Robust Standard Errors to Spatial and Time Dependence in Aggregate Panel Models

Lucciano Villacorta

February 17, 2017

Abstract

This paper studies alternative approaches to consider time and spatial dependence in aggregate panel models where neither N nor T is very large, a problem that applied researchers often face when working with country- or state-level panel data. I show that the variance of the two-way cluster standard errors (2CCE) is affected by both types of dependence. Therefore, when neither N nor T is large, these standard errors could be poorly estimated and thus unreliable. As a consequence tests based on the 2CCE might perform poorly in terms of rejection rates, even in panels with a moderate sample size. I show that the cluster estimator can be expressed as a fully flexible panel version of the spatial autoregressive model. Then, it is possible to exploit that insight and use some prior knowledge about the dependence to estimate more parsimonious models that deliver more precise standard errors. In a calibrated Monte Carlo exercise using state minimum wage data, I show that using a panel version of the spatial model yields substantially better results than using the 2CCE when N and T are as small as 50 and 30. Finally, I study the implications of considering both types of dependence within a state-year panel data of wage inequality and minimum wages in the US. When both types of error dependence are considered, the marginal effect of the minimum wage over wage inequality is no longer significant.

Keywords: Panel, spatial dependence, time series dependence, two-way cluster, spatial autoregressive model.

JEL codes: C12; C13; C23

* Special thanks go to Manuel Arellano, Stephane Bonhomme and Christian Hansen. I would also like to thank Enrique Sentana, Ivan Fernandez-Val, Luca Repetto, Azeem M. Shaikh, Timothy Vogelsang and Jeffrey Wooldridge for useful comments and discussions. Errors and omissions are exclusively my own responsibility. Also, I appreciate comments and discussion with audiences at the CEMFI PhD Workshop, the Symposium of the Spanish Economic Association 2014, the Bristol Econometric Study Group (2014) and the Workshop at the Central Bank of Chile. Support from the European Research Council/ ERC grant "Estimation of non linear panel models with unobserved heterogeneity" (grant agreement n0 263107) is gratefully acknowledged.

†Central Bank of Chile. Email: lvillacorta@bcentral.cl, lucciano.villacorta@gmail.com

1 Introduction

Many empirical economic models exhibit error dependence. Though error dependence is not usually the object of interest, computing standard errors under error dependence has become a crucial part of applied work to conduct valid inference. For example, the HAC estimator of Newey & West (1987) is very popular in time series, while clustered standard errors are routinely used in microeconometric applications (Wooldridge 2003). In empirical models with panel data, the cluster covariance matrix estimator (CCE), introduced by Arellano (1987), is widely used to deal with general forms of serial (time series) dependence under spatial (or cross-sectional) independence. The CCE is quite appealing to applied researchers as it allows for unrestricted time series dependence and it is computationally easy to implement. However, computing standard errors that are robust to both time and spatial dependence is not common in applied econometrics.[1]

A novel approach to estimate standard errors robust to both types of dependence is the two-way cluster estimator (2CCE) recently proposed by Cameron et al. (2011) and Thompson (2011). This estimator extends the CCE to a set-up of contemporaneous spatial and time dependence by decomposing the total variance into three terms: (i) the standard CCE of Arellano (1987) for panel data that only contains the time correlation, which is estimated by clustering the data into spatial clusters (from now on $CCE_{spatial}$), (ii) another CCE that only contains the spatial correlations, which is estimated by clustering the data into time clusters (CCE_{time}) and (iii) a term that is an average of the variance of each observation. Thus, the 2CCE estimator is essentially the sum of two different CCE estimators minus the White formula. The main advantage of this approach is flexibility since it does not require any specification for the dependence structure. However, the (large N and large T) statistical properties of CCE estimators have only been studied for the time series dependence case (Hansen 2007*a*).[2] Furthermore, the performance of the 2CCE might be unsatisfactory in absence of large N and large T since a cluster estimator typically relies on the number of clusters going to

[1] For instance, Bertrand et al. (2004) and Petersen (2009) provide evidence of a wide range of empirical applications where the error term of a model could exhibit dependence in more than one dimension, which is usually not taken into account when computing standard errors. Moreover, Cameron et al. 2011 show examples where standard errors increased more than threefold when both types of error dependence are taken into account.

[2] (Hansen 2007*a*) studies the large N - large T properties of the $CCE_{spatial}$ under time series dependence but spatial independence.

infinity. This is not necessarily the case in aggregate panels such as state-year panels, industry-year panels, or country-year panels, where the number of states, industries, and countries are limited as is the availability of time series data. Moreover, in these settings spatial dependence is very likely to be present, as these units might be connected through neighboring effects, economic trade or linkage in production (e.g., Barrios et al. 2012 and, Foerster et al. 2008).[3]

This paper analyses, analytically and computationally, the 2CCE estimator with the aim to provide some insights about its behavior when N and T are not large. In the spirit of Hansen (2007*a*), I find that the 2CCE estimator, appropriately normalized, is consistent when both N and T go to infinity even with strong dependence in both the time series and the spatial dimension.[4] The first new result of this paper is to show that the variance of the 2CCE is affected by both types of dependence. As a consequence and since the 2CCE converges to its asymptotic distribution at the rate $min\left(\sqrt{N},\ \sqrt{T}\right)$, when neither N nor T is large, these standard errors could be poorly estimated and thus unreliable. Monte Carlo simulations indicate rejection rates of around 15% and 10% when $N, T = 10$ and $N, T = 20$ respectively for a nominal size of 5%. [5]

I then provide some insights on why the 2CCE, despite being a consistent estimator for the two-dimensional dependence setup, might perform poorly in terms of rejection rates, even with moderate sample size. The intuition can be given by noticing that each of the two CCEs that comprise the 2CCE is an average in one of the particular dimensions of the panel. For instance, the $CCE_{spatial}$ is an average that learns from the cross-sectional dimension. In the case of time series dependence but spatial independence, the $CCE_{spatial}$ uses independent cross-sectional observations to estimate a covariance structure that present serial correlation. When there is also spatial dependence, in addition to the time dependence, each cross sectional observation possess less information and this translates into a higher variance of the $CCE_{spatial}$. The same intuition applies to the CCE_{time} that uses the time dimension to estimate a covariance structure with spatial dependence. When there is

[3] Barrios et al. (2012) illustrate the importance of considering spatial dependence in empirical studies with state level explanatory variables. They show that the state minimum wage in the U.S, a variable commonly used in the study of wage inequality (Lee 1999, David et al. (2016)), presents considerable spatial dependence across states that share a border.

[4] As long as the moment conditions are bounded. For instance, this excludes unit roots process.

[5] This is in contrast to the case of only time series dependence, where the CCE performs quite well even when N is small (and for any value of T). In fact, Hansen (2007*a*) shows in simulations that t-test based on the $CCE_{spatial}$ has a rejection rate close to the nominal size of 5% even for $N = 10$.

time dependence, besides the spatial dependence, each of the time observations is less informative and this translates into a higher variance of the CCE_{time}. Therefore, the variance of the two terms that compose the 2CCE might be large even in samples of moderate size. As was pointed out by Thompson (2011) a higher variance of a standard error estimate increases the probability of type I error. This has important implications for applied research, as it means that the use of those standard errors could lead researchers to report spurious findings as significant.

The second result shows that the 2CCE estimator is numerically equal to the GMM estimator of a two-way error component model for the moment condition. This is remarkable since the two-way error component model assumes equicorrelation within clusters while the 2CCE allows for arbitrary correlation in each of the clusters. The latter provides a simple way to calculate a bootstrap version of the 2CCE.[6]

I then explore the usefulness of imposing more structure to the covariance matrix, modeling the error dependence. To do so, I build on a panel version of the spatial autoregressive model for the error (henceforth, SAR) including terms that take into account time dependence.[7] The SAR model, while very common in spatial econometrics, is less used in empirical applications as a way to compute standard errors, since specifying a spatial model is not always straightforward. In addition, if the model is misspecified, the properties of this approach are unknown. Nevertheless, in some empirical applications there exists some prior knowledge about the dependence that might be useful in order to compute more precise estimates of the standard errors.[8] In this sense, the SAR model provides a natural framework to use this information and consider dependence in different dimensions but in a parsimonious way. [9]

Given the structure of the SAR model as a system of equations, it is possible to make a connection between the cluster for spatial dependence and the panel version of the SAR model. In this paper, I

[6] In the spatial and time dependence panel set-up, the standard bootstrap is very difficult to implement since re-sampling from a distribution with dependence in two different dimensions is not straightforward.

[7] The underlying parameters of the panel SAR model can be estimated by quasi maximum likelihood as in Lee & Yu (2010). Alternatively, I consider a non linear least square estimation that is consistent and computationally efficient in the presence of individual- and time- specific effects in panels in which neither N nor T is very large.

[8] For example, an input-output table will provide evidence on which sectors are connected.

