and all proofs are contained in Appendix A1.

Theorem 2.1. Suppose the eigenvalues of \( V \) and \( V_X \) are bounded away from both zero and infinity. Under Assumption 2.1-2.4 as \( N, T \rightarrow \infty \),

\[
\sqrt{NT}[V_X^{-1}\hat{\Sigma}V_X^{-1}]^{-1/2} (\hat{\beta}_{OLS} - \beta) \overset{d}{\rightarrow} \mathcal{N}(0, I).
\]

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

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

3 Monte Carlo evidence

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

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

where the true \( \beta_0 = 1 \). The DGP allows for serial and cross-sectional correlation in both \( x_{it} \) and \( u_{it} \), and also heteroskedasticity as follow:

\[
x_{it} = a_i \nu_{i+1,t} + \nu_{i,t} + b_i \nu_{i-1,t}, \quad \nu_{it} = \rho \nu_{i,t-1} + \epsilon_{it}, \quad \epsilon_{it} \sim N(1, 1), \quad \nu_{i0} = 0,
\]

\[
u_{it} = a_i m_{i+1,t} + m_{i,t} + b_i m_{i-1,t}, \quad m_{it} = \rho m_{i,t-1} + \epsilon_{it}, \quad \epsilon_{it} \sim N(0, 1), \quad m_{i0} = 0,
\]

\[
\alpha_i \sim N(0, 0.5), \quad \mu_t \sim N(0, 0.5),
\]

where the constants \( \{a_i, b_i\}_{i=1}^N \) are i.i.d. Uniform(0, \( \gamma \)), which introduce cross-sectional correlation, and heteroskedasticity when \( \gamma > 0 \). The regressor is uncorrelated with the error term \( u_{it} \) each other. \( \nu_{it} \) and \( m_{it} \) is modeled as AR(1) process with the autoregressive parameter \( \rho \). 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 (\( \rho = 0.3, \gamma = 1 \)).

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 \( \hat{\beta}_{OLS} \): the standard White estimator given by \( \hat{V}_{White} = \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T x_{it}x_{it}' \hat{u}_{it}^2 \), where \( \bar{x}_{it} \) is demeaned version of regressor as we discussed in Section 1. Two types of clustered standard errors, \( \hat{V}_{CX} \) and \( \hat{V}_{CT} \), as defined in Section 2. In addition, we use two types of Newey and West HAC estimators for the panel version as follows:

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

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

Note that \(\hat{V}_{HAC}\) assumes cross-sectional independence, while \(\hat{V}_{DK}\) allows arbitrary cross-sectional dependence. In addition, \(\hat{V}_{DK}\) can be obtained when \(M = 0\), and \(\hat{V}_{HAC}\) can be obtained with a larger \(M\) from our thresholding standard-error estimator.

Results are given for sample sizes \(N = 50, 100, 200\) and \(T = 100, 200\). For each \(\{N, T\}\) combination, we use different truncation parameters for our proposed estimator, \(\hat{V}_{Hard}\), and also \(\hat{V}_{DK}\) and \(\hat{V}_{HAC}\): \(L = 0.1T, 0.4T, 0.7T, 0.9T\) when there is only serial correlation. In other cases, we use \(L = 3, 7, 11, 15\) for the truncation parameters. We also use Bartlett kernel for these three estimators. For the thresholding constant parameters of our proposed estimator, we set \(M = 0.10, 0.15, 0.20, 0.25\) in all cases. The simulation is replicated for one thousand times in all cases and the nominal significance level is 0.05. Because there is only one regressor, we used hard thresholding method. Simulation results are reported in Tables 1 - 4.

### 3.2 Results

Tables 1 - 4 present the simulation results, where each table corresponds to a different \(\{\rho, \gamma\}\) combination. Each table presents result of null rejection probabilities for 5% level tests based on six different standard errors. As expected, an 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 when there is serial, cross-sectional or both correlations in the model.

In Table 1 when there is no correlation, all the estimators performs well, but especially White standard error estimators gives rejection probabilities close to 0.05. In Table 2 the performances of \(\hat{V}_{CX}\) and \(\hat{V}_{HAC}\) are markedly better than others except small sample size. Since there is only serial correlation in the error term, these estimators take this correlation into account and perform well. As the lag length increases, the standard error estimated by \(\hat{V}_{HAC}\) increases to a level similar to the results of \(\hat{V}_{CX}\) 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, \(\hat{V}_{CX}\), dominates the standard error of Newey-West panel version, \(\hat{V}_{HAC}\). In addition, our proposed estimator, \(\hat{V}_{Hard}\), also performs well when we use both larger threshold constant \(M\) and bandwidth \(L\).

Table 3 considers the case of cross-sectionally correlated errors and regressors in the model

---

1. Note that \(\hat{V}_{HAC}\) assumes cross-sectional independence, while \(\hat{V}_{DK}\) allows arbitrary cross-sectional dependence. In addition, \(\hat{V}_{DK}\) can be obtained when \(M = 0\), and \(\hat{V}_{HAC}\) can be obtained with a larger \(M\) from our thresholding standard-error estimator.