[9] For instance, as opposed to the 2CCE, in the SAR set-up it is possible to accommodate non-contemporaneous spatial dependence. There are extensions to the cluster set-up for considering non-contemporaneous spatial dependence. For instance, the multi-way cluster (Cameron et al. 2011) and the Driscoll and Kraay estimator allows for flexible forms of spatial and time dependence. However, as I show in the simulations, these approaches need even larger T.

show that the cluster can be expressed in terms of a fully flexible panel SAR model for the moment conditions. Thus, computing standard errors with a spatial panel model can be seen as a particular case of the cluster estimator, while a saturated version of this panel SAR model will result in the cluster estimator. As such, if the researcher has a prior over the dependence structure it is possible to achieve substantial gains in terms of precision and reduce the probability of type I error.[10] This connection provides a natural framework to start thinking about model selection for considering error dependence that should be explored in further research.

To explore the small sample performance of the 2CCE and the panel SAR model in a real application, I conduct a calibrated Monte Carlo study using a state-year panel model of minimum wages and wage inequality with $N = 50$ and $T = 30$. The Monte Carlo rejection rate of the 2CCE is 16% instead of the nominal size of 5%. This result indicates that the 2CCE is not a good approach to capture the dependence structure in a model like this. On the other hand, the panel SAR approach attains the nominal size of the test when the model is correctly specified and shows less variability than the 2CCE.

Finally, I study the implications of neglecting error dependence in US state-year panel data on wage inequality and minimum wages. This setup has been considered by Lee (1999) and David et al. (2016) to study the effect of changes in the state minimum wage on the state wage inequality. When both types of error dependence are considered, the marginal effect of the minimum wage over wage inequality is no longer significant. [11]

The rest of the paper is organized as follows. Section 2 presents the framework. Section 3 discusses the properties of the 2CCE estimator. Section 4 discusses the spatial and time autoregressive model. Section 5 shows Monte Carlo simulations with different simulated data generating processes to evaluate the performance of the approaches previously discussed and evaluates the implications of considering spatial and time dependence in a US state-year panel model of wage inequality and minimum wages. Section 6 concludes.

[10]Additionally, if the parametric model is correct, it is possible to estimate a feasible GLS which may have better small sample properties than the OLS estimator. This might be very useful in a setup in which N and T are not very large.

[11]David et al. (2016) open the puzzle that why did minimum wage changes have a significant effect when less than 10% of workers in the sample were paid at or below the minimum wage. Considering error dependence could help to reconcile the puzzle.

2 A panel model with spatial and time dependence

In this section I describe the panel model in a framework of contemporaneous spatial and time dependence.[12] Consider a linear panel regression model defined by:

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

where $i = 1, \ldots, N$ indexes individuals, $t = 1, \ldots, T$ indexes time and x_{it} is a vector of observable covariates that are assumed to be orthogonal to the unobservable error component u_{it}.[13]

The OLS estimator of β from equation 2.1 is defined as $\hat{\beta} = \left(\sum_i^N \sum_t^T x_{it}x_{it}'\right)^{-1} \left(\sum_i^N \sum_t^T x_{it}y_{it}\right)$. Under standard regularity conditions, $\sqrt{d_{NT}}\left(\hat{\beta} - \beta\right) \xrightarrow{d} \mathbb{N}\left(0, Q^{-1}WQ^{-1}\right)$, where $Q^{-1}WQ^{-1}$ is the asymptotic variance of the OLS estimator,

$$Q = lim_{N,T\to\infty} \frac{1}{NT}\left(\sum_i^N \sum_t^T E\left[x_{it}x_{it}'\right]\right) \tag{2.2}$$

and

$$W = lim_{N,T\to\infty}\left[\frac{1}{NT} Var\left(\sum_i^N \sum_t^T x_{it}'u_{it}\right)\right] \tag{2.3}$$

The speed of convergence $\sqrt{d_{NT}}$ and the analytic expression of the middle of the sandwich formula W depend on the correlation structure of u_{it} and x_{it}. In order to simplify notation, let us define $z_{it} = x_{it}'u_{it}$.

$$W = lim_{N,T\to\infty} \frac{1}{NT}\left[\sum_{i=1}^N Var\left(\sum_{t=1}^T z_{it}\right) + 2\sum_{i=1}^{N-1}\sum_{j=i+1}^{N} Cov\left(\sum_t^T z_{it}z_{jt}'\right)\right] \tag{2.4}$$

Assumption 1: Contemporaneous spatial and time dependence: $E\left(z_{it}z_{jt}'\right) \neq 0$, $E\left(z_{it}z_{is}'\right) \neq$

[12] For ease of exposition, I focus on the OLS estimator of a linear panel model, but the results can be extended to GMM, IV setting and non linear set-ups such as logit, probit or quantile regression.

[13] As in Hansen (2007*a*), the formulation in 2.1 can incorporate individual and time fixed effects. In this case y_{it}, x_{it}, and u_{it} should be interpreted as deviations from an individual and a time mean.

0 , $E\left(z'_{it}z_{js}\right) = 0$, $\forall i \neq j$ and $\forall t \neq s$. Assumption 1 implies unrestricted spatial dependence across individuals at the same time and unrestricted time dependence for each individual, but rules out dependence between different individuals at different moments in time. It is important to note that the dependence has been established on z_{it} rather than on u_{it}.[14]

Under assumption 1, the expression for W is the following:

$$W \;=\; lim_{N,T\to\infty} \quad \frac{1}{NT}\left[\sum_{i=1}^{N} Var\left(\sum_{t=1}^{T} z_{it}\right) + \sum_{t=1}^{T}\left(2\sum_{i=1}^{N-1}\sum_{j=i+1}^{N} Cov\left(z_{it}z'_{jt}\right)\right)\right] \tag{2.5}$$

Adding and subtracting $\sum_{i=1}^{N}\sum_{t=1}^{T} Var\left(z_{it}\right)$

$$W = lim_{N,T\to\infty}\left[\frac{1}{NT}\sum_{i=1}^{N} E\left(\Omega_i\right) + \frac{1}{NT}\sum_{t=1}^{T} E\left(\Omega_t\right) - \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} E\left(z_{it}z'_{it}\right)\right] \tag{2.6}$$

W has been split into three terms: (1) the term $\frac{1}{NT}\sum_{i=1}^{N} E\left(\Omega_i\right) = \frac{1}{N}\sum_{i=1}^{N}\left[\left(\frac{1}{T}\right)\sum_{t=1}^{T}\sum_{s=1}^{T} E\left(z_{it}z'_{is}\right)\right]$ that is an average in the cross-sectional dimension of all the accumulated time series correlation of each of the individuals, (2) the term $\frac{1}{NT}\sum_{t=1}^{T} E\left(\Omega_t\right) = \frac{1}{T}\sum_{t=1}^{T}\left[\left(\frac{1}{N}\right)\sum_{i=1}^{N}\sum_{j=1}^{N} E\left(z_{it}z'_{jt}\right)\right]$ that is an average in the time-series dimension of all the accumulated spatial correlation in a particular period, and (3) the term $\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} E\left(z_{it}z'_{it}\right)$ that is just the average of all the variances in both the cross-sectional and the time-series dimensions.

3 Analyzing the Two-Way Cluster Estimator (2CCE)

In the case of time series and contemporaneous spatial dependence, Cameron et al. (2011) and Thompson (2011) extend the cluster estimator (CCE) to a two-way cluster estimator (2CCE):

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

where

[14] The dependence in z_{it} is what matters in the formula of the variance of the OLS estimator. For instance, if the regressors are independent then W reduces to $\frac{1}{NT}\sum_{i}^{N}\sum_{t}^{T} E\left(u_{it}^2 x_{it}x'_{it}\right)$, the standard formula of the OLS variance, which can be estimated using the White formula. This is because $\mathbb{E}\left(u_{it}u_{is}x_{it}x'_{is}\right) = 0$ and $\mathbb{E}\left(u_{it}u_{jt}x_{it}x'_{jt}\right) = 0$.

$$\hat{W}_{2CCE} = \hat{W}_{CCE_{spatial}} + \hat{W}_{CCE_{time}} - \hat{W}_{white} \tag{3.1}$$

and $\hat{W}_{CCE_{spatial}} = \frac{1}{NT}\sum_i^N \hat{\Omega}_i$, $\hat{\Omega}_i = \sum_t^T \sum_s^T \left(\hat{z}_{it}\hat{z}_{is}'\right)$,$\hat{W}_{CCE_{time}} = \frac{1}{NT}\sum_t^T \hat{\Omega}_t$, $\hat{\Omega}_t = \sum_i^N \sum_j^N \left(\hat{z}_{it}\hat{z}_{jt}'\right)$, $\hat{W}_{white} = \frac{1}{NT}\sum_i^N \sum_t^T \hat{z}_{it}\hat{z}_{it}'$, $\hat{z}_{it} = x_{it}\hat{u}_{it}$ and $\hat{u}_{it}$ are the OLS residuals from 2.1. The 2CCE is essentially the sum of two different CCE minus the White formula for heteroskedastic errors.

The purpose of this section is to analyze $\hat{W}_{2CCE}$ and give some insights about its small sample behavior. To do so, I separately analyze the weak and strong dependence cases.

3.1 The weak dependence case

The weak dependence case is when both the time series dependence and the contemporaneous spatial dependence decay with some measure of distance. For instance, the former might be characterized by a stationary autorregresive model (AR), and the latter by a spatial autoregressive model (SAR) for a well defined weighting matrix.

The $\hat{W}_{2CCE}$ is consistent as long as each of its components are consistent. Hence, it is necessary to establish the conditions in which $\hat{W}_{CCE_{spatial}}$, $\hat{W}_{CCE_{time}}$ and $\hat{W}_{white}$ are consistent when there is contemporaneous spatial and time dependence.

3.1.1 Analyzing $\hat{W}_{CCE_{spatial}}$:

The term $\hat{W}_{CCE_{spatial}}$ is exactly the CCE estimator of $\frac{1}{NT}\sum_i^N E\left(\Omega_i\right)$ proposed by Arellano (1987) to consider any kind of time series dependence in a fixed T - large N panel model under cross-sectional independence. Hansen (2007*a*) studies the asymptotic properties of this CCE when T goes to infinity jointly with N in the case of time series dependence, but cross sectional independence. He finds that the CCE estimator, appropriately normalized (divided by T), is consistent without imposing any condition on the growth rate of T relative to N, no matter the amount of time series dependence allowed between the observations within each individual.[15] What is key is that $\hat{W}_{CCE_{spatial}}$ is essentially an average that learns from the cross-sectional dimension, hence it is consistent as long as N goes to infinity.

[15] The amount of time series dependence affects the behavior of $\hat{\beta}$. With weak dependence (mixing) the rate of convergence is $\sqrt{NT}$, whereas under strong dependence (non-mixing) the growth rate is $\sqrt{N}$.

In this paper, I extend Hansen (2007*a*)to analyze $\hat{W}_{CCE_{spatial}}$ when there is both time series and spatial dependence. In the case of weak spatial dependence it is possible to extend Theorem 3 of Hansen (2007*a*). [16]

$$\sqrt{N}\left[\hat{W}_{CCE_{spatial}}-\frac{1}{NT}\sum_{i=1}^{N}E\left(\Omega_i\right)\right] \overset{d}{\to} N\left(0,V_{CCE_{spatial}}\right) \tag{3.2}$$

where

$$V_{CCE_{spatial}} = lim_{N,T\to\infty}\frac{1}{NT^2}\left(\sum_{i=1}^{N}Var\left(\Omega_i\right)+2\sum_{i=1}^{N-1}\sum_{j=1}^{N}Cov\left(\Omega_i,\Omega_j\right)\right) \tag{3.3}$$

Remark : *Note that 3.2-3.3 are similar to theorem 3 in Hansen (2007a) but with the inclusion of the term* $2\sum_{i=1}^{N-1}\sum_{j=1}^{N}Cov\left(\Omega_i,\Omega_j\right)$ *in* $V_{CCE,i}$. *In the two-dimensional dependence set-up* $\hat{W}_{CCE_{spatial}}$ *has a higher variance than in Hansen (2007a). Since* $\hat{W}_{CCE_{spatial}}$ *essentially learns from the cross-sectional dimension and this is less informative due to the spatial dependence,* $\hat{W}_{CCE_{spatial}}$ *is estimated with less precision. Given that* $\hat{W}_{CCE_{spatial}}$ *is a component of* $\hat{W}_{2CCE}$, *this result shed light on the behavior of the 2CCE when N is not large.*

3.1.2 Analyzing $\hat{W}_{CCE_{time}}$:

The term $\hat{W}_{CCE_{time}}$ is estimated by grouping the data in time clusters. This estimator is analogous to the estimator described above but for the case of cross sectional dependence and time series independence. Under time series independence, this estimator is consistent for any kind of cross-sectional dependence for fixed N or large N as long as T goes to infinity.[17] This estimator is essentially an average that learns from the time series dimension, hence it is consistent as long as T goes to infinity.

As in the previous case, it is possible to derive the asymptotic distribution of $\hat{W}_{CCE,t}$ under weak

[16] Since the cross-sectional dependence is weak, the term $lim_{N,T\to\infty}\frac{1}{NT^2}\left(\sum_{i=1}^{N-1}\sum_{j=1}^{N}Cov\left(\Omega_i,\Omega_j\right)\right)$ is bounded, and it is possible to apply a central limit theorem for spatial dependence as in Conley (1999).

[17] The argument follows the analysis in Hansen (2007*a*) but for dependence in the other dimension. In the large N case, in order to derive its asymptotic distribution, the $\hat{W}_{CCE,t}$ have to be normalized, (divided by N).

spatial and time dependence. [18]

$$\sqrt{T}\left[\hat{W}_{CCE_{time}}-\frac{1}{NT}\sum_{t=1}^{T}E\left(\Omega_{t}\right)\right]\overset{d}{\rightarrow}N\left(0,V_{CCE_{time}}\right)\tag{3.4}$$

where

$$V_{CCE_{time}} = lim_{N,T\rightarrow\infty}\frac{1}{N^{2}T}\left(\sum_{t=1}^{T}Var\left(\Omega_{t}\right)+2\sum_{t=1}^{T-1}\sum_{s=1}^{T}Cov\left(\Omega_{t},\Omega_{s}\right)\right)\tag{3.5}$$

Remark : *The variance $V_{CCE_{time}}$ is augmented by the time series dependence $2\sum_{t=1}^{T-1}\sum_{s=1}^{T}Cov\left(\Omega_{t},\Omega_{s}\right)$ as opposed to case where the time dimension is independent. Since $\hat{W}_{CCE_{time}}$ essentially learns from the time series dimension and this is less informative due to the time series dependence, $\hat{W}_{CCE_{time}}$ is estimated with less precision. Because $\hat{W}_{CCE_{time}}$ is a component of $\hat{W}_{2CCE}$, this result shed light on the behavior of the 2CCE when T is not large.*

3.1.3 $\hat{W}_{2CCE}$ under weak contemporaneous spatial and time dependence.

Combining 3.1, 3.2, 3.3, 3.4 and 3.5 is possible to derive the distribution of the $\hat{W}_{2CCE}$ under weak contemporaneous spatial and time dependence.[19]

$$\sqrt{min\left(N,T\right)}\left[\hat{W}_{2CCE}-W\right]\overset{d}{\rightarrow}N\left(0,\ V\right)\tag{3.6}$$

$$\begin{aligned} V &= lim_{N,T\rightarrow\infty}\frac{1}{max(N,T)}[N\times V_{CCE_{spatial}}+T\times V_{CCE_{time}}+V_{White}\\ &+ 2Cov\left(\hat{W}_{CCE_{spatial}},\hat{W}_{CCE_{time}}\right)-2Cov\left(\hat{W}_{CCE_{spatial}},\hat{W}_{White}\right)-2Cov\left(\hat{W}_{CCE_{time}},\hat{W}_{White}\right)] \end{aligned}\tag{3.7}$$

Remark : *$\hat{W}_{2CCE}$ needs that both N and T go to infinity, since $\hat{W}_{CCE_{spatial}}$ and $\hat{W}_{CCE_{time}}$ are consis-*

[18] Since the time series dependence is weak, the term $lim_{N,T\rightarrow\infty}\frac{1}{N^{2}T}\left(\sum_{t=1}^{T-1}\sum_{s=1}^{T}Cov\left(\Omega_{t},\Omega_{s}\right)\right)$ is bounded, and it is possible to apply a central limit theorem for time series dependence.

[19] Since the white formula $\hat{W}_{white}$ is averaging in both dimensions, the estimator is $\sqrt{NT}$- consistent under weak dependence.

tent when N goes to infinity and when T goes to infinity respectively.

Remark : *The cross sectional dependence increase the variance of* $\hat{W}_{CCE_{spatial}}$ *, which might be non negligible if N is not very large. On the other hand, the time series dependence increases the variance of* $\hat{W}_{CCE_{time}}$ *, which might be non negligible when T is not very large. As a consequence, when T and N are not large enough,* $\hat{W}_{2CCE}$ *might suffer from a considerable finite sample variance and inference based on the Normal approximation using* $\left(\frac{1}{NT}\sum_i^N\sum_t^T x_{it}x_{it}'\right)^{-1}\hat{W}_{2CCE}\left(\frac{1}{NT}\sum_i^N\sum_t^T x_{it}x_{it}'\right)^{-1}$ *as the variance estimator of* $\hat{\beta}$ *may be misleading.* [20]

These two points might explain why tests based in the 2CCE might perform poorly in terms of rejection rates, even in panels with a moderate sample size as is shown in section 5.[21] This is in contrast to the one-dimensional dependence set-up, where test based on *CCE* performs quite well even when N is small (and for any value of T). [22]

3.2 The strong dependence case

The strong dependence case is when neither the time series dependence nor the contemporaneous spatial dependence decay with some measure of distance. A common example of this setting is a two-way error component model for the moment condition $z_{it} = x_{it}u_{it}$.[23]

$$z_{it} = \alpha_i + \delta_t + \varepsilon_{it} \tag{3.8}$$

where α_i is independent across individuals i with $Var(\alpha_i) = \sigma_\alpha$, δ_t is independent across t with $Var(\delta_t) = \sigma_\delta$, and ε_{it} is i.i.d with $Var(\varepsilon_{it}) = \sigma$.