2. Results are given for sample sizes \(N = 50, 100, 200\) and \(T = 100, 200\). For each \(\{N, T\}\) combination, we use different truncation parameters for our proposed estimator, \(\hat{V}_{Hard}\), and also \(\hat{V}_{DK}\) and \(\hat{V}_{HAC}\): \(L = 0.1T, 0.4T, 0.7T, 0.9T\) when there is only serial correlation. In other cases, we use \(L = 3, 7, 11, 15\) for the truncation parameters. We also use Bartlett kernel for these three estimators. For the thresholding constant parameters of our proposed estimator, we set \(M = 0.10, 0.15, 0.20, 0.25\) in all cases. The simulation is replicated for one thousand times in all cases and the nominal significance level is 0.05. Because there is only one regressor, we used hard thresholding method. Simulation results are reported in Tables 1 - 4.

3. Tables 1 - 4 present the simulation results, where each table corresponds to a different \(\{\rho, \gamma\}\) combination. Each table presents result of null rejection probabilities for 5% level tests based on six different standard errors. As expected, an 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 when there is serial, cross-sectional or both correlations in the model.

In Table 1 when there is no correlation, all the estimators performs well, but especially White standard error estimators gives rejection probabilities close to 0.05. In Table 2 the performances of \(\hat{V}_{CX}\) and \(\hat{V}_{HAC}\) are markedly better than others except small sample size. Since there is only serial correlation in the error term, these estimators take this correlation into account and perform well. As the lag length increases, the standard error estimated by \(\hat{V}_{HAC}\) increases to a level similar to the results of \(\hat{V}_{CX}\) 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, \(\hat{V}_{CX}\), dominates the standard error of Newey-West panel version, \(\hat{V}_{HAC}\). In addition, our proposed estimator, \(\hat{V}_{Hard}\), also performs well when we use both larger threshold constant \(M\) and bandwidth \(L\).

Table 3 considers the case of cross-sectionally correlated errors and regressors in the model.
with \( \rho = 0 \) and \( \gamma = 1 \). Not surprisingly, except the case of small sample size, \( \hat{V}_{CT} \) and \( \hat{V}_{DK} \) with small bandwidth \( L \) have rejection probabilities close to 0.05. However, notice that the rejection rate of \( \hat{V}_{DK} \) and also \( \hat{V}_{Hard} \) tend to over-reject substantially as the lag length \( L \) increases. In addition, as cross-section size \( N \) increases, the over-rejection problem becomes worse as we mentioned in Section 2. This tendency is easy to explain. Since \( \hat{V}_{DK} \) is an estimator based on a single time series and it is zero when full weight is given to the sample autocovariances, the bias in \( \hat{V}_{DK} \) initially falls but then increases as the lag length increases, while the variance of \( \hat{V}_{DK} \) is initially increasing but eventually becomes decreasing. Hence, \( \hat{V}_{DK} \) 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 small size of \( L \) and \( M \), \( \hat{V}_{Hard} \) gives less bias on the estimated standard error.

Table 4 allows both cross-sectional and serial correlation in the error term. Rejections are still computed using the traditional \( N(0, 1) \) critical value. Here we also use smaller size of leg length \( L = \{3, 7, 11, 15\} \) to obtain unbiased standard error estimators. Not surprisingly, all estimators except \( \hat{V}_{Hard} \) and \( \hat{V}_{DK} \) tend to over-reject substantially, compared with the case of one type dependency. When \( T = 50 \), \( \hat{V}_{Hard} \) and \( \hat{V}_{DK} \) also tend to over-reject. When \( T \) is large, however, rejection probabilities of \( \hat{V}_{Hard} \) and \( \hat{V}_{DK} \) are close to 0.05. Unreported results with larger bandwidth, \( L \), give more over sized estimator. By varying the thresholding value, \( M \), our estimators are closer to 0.05 than that of \( \hat{V}_{DK} \). This shows that we can obtain unbiased standard error estimator and appropriate rejection rate using our proposed estimator, \( \hat{V}_{Hard} \). Finally, we confirm that our suggested estimator gives unbiased standard error and rejections close to 0.05 by choosing thresholding and bandwidth in different forms of correlations.

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 the divorce rates. Friedberg (1998) found that state law reforms significantly increased divorce rates with controls for state and year fixed effects. However, her assumption is that unilateral divorce affected the divorce rate by a permanent constant. On the other hand, Wolfers (2006) investigated the question of whether law reform continue to have an impact on the divorce 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

\[
y_{it} = \alpha_i + \mu_t + \sum_{k=1}^{8} \beta_k X_{it,k} + \delta_i t + \delta_{it} + u_{it},
\]

(4.1)
where $y_{it}$ is the divorce rate for state $i$ and year $t$; $\alpha_i$ and $\mu_t$ are the state and year fixed effects; $X_{it,k}$ is a binary regressor that representing the treatment effect $2k$ years after the reform; $\delta_{it}$ a linear time trend. Based on the results, Wolfers (2006) found 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”. He 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 of these correlations might lead to bias in the standard error estimators. Therefore, 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 the models both with and without linear time trend, and employ least square estimation. We 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 existence of cross-sectional correlations. The estimated $\beta_1, \cdots , \beta_8$ with and without linear time trend and their different types of standard errors are presented respectively in Table 5 below. 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 5, 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 after the law reform. However within a decade, divorce rates had fallen over time. Most of 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 is 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 Conclusions