The model in 3.8 has a equicorrelated time series dependence that does not vanish, even for observations that are far away. It also presents a equicorrelated spatial dependence that does not vanish. This model has been widely studied (see for example Amemiya 1971).

[20] This problem might be worsened if the original model presents individual- and time- specific effects as controls for unobserved heterogeneity, given that the demeaning induces additional dependence in both dimensions. This additional dependence is strong when N and T are not large but disappears when N and T go to infinity.

[21] In fact, Thompson (2011) shows in simulations that the 2CCE has a standard deviation that is more than the double of the standard deviation of the White formula and 20% more than the CCE. He also shows, using Jensen's inequality, that less precision in the variance estimator translates into a higher expected value of the t-statistic and therefore a higher probability of type I error.

[22] Hansen (2007*a*) shows in simulations that t-test based on the CCE_i has a rejection rate close to the nominal size of 5% even for $N = 10$.

[23] For the sake of simplicity I will consider a scalar z_{it}, but the analysis can be easily extended to a vector of covariates.

$$E\left(z_{it}z_{js}\right) = \begin{cases} \sigma_\alpha & i = j, t \neq s \\ \sigma_\delta & i \neq j, t = s \\ \sigma_\alpha + \sigma_\delta + \sigma & i = j, t = s \\ 0 & i \neq j, t \neq s \end{cases} \tag{3.9}$$

The idea of this subsection is to see whether the 2CCE estimator is consistent for an error dependence structure as in 3.8. To do so, I will follow a different approach than in subsection 3.1. In particular, I show that the 2CCE estimator is numerically the same as the GMM estimator of the two-way error component model for z_{it} in (3.2).

Under the dependence structure in (3.2), the middle of the sandwich formula in (2.2) takes the form of:

$$W = lim_{N,T\rightarrow\infty} T\sigma_\alpha + N\sigma_\delta + \sigma \tag{3.10}$$

Defining:

$$\bar{z}_i \equiv \frac{1}{T}\sum_{t=1}^{T} z_{it} = \alpha_i + \frac{1}{T}\sum_{t=1}^{T}\delta_t + \frac{1}{T}\sum_{t=1}^{T}\varepsilon_{it} \tag{3.11}$$

$$\bar{z}_t \equiv \frac{1}{N}\sum_{i=1}^{N} z_{it} = \frac{1}{N}\sum_{i=1}^{N}\alpha_i + \delta_t + \frac{1}{N}\sum_{t=1}^{T}\varepsilon_{it} \tag{3.12}$$

Then

$$Var\left(\bar{z}_i\right) = \sigma_\alpha + \frac{1}{T}\sigma_\delta + \frac{1}{T}\sigma \tag{3.13}$$

$$Var\left(\bar{z}_t\right) = \frac{1}{N}\sigma_\alpha + \sigma_\delta + \frac{1}{N}\sigma \tag{3.14}$$

$$Var\left(z_{it}\right) = T\sigma_\alpha + N\sigma_\delta + \sigma \tag{3.15}$$

Combining equations 3.10, 3.13, 3.14 and 3.15

$$W = lim_{N,T\to\infty} TVar(\bar{z}_i) + NVar(\bar{z}_i) - Var(z_{it}) \tag{3.16}$$

A natural plug-in (GMM) estimator of W in this two-way error component model is to replace each population variance by their sample analogues:

$$\hat{W} = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T} z_{it}z_{is}\right) + \frac{1}{T}\sum_{t=1}^{T}\left(\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{N} z_{it}z_{jt}\right) - \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} z_{it}^2 \tag{3.17}$$

which is exactly the 2CCE estimator in 3.1. The following conclusions emerge from this result.

Remark: *Since the 2CCE is numerically equal to the GMM estimator of the two way error component model, we can conclude that the 2CCE is consistent even with strong dependence in both the spatial and time series dimensions, as long as z_{it} has bounded moments. The asymptotic properties of the GMM estimator of the two-way error component model are studied in Amemiya (1971).*

Remark: *Imposing within-cluster equicorrelation in each of the dimensions is essentially innocuous for the purpose of calculating cluster-robust standard errors to contemporaneous spatial and time dependence.*

Remark: *This equality provides a simple way of computing a bootstrap version of the 2CCE variance estimator by re-sampling the individual effects, the time effects and the i.i.d term from a two-way error component model estimated for the moment condition z_{it}. The 2CCE computed using the bootstrap delivers a consistent estimator of the variance even when the spatial and time dependence do not come from a structure like the two-way error component model.* [24] *This procedure would be very useful, for instance, in complicated non-linear models where an analytical expression for the variance is not available, considering that the standard block bootstrap of Cameron & Miller (2015) is very difficult to implement in a two-dimensional dependence set-up since there is not a straightforward way of defining blocks under spatial and time dependence.*

[24] This procedure is valid for any type of time series and spatial dependence but only for models without individuals and time effects in the specification of $E(y_{it} \mid x_{it})$ as controls for unobserved heterogeneity. The inclusion of those effects in the model of $E(y_{it} \mid x_{it})$ prevents the estimation of a two-way error component model for z_{it}.

4 A parametric model for spatial and time dependence:

In this section I consider a model for the error term that allows for spatial and time series dependence. The model is a panel version of the spatial autoregressive model (SAR), commonly used in spatial econometrics (Anselin 1988, 2001, Baltagi et al. 2003, 2007, Baltagi & Pirotte 2010, Kelejian & Prucha 2004, 2010, Lee & Yu 2010, Kelejian & Prucha 1998), and includes lags as in the standard time-series autoregressive model (AR) . The aim of this section is twofold. On the one hand, I present an alternative way of considering error dependence and show how to estimate the parameters of this model in order to compute robust standard errors. On the other hand, I make a connection between the panel version of the SAR model and the cluster estimator.

Consider the following panel version of a spatial and time autorregresive model (SAR-TS):

$$u_{it} = \lambda \sum_{j=1}^{N} w_{ij} u_{jt} + \rho_1 u_{it-1} + \rho_2 \sum_{j=1}^{N} w_{ij} u_{jt-1} + \varepsilon_{it} \tag{4.1}$$

or in matrix form:

$$U_t = \lambda W_N U_t + \rho_1 U_{t-1} + \rho_2 W_N U_{t-1} + \epsilon_t \tag{4.2}$$

where U_t is a $Nx1$ vector which contains each of the shocks u_{it} for the N observations in time $t = 1....T$, ϵ_t is a $Nx1$ vector of innovations and W_N is a NxN matrix which is determined by the researcher before the estimation.[25] Each element w_{ij} of W_N that is different from zero allows for direct spatial correlation between two different cross sectional units. Typically w_{ij} is nonzero when the unit i and the unit j are not so far in some "measure of distance" but tend to zero when the distance between units increases.[26]

[25]This assumption is crucial in the cross-sectional setting, since it is otherwise impossible to identify the NxN elements of W_N. In the large T - fixed N panel setup it is possible to use the time series dimension to estimate the elements of W_N under some restrictions defined before the estimation as in the case of a Structural Vector Autoregressive model. In the large N - large T data Manresa (2016)) shows that it possible to recover the structure of the interactions by using a panel version of the Lasso estimator under a sparsity assumption. Her spillover model can be interpreted as a particular spatial model, where the dependence comes from the covariates of the model (i.e $Y(t) = \beta X(t) + \lambda W_N X(t) + \varepsilon(t)$) and the elements of W_N are not specified. In this setup there is no identification problem since $E(\varepsilon(t) X(t)) = 0$ so she can recover the elements of W_N by OLS, assuming stationarity in the spillover process (i.e W_N does not change over time) and sparcity of the model (some elements of W_N are zero).

[26]This assumption is crucial to have a finite variance-covariance matrix for the asymptotic distribution of the OLS estimator when N goes to infinity, like an ergodicity property for a time series process.

The model in (4.2) is a parsimonious way to allow for contemporaneous spatial and time series dependence but also allows for non-contemporaneous spatial dependence.[27] Therefore, the 2CCE estimator discussed in the previous section would not be a consistent estimator of the variance under the dependence structure implied by (4.2), as it is shown in the Monte Carlo simulation of section 5.[28]

An important assumption in (4.2) is that the spatial dependence implied by λW_N is constant over time. This assumption enables a $\sqrt{NT}$- consistent estimator of the variance that is crucial for setups where N and T are not that large separately but NT is, as in some aggregate panel models. For instance, the latter assumption might be sensible when the measure of distance implied in W_N is, for instance, geographical distance or the dependence structure has reached a stationary level (see Manresa (2016)).

For instance, if the cross sectional units were states, it might be sensible to allow for direct connection between neighboring states and indirect connection between non-neighboring states through common neighbors (Barrios et al. 2012). This kind of dependence can be accommodated in (4.2) by letting $w_{ij} = 1$ if state i and state j share a border and 0 otherwise.