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. The estimator is asymptotically unbiased with an improved convergence rate, and is more general than existing methods in the panel data literature. Simulated experiments demonstrate the robustness of the new standard-error estimator to various correlation structures.
Table 1: Null Rejection Probabilities, 5% level, Two-Tailed Test of $H_0 : \beta = 1$. No serial and cross-sectional correlations ($\rho = 0, \gamma = 0$).

| N  | T  | L\ M | 0.10 | 0.15 | 0.20 | 0.25 | $V_{Hard}$ | $V_{HAC}$ | $V_{DK}$ | $V_{CX}$ | $V_{CT}$ | $V_{W}$ |
|----|----|-----|------|------|------|------|-------------|-------------|----------|---------|---------|--------|
| 50 | 100| 3   | .067 | .067 | .065 | .062 | .054        | .067        | .059     | .055    | .053    |        |
| 7  | .078 | .079 | .072 | .063 |        | .054 | .083        | .059        | .055     | .053    |        |
| 11 | .091 | .072 | .066 | .060 |        | .054 | .091        | .059        | .055     | .053    |        |
| 15 | .091 | .071 | .058 | .053 |        | .053 | .105        | .059        | .055     | .053    |        |
| 50 | 200| 3   | .055 | .057 | .055 | .054 | .048        | .054        | .055     | .043    | .049    |        |
| 7  | .056 | .055 | .053 | .051 |        | .049 | .061        | .055        | .043     | .049    |        |
| 11 | .058 | .058 | .052 | .050 |        | .049 | .066        | .055        | .043     | .049    |        |
| 15 | .055 | .059 | .048 | .049 |        | .049 | .069        | .055        | .043     | .049    |        |
| 100| 100| 3   | .063 | .060 | .059 | .060 | .049        | .064        | .054     | .056    | .050    |        |
| 7  | .075 | .065 | .064 | .056 |        | .049 | .077        | .054        | .056     | .050    |        |
| 11 | .075 | .065 | .056 | .048 |        | .048 | .083        | .054        | .056     | .050    |        |
| 15 | .081 | .059 | .052 | .049 |        | .049 | .092        | .054        | .056     | .050    |        |
| 100| 200| 3   | .048 | .042 | .045 | .046 | .033        | .041        | .038     | .038    | .033    |        |
| 7  | .047 | .049 | .046 | .040 |        | .033 | .054        | .038        | .038     | .033    |        |
| 11 | .056 | .048 | .044 | .038 |        | .034 | .060        | .038        | .038     | .033    |        |
| 15 | .058 | .046 | .036 | .034 |        | .034 | .070        | .038        | .038     | .033    |        |
| 200| 100| 3   | .055 | .052 | .050 | .054 | .036        | .054        | .041     | .043    | .037    |        |
| 7  | .064 | .058 | .056 | .043 |        | .038 | .072        | .041        | .043     | .037    |        |
| 11 | .068 | .059 | .041 | .035 |        | .037 | .087        | .041        | .043     | .037    |        |
| 15 | .070 | .044 | .038 | .037 |        | .037 | .097        | .041        | .043     | .037    |        |
| 200| 200| 3   | .057 | .058 | .051 | .049 | .048        | .053        | .048     | .051    | .048    |        |
| 7  | .063 | .057 | .052 | .051 |        | .048 | .058        | .048        | .051     | .048    |        |
| 11 | .064 | .058 | .057 | .050 |        | .048 | .069        | .048        | .051     | .048    |        |
| 15 | .070 | .053 | .047 | .048 |        | .048 | .082        | .048        | .051     | .048    |        |
Table 2: Null Rejection Probabilities, 5% level, Two-Tailed Test of $H_0 : \beta = 1$. Only serial correlation ($\rho = 0.5, \gamma = 0$).

| N  | T  | L\M | $V_{Hard}$ | $V_{HAC}$ | $V_{DK}$ | $V_{CX}$ | $V_{CT}$ | $V_{W}$ |
|----|----|-----|------------|-----------|---------|---------|---------|--------|
| 50 | 100| 0.1T | .112       | .107      | .065    | .146    | .140    |
|    |    | .04T | .069       | .069      | .069    | .211    | .065    | .146    |
|    |    | .06T | .070       | .070      | .070    | .267    | .065    | .146    |
|    |    | .09T | .068       | .068      | .068    | .337    | .065    | .146    |
| 50 | 200| 0.1T | .071       | .064      | .089    | .061    | .131    | .132    |
|    |    | .04T | .069       | .069      | .069    | .192    | .061    | .131    |
|    |    | .06T | .068       | .068      | .068    | .251    | .061    | .131    |
|    |    | .09T | .064       | .064      | .064    | .324    | .061    | .131    |
| 100| 100| 0.1T | .079       | .053      | .082    | .040    | .109    | .110    |
|    |    | .04T | .048       | .048      | .048    | .183    | .040    | .109    |
|    |    | .06T | .046       | .046      | .046    | .246    | .040    | .109    |
|    |    | .09T | .046       | .046      | .046    | .323    | .040    | .109    |
| 100| 200| 0.1T | .074       | .057      | .092    | .055    | .119    | .118    |
|    |    | .04T | .058       | .058      | .058    | .187    | .055    | .119    |
|    |    | .06T | .057       | .057      | .057    | .239    | .055    | .119    |
|    |    | .09T | .056       | .056      | .056    | .321    | .055    | .119    |
| 200| 100| 0.1T | .093       | .062      | .103    | .048    | .134    | .122    |
|    |    | .04T | .056       | .056      | .056    | .189    | .048    | .134    |
|    |    | .06T | .053       | .053      | .053    | .251    | .048    | .134    |
|    |    | .09T | .051       | .051      | .051    | .328    | .048    | .134    |
| 200| 200| 0.1T | .063       | .049      | .050    | .050    | .079    | .039    |
|    |    | .04T | .045       | .045      | .045    | .191    | .039    | .132    |
|    |    | .06T | .042       | .042      | .042    | .233    | .039    | .132    |
|    |    | .09T | .041       | .041      | .041    | .319    | .039    | .132    |
Table 3: Null Rejection Probabilities, 5% level, Two-Tailed Test of $H_0 : \beta = 1$. Only cross-sectional correlation ($\rho = 0, \gamma = 1$).

| N   | T   | L \ M | $V_{Hard}$ | $V_{HAC}$ | $V_{DK}$ | $V_{CX}$ | $V_{CT}$ | $V_{W}$ |
|-----|-----|-------|------------|-----------|----------|----------|----------|--------|
| 50  | 100 | 3     | .062       | .061      | .061     | .062     | .148     | .061   |
|     |     |       | .071       | .070      | .067     | .069     | .146     | .071   |
|     |     |       | .077       | .072      | .082     | .108     | .143     | .083   |
|     |     |       | .090       | .085      | .119     | .143     | .143     | .097   |
| 50  | 200 | 3     | .047       | .049      | .048     | .051     | .142     | .047   |
|     |     |       | .057       | .056      | .054     | .061     | .144     | .054   |
|     |     |       | .058       | .055      | .065     | .072     | .145     | .059   |
|     |     |       | .062       | .063      | .075     | .094     | .145     | .064   |
| 100 | 100 | 3     | .048       | .048      | .052     | .046     | .138     | .048   |
|     |     |       | .060       | .061      | .056     | .058     | .138     | .060   |
|     |     |       | .068       | .069      | .079     | .099     | .138     | .078   |
|     |     |       | .078       | .079      | .111     | .139     | .139     | .086   |
| 100 | 200 | 3     | .048       | .050      | .050     | .048     | .148     | .049   |
|     |     |       | .055       | .057      | .050     | .055     | .146     | .051   |
|     |     |       | .061       | .057      | .064     | .077     | .146     | .052   |
|     |     |       | .066       | .068      | .084     | .113     | .146     | .062   |
| 200 | 100 | 3     | .066       | .065      | .065     | .061     | .157     | .067   |
|     |     |       | .070       | .070      | .070     | .066     | .158     | .075   |
|     |     |       | .079       | .077      | .082     | .108     | .159     | .085   |
|     |     |       | .087       | .081      | .126     | .160     | .160     | .105   |
| 200 | 200 | 3     | .063       | .060      | .061     | .059     | .147     | .062   |
|     |     |       | .067       | .067      | .063     | .065     | .145     | .071   |
|     |     |       | .071       | .071      | .073     | .083     | .146     | .075   |
|     |     |       | .082       | .087      | .087     | .118     | .146     | .083   |
Table 4: Null Rejection Probabilities, 5% level, Two-Tailed Test of $H_0 : \beta = 1$. Both serial and cross-sectional correlations ($\rho = 0.3, \gamma = 1$).