4.1 Estimation

There are several approaches to estimate model (4.2), see for instance, (Kelejian & Prucha 1998, 2004, 2010, Lee & Yu 2010, Yu et al. 2008). One way to estimate $\theta = [\lambda, \rho_1, \rho_2]$ is to use a conditional pseudo maximum likelihood (QML) approach as if it were the reduced form of a vector autoregressive model.[29]However, if the original model presents individual and time specific fixed effects (α_i and δ_t) and neither N or T are large, the QML estimator using the estimated residual might be affected by an incidental parameter problem in both dimension as N are T are not sufficiently large. When N and T are not large, the subtractions of the time and group means from the data to eliminate α_i and δ_t will alter the variance structure of the data and, as a result, the estimator of θ will be biased.

Non-linear least squares. Another way to estimate the parameters in θ is to use all the moment conditions inside the variance-covariance matrix of the error term of the model. The $NT \times NT$

[27]Note that a model where $\rho_2 = 0$ also allows for non-contemporaneous spatial dependence by the interaction between w_{ij} and ρ_1.

[28]A multi-way cluster for non-contemporaneous spatial dependence might be consistent. However as is shown in section 5, the multi-way cluster also performs poorly for moderate N and T.

[29]The properties of consistency and asymptotic normality of the estimator will be satisfied as shown in Yu et al. (2008).

variance-covariance matrix of the error term is :

$$\Upsilon = E\begin{bmatrix} U_1U_1' & U_1U_2' & \cdots\cdots\cdots & U_1U_T' \\ U_2U_1' & U_2U_2' & \cdots\cdots\cdots & U_2U_T' \\ \vdots & \cdots\cdots\cdots & \cdots\cdots\cdots & \vdots \\ \vdots & \cdots\cdots\cdots & \cdots\cdots\cdots & \vdots \\ \vdots & \vdots & \vdots & \vdots \\ U_TU_1' & U_TU_2' & \cdots\cdots\cdots & U_TU_T' \end{bmatrix} \tag{4.3}$$

Defining $Q_1 = I_N \otimes (I_T - \frac{1}{T}\iota_T\iota_T\prime)$ as the matrix which demeans the variables with respect to their time mean and $Q_2 = (I_N - \frac{1}{N}\iota_N\iota_N\prime) \otimes I_T)$ as the matrix which demeans the variables with respect to their state mean, I transform the data using Q_1 and Q_2 in order to eliminate the time-specific and the state- specific effects of the regression.

Using the residuals that come from the transformed model, I can minimize the quadratic distance between all the moment conditions inside $Q_2Q_1\Upsilon(\theta)Q_1'Q_2'$ and the sample counterpart of $E[\hat{U}^*\hat{U}^{*\prime}]$, where $\hat{U}^*$ is a $NT \times 1$ vector that contains the estimated residuals of the transformed model. Defining $\zeta_p = vech(\hat{U}^*\hat{U}^{*\prime})$, and $\xi_p(\theta) = vech(Q_2Q_1\Upsilon(\theta)Q_1'Q_2')$, I can estimate θ as the nonlinear least squares from the following model: $\zeta_p = \xi_p(\theta) + \upsilon_p$, where $p = NT * \frac{(NT+1)}{2}$.[30]

$$\widetilde{\theta^{NLS}} = \underset{c}{argmin}\{[\zeta_p - \xi_p(\theta)]'[\zeta_p - \xi_p(\theta)]\} \tag{4.4}$$

Estimating the model in this way respects the real dependence structure of the data, as opposed to QML, therefore $\widetilde{\theta^{NLS}}$ does not suffer from the bias affecting the QML. Intuitively, the estimator is explicitly considering the Nickell biases in both dimensions when estimating θ as in Hansen (2007*b*) for the time series dependence case.

[30]The estimator is inefficient because it is not using a correct weighting matrix for the moments in $\Upsilon(\theta)$.For simplicity I am using the identity matrix as a weighting matrix instead $Var(\zeta_p)$. For use the correct matrix I should know the fourth moments of U_{it}.

4.2 Cluster estimation as a flexible model of spatial dependence.

The purpose of this subsection is to show that there is a connection between the panel version of the SAR model and the CCE. The CCE can be expressed in terms of a fully flexible panel version of the SAR model for the moment condition $z_{it} = u_{it} x_{it}$. Therefore, if the researcher has some prior knowledge about the dependence structure, she could estimate a more parsimonious SAR model than the one implied by the CCE and compute more precise estimates of the standard errors. Also, expressing the CCE as a fully flexible SAR model give some guidance on how to use model selection for estimating more precise standard errors than the cluster.

To make this connection, consider the cluster estimator that takes into account the spatial dependence studied in section 3.1.2:[31]

$$\hat{W}_{CCE_{time}} = \frac{1}{NT} \sum_{t}^{T} \left(\sum_{i}^{N} \sum_{j}^{N} \left(z_{it} z_{jt} \right) \right)$$

or

$$\hat{W}_{CCE_{time}} = \frac{1}{N} \left(\frac{1}{T} \sum_{t=1}^{T} z_{1t}^{2} + \frac{1}{T} \sum_{t=1}^{T} z_{1t} z_{2t} + \cdots \frac{1}{T} \sum_{t=1}^{T} z_{N-1t} z_{Nt} + \frac{1}{T} \sum_{t=1}^{T} z_{Nt}^{2} \right) \tag{4.5}$$

The cluster estimator robust to spatial dependence can be seen as the sum of each of the sample analogue estimators of the $N \times N$ spatial covariances in:

$$\begin{bmatrix} E\left(z_{1t}^{2}\right) = \rho_{1,1} \\ E\left(z_{1t} z_{2t}\right) = \rho_{1,2} \\ \vdots \\ E\left(z_{1t} z_{Nt}\right) = \rho_{1,N} \\ \vdots \\ E\left(z_{Nt}^{2}\right) = \rho_{N,N} \end{bmatrix} \tag{4.6}$$

The CCE_t uses the information in the time series to estimate each of the moment conditions in 4.6, and is consistent even if the spatial correlation between two observations is not stable over time,

[31] For the sake of simplicity let us consider when z_{it} is scalar and observable.

(i.e $E(z_{1t}z_{2t}) \neq E(z_{1s}z_{2s})$ for $t \neq s$). Remember that $\hat{W}_{CCE_{time}}$ is $\sqrt{T}$−consistent.

It is possible to accommodate all the dependence structure considered in the CCE_{time} by using a panel version of the SAR model. To do so, lets group all the individual values of z_{it} in a $N \times 1$ vector $Z_t = \begin{bmatrix} z_{1t} & z_{2t} & \cdots & z_{Nt} \end{bmatrix}'$ and define the following linear equation:

$$Z_t = \Gamma^{1/2}\epsilon_t \tag{4.7}$$

where Γ is an $N \times N$ symmetric matrix that contains the $N \times N$ spatial correlations of 4.6 and ϵ_t is a $N \times 1$ vector of innovations with $E[\epsilon_t] = 0$, $E\left[\epsilon_t\epsilon_t'\right] = I_N$. Then it is possible to express 4.7 as:

$$Z_t = \Lambda Z_t + \epsilon_t \tag{4.8}$$

where $\Lambda = I_N - \Gamma^{-1/2}$ is a NxN matrix.

Let us define λ_{ij} as the ij element of Λ and W_{ij}^{CCE} as a NxN matrix in which only the element ij is equal to one and the other elements are zero. Then we can reformulate 4.8 in terms of λ_{ij} and W_{ij}^{CCE}

$$Z_t = \sum_{i=1}^{N-1}\sum_{j=i+1}^{N} \lambda_{ij} W_{ij}^{CCE} Z_t + \epsilon_t \tag{4.9}$$

Equation 4.9 is a fully flexible panel version of a SAR model of order NxN with spatial matrices W_{ij}^{CCE} and spatial parameters λ_{ij}. Using the time series dimension of the panel to estimate the NxN parameters λ_{ij} and then use the structure of the model in 4.9 for computing standard errors is numerically the same as computing the cluster estimator CCE_{time}. However if the researcher possess some prior information about the dependence she might estimate a parsimonious specification of 4.9. For instance, if the cross sectional units are states and the researcher believes that (i) the source of spatial dependence comes from states that share a border and (ii) the intensity of the spatial dependence are stable across states, she can estimate a restricted version of 4.9:

$$Z_t = \lambda W Z_t + \epsilon_t \tag{4.10}$$

where the element w_{ij} of W is one if state i and j share a border and zero otherwise.