| N  | T   | L \ M | $V_{Hard}$ | $V_{HAC}$ | $V_{DK}$ | $V_{CX}$ | $V_{CT}$ | $V_{W}$ |
|----|-----|-----|-----------|-----------|----------|----------|----------|--------|
|    |     |     | .010 | .015 | .020 | .025 |         |        |        |
| 50 | 100 | 3   | .062 | .063 | .064 | .065 | .164 | .064 | .166 | .072 | .187 |
|    |     | 7   | .063 | .068 | .070 | .071 | .162 | .065 | .166 | .072 | .187 |
|    |     | 11  | .079 | .078 | .085 | .117 | .160 | .081 | .166 | .072 | .187 |
|    |     | 15  | .088 | .094 | .125 | .156 | .159 | .097 | .166 | .072 | .187 |
| 50 | 200 | 3   | .056 | .055 | .056 | .055 | .156 | .056 | .174 | .069 | .192 |
|    |     | 7   | .063 | .061 | .065 | .070 | .151 | .061 | .174 | .069 | .192 |
|    |     | 11  | .067 | .070 | .071 | .085 | .150 | .068 | .174 | .069 | .192 |
|    |     | 15  | .076 | .074 | .092 | .106 | .147 | .073 | .174 | .069 | .192 |
| 100| 100 | 3   | .070 | .069 | .068 | .069 | .158 | .071 | .154 | .080 | .188 |
|    |     | 7   | .080 | .072 | .077 | .084 | .155 | .085 | .154 | .080 | .188 |
|    |     | 11  | .089 | .085 | .092 | .109 | .157 | .095 | .154 | .080 | .188 |
|    |     | 15  | .098 | .097 | .121 | .158 | .158 | .106 | .154 | .080 | .188 |
| 100| 200 | 3   | .059 | .058 | .057 | .054 | .154 | .060 | .159 | .066 | .181 |
|    |     | 7   | .059 | .059 | .059 | .057 | .151 | .062 | .159 | .066 | .181 |
|    |     | 11  | .066 | .062 | .065 | .073 | .149 | .067 | .159 | .066 | .181 |
|    |     | 15  | .066 | .069 | .085 | .112 | .148 | .073 | .159 | .066 | .181 |
| 200| 100 | 3   | .074 | .073 | .073 | .068 | .160 | .072 | .158 | .087 | .203 |
|    |     | 7   | .077 | .073 | .073 | .075 | .158 | .079 | .158 | .087 | .203 |
|    |     | 11  | .086 | .079 | .090 | .116 | .158 | .088 | .158 | .087 | .203 |
|    |     | 15  | .096 | .092 | .132 | .156 | .156 | .102 | .158 | .087 | .203 |
| 200| 200 | 3   | .060 | .056 | .057 | .057 | .149 | .061 | .151 | .071 | .174 |
|    |     | 7   | .060 | .059 | .056 | .063 | .147 | .056 | .151 | .071 | .174 |
|    |     | 11  | .063 | .061 | .068 | .079 | .146 | .065 | .151 | .071 | .174 |
|    |     | 15  | .068 | .069 | .084 | .107 | .146 | .071 | .151 | .071 | .174 |
Table 5: 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 |
| 12 years | .256                 | .086*      | .140   | .139      | .172      | .155      | .148     |
| 34 years | .209                 | .086*      | .081*  | .159      | .075*     | .114      | .104*    | .089*    |
| 56 years | .126                 | .086       | .073   | .168      | .064*     | .105      | .088     | .069     |
| 78 years | .105                 | .086       | .070   | .165      | .059      | .100      | .065     | .040*    |
| 910 years | -.122                | .085       | .060*  | .161      | .041*     | .088      | .058*    | .054*    |
| 1112 years | -.344                | .085*      | .071*  | .173*     | .043*     | .101*     | .056*    | .075*    |
| 1314 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 |
| 12 years | .286                 | .064*      | .152   | .206      | .143*     | .185      | .145*    | .140*    |
| 34 years | .254                 | .071*      | .099*  | .171      | .102*     | .140      | .134     | .126*    |
| 56 years | .186                 | .079*      | .102   | .206      | .110      | .145      | .148     | .143     |
| 78 years | .177                 | .086*      | .109   | .230      | .120      | .153      | .155     | .146     |
| 910 years | -.037                | .093       | .111   | .241      | .120      | .156      | .164     | .154     |
| 1112 years | -.247                | .100*      | .128   | .268      | .141      | .179      | .196     | .183     |
| 1314 years | -.386                | .108*      | .137*  | .296      | .164*     | .193*     | .218     | .209     |
| 15+ years | -.414                | .120*      | .158*  | .337      | .186*     | .221      | .251     | .243     |

**Note:** Standard errors with asterisks indicate significance at 5% level using $N(0,1)$ critical values; $se_{OLS}$ 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).
A Appendix

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

A.1 Proof of Theorem 2.1

First let

\[
V_L = \frac{1}{NT} \sum_t {E_t u_t u'_t x_t} + \frac{1}{NT} \sum_{h=1}^{T-1} \sum_{t=h+1}^{T} {E_t u_t u'_{t-h} x_{t-h}} + {E_{t-h} u_{t-h} u'_{t-h} x_t}.
\]

We need following lemmas to prove the main results.

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

**Proof.** First note that

\[
\|E_t u_t u'_t x_{t-h} + E_{t-h} u_{t-h} u'_t x_t\| \leq \|x_t\| \|E(u_t u'_t x_{t-h})\| \|x_{t-h}\| + \|x_{t-h}\| \|E(u_{t-h} u'_t x_t)\| \|x_{t}\|
\leq \alpha_{NT} \sqrt{\|x_t\|^2 \|x_{t-h}\|^2} = \alpha_{NT} \|x_{t}\|^2 \leq NC \alpha_{NT} \|x_{t}\|^2.
\]

Hence

\[
\|V - V_L\| \leq \frac{1}{NT} \sum_{h=L}^{T-1} \sum_{t=h}^{T} \|E_t u_t u'_t x_{t-h} + E_{t-h} u_{t-h} u'_t x_t\|
+ \frac{1}{NT} \sum_{h=1}^{L} (1 - \omega(h, L)) \sum_{t=h+1}^{T} \|E_t u_t u'_t x_{t-h} + E_{t-h} u_{t-h} u'_t x_t\|
\leq \frac{NC}{NT} \sum_{h=L}^{T-1} \sum_{t=1}^{T} \alpha_{NT}(h) + \frac{NC}{NT} \sum_{h=1}^{L} (1 - \omega(h, L)) \sum_{t=1}^{T} \alpha_{NT}(h)
\leq C \sum_{h=L}^{T-1} \alpha_{NT}(h) + C \sum_{h=1}^{L} (1 - \omega(h, L)) \alpha_{NT}(h) = o_P(1).
\]

Second term of the last equation goes to zero due to Assumption 2.2(iii) and the dominated convergence theorem.
Lemma A.2. Suppose $\log N = o(T)$. For $f(t, h, L) = \omega(h, L)1\{t > h\}$,

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

Proof. The left hand side can be written as

$$\max_h \max_{i,j} \left\| \frac{1}{T} \sum_t Z_{h,i,j,t} \right\|,$$

where $Z_{h,i,j,t} = f(t, h, L)(x_{it}u_{it}x_{j,t-h} - Ex_{it}u_{it}x_{j,t-h}).$

For convenience, assume that $\dim(Z_{h,i,j,t}) = 1$ and there is no serial correlation. Set $\alpha_n = \sqrt{\log(LN)N}T$ and $c^2 = 2C$ for $c, C > 0$. Then, by using Bernstein Inequality and exponential tail conditions, and that $f(t, h, L)$ is bounded,

$$P(\max_h \max_{i,j} \left\| \frac{1}{T} \sum_{t=1}^T Z_{h,i,j,t} \right\| > c\alpha_n) \leq LN^2 \max_{i,j} P\left( \left\| \frac{1}{T} \sum_{t=1}^T Z_{h,i,j,t} \right\| > c\alpha_n \right) \leq LN^2 \exp\left( -\frac{c^2\alpha_n^2}{C} \right)$$

$$\leq \exp\left( \log(LN) - \frac{c^2\alpha_n^2}{C} \right)$$

$$= \exp(-\log(LN))$$

$$= \frac{1}{LN} \to 0.$$

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

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

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

$$a_1 = \max_h \max_{i,j} \left\| \frac{1}{T} \sum_{t=1}^T x_{it}(\tilde{u}_{it} - u_{it})(\tilde{u}_{j,t-h} - u_{j,t-h})x_{j,t-h}f(t, h, L) \right\|$$

$$a_2 = \max_h \max_{i,j} \left\| \frac{1}{T} \sum_{t=1}^T x_{it}u_{it}(\tilde{u}_{j,t-h} - u_{j,t-h})x_{j,t-h}f(t, h, L) \right\|$$

$$a_3 = \max_h \max_{i,j} \left\| \frac{1}{T} \sum_{t=1}^T x_{it}(\tilde{u}_{it} - u_{it})u_{j,t-h}x_{j,t-h}f(t, h, L) \right\|. $$
For simplicity, let’s assume $\dim(x_{it}) = 1$. Then

$$a_1 \leq \|\hat{\beta} - \beta\|^2 \max_h \max_{i,j} \| \frac{1}{T} \sum_{t=1}^{T} x_{it}x_{jt,t-h}f(t, h, L) \|
$$

$$\leq O_P\left( \frac{1}{NT} \right) \max_{i,j} \| x_{it} \|^4 = O_P\left( \frac{1}{NT} \right).$$

By using Bernstein Inequality for weakly dependent data and exponential tail conditions, and that $f(t, h, L)$ is bounded,

$$a_2 \leq \|\hat{\beta} - \beta\| \max_h \max_{i,j} \| \frac{1}{T} \sum_{t=1}^{T} x_{it}u_{it}x_{jt,t-h}x_{jt,t-h}f(t, h, L) \|
$$

$$\leq O_P\left( \frac{1}{\sqrt{NT}} \right) O_P\left( \sqrt{\frac{\log(LN)}{T}} \right)
$$

$$= O_P\left( \frac{1}{T} \sqrt{\frac{\log(LN)}{N}} \right).$$

$a_3$ is bounded using the same argument. Together,

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

□

**Proof of Theorem 2.1.** It suffice to prove $\|\hat{V} - V\| = o_P(1)$. We have

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

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

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

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

$$\|\hat{V} - V_L\| \leq \frac{1}{N} \sum_{\hat{S}_{u,ij} = 0} \| V_{u,ij} - \hat{S}_{u,ij} \| + \frac{1}{N} \sum_{\hat{S}_{u,ij} \neq 0} \| V_{u,ij} - \hat{S}_{u,ij} \|.$$
Note that \( \| S_{u,ij} - V_{u,ij} \| < \frac{1}{2} \lambda_{ij} \) for \( \forall (i, j) \) and \( C_1 > 0 \)

\[
\| S_{u,ii} \| \geq \| V_{u,ii} \| - \max_{ij} \| S_{u,ii} - V_{u,ii} \|
\geq \| V_{u,ii} \| - C \omega_{NT} > C_1.
\]