It is also possible to capture the unrestricted spatial dependence implied by 4.9 with other sequences of weighting matrices that the researcher might think are sensible for the empirical application. To see this, let us define the following $SAR(K)$ model:

$$Z_t = \sum_{k=1}^{K(N)} \lambda_k W_k Z_t + \epsilon_t \tag{4.11}$$

where $K(N)$ is the maximum number of spatial lags and is a function of the sample size N and $\{W_1, \cdots\cdots, W_K\}$ is a sequence of weighting matrices. It is possible to mimic the model 4.9 and capture all the dependence implied by the matrices W_{ij}^{CCE} in 4.9 by saturating the $SAR(K)$ in 4.11 with $K = NxN$ and a suitable selection of independent $\{W_1, \cdots\cdots, W_K\}$.[32] If $K < NxN$, but grows with N, when $N \to \infty$, the model in 4.11 will be asymptotically equivalent to the 4.9 and, therefore, to the cluster CCE_t, which is badly estimated for T not very large. However, if K is fixed the estimator of the variance may learn from the time series and the cross section.

The model in 4.11 provides a natural framework for thinking about how to do model selection for error dependence. In many empirical applications it is sensible to think that just a finite number of weighting matrices (measures of distance) generate all the dependence structure. Therefore, in those settings we can use, for instance, the Lasso estimator to select a small number of weighting matrices in a model like 4.11. The latter is an idea that should be explored in further research.

4.3 Other alternatives to handle error dependence in both dimensions

There are other ways of conducting inference in models with error dependence that do not rely on the limiting Normal approximation. For example, in a recent paper, Bester et al. (2011) proposed a test based on the cluster estimator when the number of clusters is small. Rather than using the normal approximation, they derive a limiting distribution for the t-statistic, treating the cluster estimator as a random variable. On the other hand, Ibragimov & Müller (2010) propose to split the data into groups and construct a t statistic using as the numerator the average across groups of the OLS estimates in each group and as the denominator the standard deviation of the estimates across groups.

However, both approaches need (i) to group the date in such a way that the groups are approximately independent which is not straightforward in a two-dimensional dependence set-up, and (ii)

[32] The spatial dependence can be summarized in K independent underlying weighting matrices that generate all the spatial dependence in the model. Some examples might be weighting matrices that capture Euclidean distances, measures of economic distance like trade, etc.

need that the number of elements in each cluster is large relative to the number of clusters, while I am focusing on settings where N are T are of similar size and neither of the two sample sizes is large.

5 Simulation examples

In this section, I provide evidence on the inference properties of tests based on the approaches previously discussed (the $CCE_{spatial}$, CCE_{time}, $2CCE$ and the panel SAR model), first using simulation experiments in entirely simulated data, and then for an experiment calibrated to a model of U.S. state inequality and minimum wage. The latter experiments are conducted in a panel data setting with $N = 50$ and $T = 31$ where time and state-level fixed effects are included. In these simulations, we consider inference about a slope coefficient from a linear regression model with point estimates obtained using OLS. For the $CCE_{spatial}$, CCE_{time} and $2CCE$ estimators, a $t_{(N-1)}$, $t_{(T-1)}$ and $t_{(min(N,T)-1)}$ distributions are used respectively as the reference distribution.

5.1 Simulated data

All the DGPs used in the simulation exercises come from the following single framework, where the dependence of the error term u_{it} (and the regressor x_{it}) has the subsequent structure:

$$y_{it} = \beta x_{it} + \tilde{u_{it}} \tag{5.1}$$

$$\tilde{u_{it}} = u_{it} + \alpha_i + \delta_t \tag{5.2}$$

$$u_{it} = \lambda_1 \sum_{j=1}^{N} w_{ij} u_{jt} + \rho u_{it-1} + \varepsilon_{it} + v_{it} \tag{5.3}$$

$$v_{it} = \lambda_2 \sum_{j=1}^{N} w_{ij} v_{jt} + \epsilon_{it} \tag{5.4}$$

$$\varepsilon_{it} = \rho_2 \varepsilon_{it-1} + e_{it} \tag{5.5}$$

$$\epsilon_{it} \sim N(0,1) \tag{5.6}$$

$$\upsilon_{it} \sim N(0,1) \tag{5.7}$$

where $\lambda_2 < 1$, $\rho_2 < 1$ $\rho + \lambda \leq 1$ to ensure stationarity (see Lee & Yu 2010). For all the simulations the elements of the weighting matrix (w_{ij}) were row normalized to add up to one.

This is a general model that accommodates different spatial and time dependence regimes. I evaluate different regimes of dependence by working with some specific scenarios of 5.3. For the cases 1 to 3, I include individual fixed effects and time fixed effects in the main regression as controls for unobserved heterogeneity in both dimensions as do most difference in difference papers that use panel data (i.e $y_{it} = \beta x_{it} + \alpha_i + \delta_t + u_{it}$). The goal in these scenarios is focusing on the case where the error of the model u_{it} do not present strong dependences. In case 4, the main equation does not consider neither individual fixed effects nor time fixed effects (i.e $y_{it} = \beta x_{it} + \tilde{u_{it}}$), therefore the error term $\tilde{u_{it}} = u_{it} + \alpha_i + \delta_t$ will presents strong dependence.

Case 1: Additive errors

In this particular specification $\lambda_1 = 0$ and $\rho = 0$. Therefore, the model has an additive error ($u_{it} = \varepsilon_{it} + v_{it}$) that combines a source of time dependence, the AR (1) in $\varepsilon_{it} = \rho_2 \varepsilon_{it-1} + e_{it}$, and a source of contemporaneous spatial dependence, the SAR (1) in $v_{it} = \lambda_2 \sum_{j=1}^{N} w_{ij} v_{jt} + \epsilon_{it}$. Under this specification the 2CCE is a consistent estimator of the variance, thus this is the baseline model for evaluating the performance of the 2CCE in small samples.

Case 2: Time and spatial autoregressive model (TS-SAR)

In this particular specification $\lambda_2 = 0$ and $\rho_2 = 0$. In this model we have a unique process for the error term that allows for time series dependence through an AR(1) component and for spatial dependence through a spatial autoregressive component: $u_{it} = \lambda_1 \sum_{j=1}^{N} w_{ij} u_{jt} + \rho u_{it-1} + \xi_{it}$ where $\xi_{it} = \epsilon_{it} + \upsilon_{it}$. This is the model that has been discussed in the previous section.

As opposed to the DGP in case 1, this DGP allows for non-contemporaneous spatial dependence through the combination of λ, ρ, and w_{ij}. Therefore, the 2CCE will not be consistent. Still, I simulate a model like this to see how the 2CCE will perform for this class of processes. I also check results with an extension to the 2CCE that can capture non-contemporaneous spatial dependence.

Case 3: Autoregressive model and spatial moving average

In this specification $\lambda_1 = 0$ and $\rho_2 = 0$. This model has a unique process for the error term that allows for time series dependence through an AR(1) component and for spatial dependence through a spatial moving average component (SMA): $u_{it} = \rho u_{it-1} + \lambda_2 \sum_{j=1}^{N} w_{ij} v_{jt} + \epsilon_{it}$
What is particularly useful in a setup where the spatial dependence comes from a spatial moving average its that we can get rid of the time series dependence by a quasi difference. In this sense, if the researcher does not have a clear idea of the weighting matrix that governs the spatial dependence, she can remove the time series dependence by a quasi-difference and then cluster by time to account for the spatial dependence. The latter is particularly useful for setups where T is not that large and the CCE estimator that takes into account the spatial dependence is poorly estimated because it is also affected by the time dependence as has been shown in section 3.

Case 4: Strong dependence

For the strong dependence case, I model both the error term(and the regressor) as a two-way error component model. In this setup, the estimation of β does not include fixed effects as in the case where the unobservable fixed effects (individual and time) are not correlated with the regressor.

$y_{it} = x_{it}\beta + u_t$

$u_{it} = \alpha_i + \delta_t + \varepsilon_{it}$

where $\alpha_i \sim N(0, \sigma_\alpha)$, $\delta_t \sim N(0, \sigma_\delta)$, $\varepsilon_{it} \sim N(0, 1)$

Simulation results for all these cases and for different sample sizes of the cross-section (N) and the time-series (T) are reported in table 1 and 2. The simulations are based on 10,000 replications. Rows labeled White use the conventional heteroskedasticity standard errors for iid observations. Rows leveled $CCE_{spatial}$ and CCE_{time} use the cluster estimator for time series dependence and spatial dependence, respectively. Row $2CCE$ uses the two-way cluster estimator and the row labeled $multi-way$

uses an extension of the $2CCE$ that also takes into account non-contemporaneous spatial dependence. Finally, the parametric row uses the the variance estimator constructed from the estimation of the parametric model. The estimates of the parametric model come from the NLS estimator discussed in the previous section in order to avoid the incidental parameter problem in both dimensions.

Table 1 reports rejection rates for the data generating processes in case 1 using a test with nominal size of 5%. The most remarkable result of this table is that even though the $2CCE$ estimator is a consistent estimator of the standard errors under the DGP, the rejection rates are about 15% when $N = 10$ and $T = 10$ and about 10% when $N = 20$ and $T = 20$. These over-rejections suggest a bad behavior of the t-test based on the $2CCE$.

Tables 2 and 3 report the rejection rates for cases 2 and 3. The main message of these tables is that the $2CCE$ is not capable of capturing all the dependence structure in those models and as a consequence the test based on the $2CCE$ reports high rejection rates even when $N = 50$ and $T = 50$. The multi-way cluster that should handle non-contemporaneous spatial dependence also delivers high rejection rates in small and moderate sample sizes, whereas the parametric model almost attains the nominal size of the test.