From Assumption 2.2, \( \| V_{u,ii} \| > c_1 > 0 \), then, \( \lambda_{ij} = M \omega_{NT} \sqrt{\| S_{u,ii} \| \| S_{u,jj} \| } > C \omega_{NT} > 2c_1 \omega_{NT} \). 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 \( \rho_{ij,h} = \sup_x \max_t |E(u_{it}u_{jt} - h|X) + |E(u_{it}u_{jt} - h|X) + E(u_{it}u_{jt} - h|X)| \). Then,

\[
\| V_{u,ij} \| \leq \frac{1}{T} \sum_t Ex_i u_{it} u_{jt} x_{jt}' + \frac{1}{T} \sum_{h=1}^L \omega(h, L) \sum_{t=h+1}^T [Ex_i u_{it} u_{jt} - h x_{jt}' + Ex_i u_{it} - h u_{jt} x_{jt}]
\]

\[
\leq C \rho_{i,j,0}/2 + C \frac{1}{T} \sum_{h=1}^L \omega(h, L) \sum_{t=h+1}^T \rho_{i,j,h} \leq C \sum_{h=0}^L \rho_{i,j,h}.
\]

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

\[
\frac{1}{N} \sum_{S_{u,ij}=0} \| V_{u,ij} - \hat{S}_{u,ij} \| \leq \frac{1}{N} \sum_{S_{u,ij}=0} \| V_{u,ij} \|
\leq \frac{1}{N} \sum_{ij} \| V_{u,ij} \| \{ \| S_{u,ij} \| < \lambda_{ij} \}
\]

\[
= \frac{1}{N} \sum_{ij} \| V_{u,ij} \| \{ \| V_{u,ij} \| < \| S_{u,ij} \| + \| S_{u,ij} - V_{u,ij} \|, \| S_{u,ij} \| < \lambda_{ij} \}
\]

\[
= \frac{1}{N} \sum_{ij} \| V_{u,ij} \| \{ \| V_{u,ij} \| < \lambda_{ij} + 0.5 \lambda_{ij} \}
\]

\[
\leq \frac{1}{N} \sum_{ij} \| V_{u,ij} \| (1.5 \lambda_{ij})^{1-q} \| V_{u,ij} \|^{1-q} \{ \| V_{u,ij} \| < 1.5 \lambda_{ij} \}
\]

\[
\leq \frac{1}{N} \sum_{ij} \| V_{u,ij} \|^q (1.5 \lambda_{ij})^{1-q} \leq \frac{1}{N} \sum_{ij} \| V_{u,ij} \|^q \omega_{NT}^{1-q}
\]

\[
\leq C \omega_{NT}^{1-q} \max_i \sum_j \{ \sum_{h=0}^L \rho_{i,j,h} \}^q,
\]
On the other hand, on the event \( \max_{ij} \| S_{u,ij} - V_{u,ij} \| \leq C \omega_{NT} \),

\[
\frac{1}{N} \sum_{\hat{S}_{u,ij} \neq 0} \| V_{u,ij} - \hat{S}_{u,ij} \| \leq \frac{1}{N} \sum_{\hat{S}_{u,ij} \neq 0} \| V_{u,ij} - S_{u,ij} \| + \frac{1}{N} \sum_{\hat{S}_{u,ij} \neq 0} \| S_{u,ij} - \hat{S}_{u,ij} \|
\]

\[
\leq \frac{1}{N} \sum_{ij} 0.5 \lambda_{ij} + \frac{1}{N} \sum_{ij} \lambda_{ij} \leq \frac{1}{N} \sum_{ij} 1.5 \lambda_{ij} 1\{\| S_{u,ij} \| > \lambda_{ij} \}
\]

\[
= \frac{1}{N} \sum_{ij} 1.5 \lambda_{ij} 1\{\| V_{u,ij} \| > \| S_{u,ij} \| - \| S_{u,ij} - V_{u,ij} \|, \| S_{u,ij} \| > \lambda_{ij} \}
\]

\[
= \frac{1}{N} \sum_{ij} 1.5 \lambda_{ij} 1\{\| V_{u,ij} \| > \lambda_{ij} - 0.5 \lambda_{ij} \}
\]

\[
\leq \frac{1}{N} \sum_{ij} \lambda_{ij} 1\{|V_{u,ij}|^q \leq (0.5 \lambda_{ij})^q \} 1\{|V_{u,ij}| > 0.5 \lambda_{ij} \}
\]

\[
\leq \frac{1}{N} \sum_{ij} C \lambda_{ij} 1-q |V_{u,ij}|^q \leq \frac{1}{N} \sum_{ij} |V_{u,ij}|^q C \omega_{NT}^{1-q}
\]

\[
\leq C \omega_{NT}^{1-q} \max_i \sum_j \left( \sum_{h=0}^{L} \rho_{ij,h} \right)^q.
\]

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

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

**Remaining proofs**: \( \max_{ij} \| S_{u,ij} - V_{u,ij} \| = O_P(\omega_{NT}) \)

Recall

\[
S_{u,ij} := \frac{1}{T} \sum_t x_{it} \tilde{u}_{it} \tilde{u}_{jt} x'_{jt} + \frac{1}{T} \sum_{h=1}^{L} \omega(h, L) \sum_{t=h+1}^{T} [x_{it} \tilde{u}_{it} \tilde{u}_{jt} x'_{jt} + x_{i,t-h} \tilde{u}_{i,t-h} \tilde{u}_{jt} x'_{jt}],
\]

\[
V_{u,ij} := \frac{1}{T} \sum_t E x_{it} u_{it} u_{jt} x'_{jt} + \frac{1}{T} \sum_{h=1}^{L} \omega(h, L) \sum_{t=h+1}^{T} [E x_{it} u_{it} u_{jt} x'_{jt} + E x_{i,t-h} u_{i,t-h} u_{jt} x'_{jt}],
\]

Let

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

We first show the boundedness of \( \max_{ij} \| M_{u,ij} - V_{u,ij} \| \), then that of \( \max_{ij} \| S_{u,ij} - M_{u,ij} \| \).

**Proof of \( \max_{ij} \| M_{u,ij} - V_{u,ij} \| = O_P(L \sqrt{\frac{\log(LN)}{T}}) \)**
Given Lemma A.2, we have

\[
\max_{ij} \|M_{u,ij} - V_{u,ij}\| \leq O_P\left(\sqrt{\frac{\log N}{T}}\right) \\
+ 2 \max_{ij} \left\| \frac{1}{T} \sum_{h=1}^{L} \sum_{t=1}^{T} \left[ x_{it} u_{it} u_{t-h} x'_{j,t-h} f(t, h, L) - E x_{it} u_{it} u_{t-h} x'_{j,t-h} f(t, h, L) \right] \right\| \\
\leq O_P\left(\sqrt{\frac{\log N}{T}}\right) + O_P\left(\sqrt{\frac{\log(LN)}{T}}\right) = O_P\left(\sqrt{\frac{\log(LN)}{T}}\right). \quad \Box
\]

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

Given Lemma A.3, we have

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

Together,

\[
\max_{ij} \|V_{u,ij} - S_{u,ij}\| = O_P\left(\frac{L}{T} \sqrt{\frac{\log(LN)}{T}}\right) + O_P\left(\frac{L}{T} \sqrt{\frac{\log(LN)}{N}}\right) = O_P\left(\frac{L}{T} \sqrt{\frac{\log(LN)}{T}}\right). \quad \Box
\]
References

Abadie, A., S. Athey, G. W. Imbens, and J. Wooldridge (2017): “When should you adjust standard errors for clustering?,” Discussion paper, National Bureau of Economic Research.

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

Andrews, D. W. (1991): “Heteroskedasticity and autocorrelation consistent covariance matrix estimation,” Econometrica: Journal of the Econometric Society, 59(3), 817–858.

Angrist, J. D., and J.-S. Pischke (2008): Mostly harmless econometrics: An empiricist’s companion. Princeton university press.

Arellano, M. (1987): “Computing Robust Standard Errors for Within-groups Estimators,” Oxford bulletin of Economics and Statistics, 49(4), 431–434.

— (2003): Panel data econometrics. Oxford university press.

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

Bickel, P. J., and E. Levina (2008): “Covariance regularization by thresholding,” The Annals of Statistics, 36(6), 2577–2604.

Cameron, A. C., and D. L. Miller (2015): “A practitioners guide to cluster-robust inference,” Journal of Human Resources, 50(2), 317–372.

Cameron, A. C., and P. K. Trivedi (2005): Microeconometrics: methods and applications. Cambridge university press.

Driscoll, J. C., and A. C. Kraay (1998): “Consistent covariance matrix estimation with spatially dependent panel data,” Review of economics and statistics, 80(4), 549–560.

Fan, J., Y. Liao, and M. Mincheva (2013): “Large covariance estimation by thresholding principal orthogonal complements,” Journal of the Royal Statistical Society: Series B (Statistical Methodology), 75(4), 603–680.

Friedberg, L. (1998): “Did unilateral divorce raise divorce rates? Evidence from panel data,” American Economic Review, 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,” *Journal of Econometrics*, 141(2), 597–620.

Kim, D., and T. Oka (2014): “Divorce law reforms and divorce rates in the USA: An interactive fixed-effects approach,” *Journal of Applied Econometrics*, 29(2), 231–245.

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

Newey, W. K., and K. D. West (1987): “A simple, positive semi-definite, heteroskedasticity and autocorrelationconsistent covariance matrix,” *Econometrica: Journal of the Econometric Society*, 55, 703–708.

——— (1994): “Automatic lag selection in covariance matrix estimation,” *The Review of Economic Studies*, 61(4), 631–653.

Petersen, M. A. (2009): “Estimating standard errors in finance panel data sets: Comparing approaches,” *The Review of Financial Studies*, 22(1), 435–480.

Stock, J. H., and M. W. Watson (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,” *Journal of Econometrics*, 166(2), 303–319.

White, H. (1980): “A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity,” *Econometrica: Journal of the Econometric Society*, pp. 817–838.

Wolfers, J. (2006): “Did unilateral divorce laws raise divorce rates? A reconciliation and new results,” *American Economic Review*, 96(5), 1802–1820.

Wooldridge, J. M. (2003): “Cluster-sample methods in applied econometrics,” *American Economic Review*, 93(2), 133–138.

——— (2010): *Econometric analysis of cross section and panel data*. MIT press.