One nice thing to notice is that in case 3 (where the spatial dependence comes from a spatial moving average) a hybrid between the parametric model for the time series dependence and the cluster for the spatial dependence performs pretty well. The idea is to get rid of the time series dependence by using a quasi-difference and then handle the spatial dependence by clustering in the time series for the transformed model. This procedure outperforms the $2CCE$ and the multi-way because once the time series dependence is eliminated, the CCE_{time} is consistent and is well estimated even for moderate T. This procedure is also recommendable when the researcher thinks that this structure is sensible for modeling the dependence but does not have a prior on what the weighting matrix W should look like.

5.2 US state-year panel model of wage inequality and minimum wage

In the second set of simulations, I model error dependence in a US state-year panel model of minimum wages and wage inequality with $N = 50$ and $T = 30$. The purpose of this section is twofold. First, I use this model to conduct a calibrated Monte Carlo simulation and study the small sample performance of the approaches studied in the previous sections. Second, I study the implications of neglecting error dependence outright in a widely used data set.[33]

5.2.1 Calibrated Monte Carlo simulation

In this section I consider the state-year panel model of minimum wages and wage inequality of David et al. (2016):

$$y_{it} = \alpha_i + \delta_t + \beta_1 x_{it} + \beta_2 x_{it}^2 + u_{it} \tag{5.8}$$

where y_{it} is the log of the 10^{th} percentile state wage relative to the log of the median state wage (a proxy of wage inequality), α_i and δ_t are state and year fixed effects and x_{it} is the minimum wage in state i in year t.

The idea is to study the performance of the 2CCE and the spatial panel model in a model that possess an error dependence structure as in David et al. (2016). To do so, I first estimate a spatial autoregressive panel model for the residual of 5.8.

$$u_{it} = \rho u_{it-1} + \lambda \sum_{j=1}^{N} w_{ij} u_{jt} + e_{it} \tag{5.9}$$

where $E(e_{it}) = \sigma^2$. Given the evidence in Barrios et al. (2012) I define w_{ij} as taking the value of one if the states share a border and zero otherwise.[34] Estimates of $\theta = [\lambda, \rho_1, \sigma^2]$ are obtained both by conditional QML and the NLS described in section 4 (residuals come from the OLS estimated errors

[33] This model has been applied by Lee (1999) and David et al. (2016) to study the effect of changes in state minimum wages over state wage inequality in the U.S. In an influential paper, Lee (1999) used cross-state variation in the state minimum wage and found an important effect of minimum wages on wage inequality from 1979 through 1988. David et al. (2016) reassess this evidence by extending the sample until 2009 and using an instrument to resolve an existing bias in the literature. Relying on variation in the statutory minimum wage across state and time, they find a modest but still significant effect on state wage inequality. They point out that finding a significant impact of the minimum wage opens a puzzle, since between 1979-2012 there is no year in which more than 10% of hours were paid at or below the federal or applicable state minimum wage. They provide evidence of spillover effects as a possible explanation. I argue that considering error dependence could also help to reconcile the puzzle, since this is a setting where dependence is very likely to exist. In fact, Barrios et al. (2012) show that yearly earnings and state regulatory variables such as minimum wages exhibit substantial correlation across neighboring states. This dependence could be explained by geographical or local labor market features.

[34] I estimate the model using other definitions of the spatial matrix and also including higher spatial lags. The extra lags are not significant.

of model David et al. (2016)) and are summarized in Table 5. As we can see from these tables, the conditional QML (from now on θ_{QML}) and the NLS (from now on θ_{NLS}) estimators of the parameters are highly significant. The difference between θ_{QML} and the θ_{NLS} might be due to the incidental parameter problem that affects θ_{QML}.

I then, use θ_{NLS} to generate replications of simulated data with the same characteristics as the error model in 5.9. En each replication s I run a regression between the simulated process u_{it}^{s} and the actual state minimum wage x_{it} and its square x_{it}^{2} and compute the OLS estimator and different standard errors (White, CCE, 2WCCE, spatial panel model). Then I separately evaluate, using individual tests, how many times I reject the true null hypothesis of $\beta_1 = 0$ and $\beta_2 = 0$, with a 5% confidence level.

The results of the simulation are summarized in table 7. The type I error is around 70% for the OLS estimators of β_1 and β_2 if I do not control for any type of dependence (White). The type I errors calculated using different cluster variances are far away from the nominal size, even when I use 2CCE or a multi-way cluster. It is remarkable that the 2CCE and the multi-way cluster approach, which control for both type of dependences, still have a high probability of rejecting the true null hypothesis (around 16%). These results reinforce the importance of having large N and large T, in order to use the cluster approach to control for general forms of dependence. Finally, the spatial panel standard errors is the only one which almost attained the real nominal size of the test for β_1 and β_2.

5.2.2 Robust standard errors for the US state-year panel model of wage inequality and minimum wages

In order to study the effects of neglecting error dependence is this setting, I re-estimate the model of David et al. (2016) and compute standard errors using the methodologies previously discussed in this article (2CCE and the panel SAR model). Then, I analyze how the results change once robust standard errors are used in this context where there is error dependence in more than one dimension.

I separately discuss the variance estimates for the marginal effects arising from: (i) OLS estimation of US state inequality over the state effective minimum and its square, plus fixed effects and time effects as controls (ii) 2SLS estimation after instrumenting both the state effective minimum wage and the square of the state effective minimum wage, plus fixed effects and time effects as controls.[35]

As can be seen in Table 8, the standard errors for the OLS estimates of the minimum wage and its square increase by approximately 200% when I switch from non-cluster (White) to one way cluster by state ($CCE_{spatial}$) or one way cluster by time (CCE_{time}). When I only cluster by one dimension, either by state or by time, the OLS estimator of the minimum wage variable remains significant at the 1% confidence level. However, the OLS estimator of the square of the minimum wage reduces its individual significance at 5% level with respect to the non-cluster case. However, when I use variance estimators that take into account the dependence along more than one dimension, the OLS estimator of the square of the effective minimum wage, $\hat{\beta}_2$, loses significance at all levels, while the OLS estimator of the minimum wage, $\hat{\beta}_1$, remains significant, but this time at the 5% for the 2CCE and the multi-way cluster, and at the 10% using the spatial panel standard errors.

Given that the object of interest is the total marginal effect of the effective minimum wage, I present in Table 9 the results of the estimated average marginal effect of the minimum wage over US state inequality, which is given by $\hat{\beta}_1 + 2\hat{\beta}_2 \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} [logwage_{it}^{Min} - logwage_{it}(50)]$. The standard error of the estimated marginal effect increases by 200% when I use a one way-cluster (either by time or space) and by 300% when I use a 2CCE, compared to no cluster. Despite this fact, the estimated marginal effect still remains significant at the 5% level. Finally, when I use the panel SAR model to account for the error dependence, the marginal effect of the minimum wage over US state inequality becomes insignificant.

In the 2SLS estimation, once I control for error dependence using either the 2CCE or the panel SAR, the marginal effect is not significant anymore. Table 11 shows the results for the estimated average marginal effect. Similar to the OLS estimation, the standard error of the estimated average marginal effect increases by 100% and 200% when I use the 2CCE and the panel SAR, respectively. [36]

[35]The instruments are, as in David et al. (2016), the statutory minimum and the square of the predicted value from a regression between the state effective minimum wage and the statutory minimum plus fixed and time effects, respectively.

[36] It is important to consider that the instrument does not have much variation and this could be reflected in the higher standard errors (with respect to the OLS estimators). The federal minimum wage, which is an important input of the instrument used in the 2SLS, does not vary across states, and changes over time in only 9 of the 31 years.

David et al. (2016) stress that the large significance found in their estimations are not in line with the fact that between 1979 and 2009, no more than nine percent of all workers were paid at or below the federal or applicable state minimum wage. They explain this puzzle by arguing that the effect found with the OLS and the 2SLS estimators captures a spillover effect; whenever the minimum wage rises, the wages of workers earning above the minimum also rise. It is possible however that the puzzle of the significance found by the OLS and 2SLS estimation is due to underestimated standard errors, as a consequence of not considering all possible error dependence as it shown in tables 8, 9, 10, and 11.

6 Conclusions

This article studies different approaches to conduct robust inference in panel models with spatial and time error dependence. These two types of dependence are likely to be present in aggregate panels (e.g state-year or industry-year panels). However, considering robust standard errors to both types of dependence in not usual in empirical applications. The article makes two contributions. First, it analyses, analytically and computationally, some approaches to deal with error dependence in two dimensions. Specifically, it evaluates the performance of each of the approaches in aggregate panels in which neither N nor T is very large. Second, it shows the importance of considering spatial and time error dependence when working with a US state minimum wage panel model.

The first result of this article shows that the two-way cluster (2CCE) is a consistent estimator of the variance when there is contemporaneous spatial and time dependence, even when both types of error dependence are strong. Therefore, this estimator can be also used when the error term of the model presents individual- and time- fixed effects as in a two-way error component model. However, it is shown that the 2CCE might suffer from considerable finite variance as is affected by the two types of dependence. As a consequence of the latter and since the 2CCE converges to its asymptotic distribution at the rate $min\left(\sqrt{N}, \sqrt{T}\right)$, tests based on the normal approximation might perform poorly in terms of rejection rates, even in panels with a moderate sample size. The simulation results suggest that alternative approaches such as a panel version of the SAR model for the dependence or a hybrid between the cluster and the panel SAR approach, outperform the cluster approach in settings where N are T are not that large.

This article also shows that the cluster estimator for spatial dependence can be expressed as a fully flexible spatial autoregressive panel model. Then, it is possible to exploit that insight and use some prior knowledge about the dependence to estimate more parsimonious models that deliver more precise standard errors. Moreover, this connection provides a natural framework to start thinking about model selection for error dependence that should be explored in further research.

Finally, the results of the empirical application highlight the consequences of neglecting spatial and time dependence when working with US state-year panel data of wage inequality and minimum wages. When standard errors without considering any kind of dependence are computed, the marginal effect of the minimum wage on wage inequality is highly significant. However, once both types of error dependence are considered, the marginal effect is no longer significant. These result encourage the exploration of error dependence and the use of robust standard errors when working with aggregate panels such as state-year, country-year or industry-year panels, as these units might be connected through neighboring effects, economic trade or linkages in production.

References

Amemiya, T. (1971), 'The estimation of the variances in a variance-components model', *International Economic Review* pp. 1–13.

Anselin, L. (1988), *Spatial econometrics: methods and models*, Vol. 4, Springer.

Anselin, L. (2001), 'Spatial econometrics', *A companion to theoretical econometrics* **310330**.

Arellano, M. (1987), 'Computing robust standard errors for within-groups estimators*', *Oxford bulletin of Economics and Statistics* **49**(4), 431–434.

Baltagi, B. H., Heun Song, S., Cheol Jung, B. & Koh, W. (2007), 'Testing for serial correlation, spatial autocorrelation and random effects using panel data', *Journal of Econometrics* **140**(1), 5–51.

Baltagi, B. H. & Pirotte, A. (2010), 'Panel data inference under spatial dependence', *Economic Modelling* **27**(6), 1368–1381.

Baltagi, B. H., Song, S. H. & Koh, W. (2003), 'Testing panel data regression models with spatial error correlation', *Journal of econometrics* **117**(1), 123–150.

Barrios, T., Diamond, R., Imbens, G. W. & Kolesar, M. (2012), 'Clustering, spatial correlations, and randomization inference', *Journal of the American Statistical Association* **107**(498), 578–591.

Bertrand, M., Duflo, E. & Mullainathan, S. (2004), 'How much should we trust differences-in-differences estimates?', *Quarterly Journal of Economics* **119**(1).

Bester, C. A., Conley, T. G. & Hansen, C. B. (2011), 'Inference with dependent data using cluster covariance estimators', *Journal of Econometrics* **165**(2), 137–151.

Cameron, A. C., Gelbach, J. B. & Miller, D. L. (2011), 'Robust inference with multiway clustering', *Journal of Business & Economic Statistics* **29**(2).

Cameron, A. C. & Miller, D. L. (2015), 'A practitioner guide to cluster-robust inference', *Journal of Human Resources* **50**(2), 317–372.

Conley, T. G. (1999), 'Gmm estimation with cross sectional dependence', *Journal of econometrics* **92**(1), 1–45.

David, H., Manning, A. & Smith, C. L. (2016), 'The contribution of the minimum wage to us wage inequality over three decades: a reassessment', *American Economic Journal: Applied Economics* **8**(1), 58–99.

Foerster, A. T., Sarte, P.-D. G. & Watson, M. W. (2008), Sectoral vs. aggregate shocks: A structural factor analysis of industrial production, Technical report, National Bureau of Economic Research.

Hansen, C. B. (2007*a*), 'Asymptotic properties of a robust variance matrix estimator for panel data when t is large', *Journal of Econometrics* **141**(2), 597–620.

Hansen, C. B. (2007*b*), 'Generalized least squares inference in panel and multilevel models with serial correlation and fixed effects', *Journal of Econometrics* **140**(2), 670–694.

Ibragimov, R. & Müller, U. K. (2010), 't-statistic based correlation and heterogeneity robust inference', *Journal of Business & Economic Statistics* **28**(4), 453–468.

Kelejian, H. H. & Prucha, I. R. (1998), 'A generalized spatial two-stage least squares procedure for estimating a spatial autoregressive model with autoregressive disturbances', *The Journal of Real Estate Finance and Economics* **17**(1), 99–121.

Kelejian, H. H. & Prucha, I. R. (2004), 'Estimation of simultaneous systems of spatially interrelated cross sectional equations', *Journal of Econometrics* **118**(1), 27–50.

Kelejian, H. H. & Prucha, I. R. (2010), 'Specification and estimation of spatial autoregressive models with autoregressive and heteroskedastic disturbances', *Journal of Econometrics* **157**(1), 53–67.

Lee, D. S. (1999), 'Wage inequality in the united states during the 1980s: Rising dispersion or falling minimum wage?', *Quarterly Journal of Economics* pp. 977–1023.

Lee, L.-f. & Yu, J. (2010), 'A spatial dynamic panel data model with both time and individual fixed effects', *Econometric Theory* **26**(02), 564–597.

Manresa, E. (2016), 'Estimating the structure of social interactions using panel data', *Unpublished Manuscript. CEMFI, Madrid* .

Newey, W. K. & West, K. D. (1987), 'A simple, positive semi-definite, heteroskedasticity and autocorrelationconsistent covariance matrix'.

Petersen, M. A. (2009), 'Estimating standard errors in finance panel data sets: Comparing approaches', *Review of financial studies* **22**(1), 435–480.

Thompson, S. B. (2011), 'Simple formulas for standard errors that cluster by both firm and time', *Journal of Financial Economics* **99**(1), 1–10.

Wooldridge, J. M. (2003), 'Cluster-sample methods in applied econometrics', *American Economic Review* pp. 133–138.

Yu, J., de Jong, R. & Lee, L.-f. (2008), 'Quasi-maximum likelihood estimators for spatial dynamic panel data with fixed effects when both n and t are large', *Journal of Econometrics* **146**(1), 118–134.

Table 1: Simulation Results. T test rejection rates for a 5% level test

$$u_{it} = \varepsilon_{it} + v_{it}, \varepsilon_{it} = \rho\varepsilon_{it} + e_{it}, v_{it} = \lambda \sum_{j=1}^{N} w_{ij} v_{jt} + \epsilon_{it}$$

	$\rho = 0, \lambda = 0$	$\rho = 0.9, \lambda = 0$	$\rho = 0, \lambda = 0.9$	$\rho = 0.45, \lambda = 0.45$	$\rho = 0.9, \lambda = 0.9$
			$N = 10, T = 10$		
White	0.075	0.121	0.14	0.15	0.19
$CCE_{spatial}$	0.113	0.112	0.17	0.18	0.21
CCE_{time}	0.119	0.142	0.13	0.13	0.14
2CCE	0.153	0.154	0.15	0.15	0.16
Panel SAR	0.061	0.065	0.061	0.059	0.054
			$N = 20, T = 20$		
White	0.055	0.213	0.184	0.112	0.174
$CCE_{spatial}$	0.071	0.062	0.189	0.141	0.164
CCE_{time}	0.084	0.262	0.075	0.082	0.120
2CCE	0.091	0.070	0.082	0.091	0.105
Panel SAR	0.061	0.065	0.061	0.059	0.054
			$N = 50, T = 50$		
White	0.060	0.374	0.161	0.151	0.194
$CCE_{spatial}$	0.067	0.067	0.175	0.133	0.150
CCE_{time}	0.070	0.403	0.051	0.091	0.115
2CCE	0.073	0.072	0.062	0.077	0.084
Panel SAR	0.055	0.057	0.061	0.55	0.045

Note: The table reports rejection rates for 5% level tests from a Monte Carlo simulation experiment for different sample sizes. The simulations are based on 1 000 simulation replications. Row labels indicate which standard errors is used. Columns indicates different values for the parameters of th DGP. Row labeled White uses heteroskedasticity robust standard errors. Rows labeled $CCE_{spatial}$ *and* CCE_{time} *use cluster standard errors, clustering in the cross sectional and the time series dimension, respectively. Row labeled 2CCE uses the two-way cluster standard errors. Finally, row labeled Panel SAR uses standard errors that comes from the estimation of a Panel SAR model.*

Table 2: Simulation Results. T test rejection rates for a 5% level test

$$u_{it} = \lambda \sum_{j=1}^{N} w_{ij} u_{jt} + \rho u_{it-1} + \epsilon_{it}$$

	$N = 10, T = 10$	$N = 20, T = 20$	$N = 50, T = 50$
White	0.075	0.121	0.14
$CCE_{spatial}$	0.113	0.112	0.17
CCE_{time}	0.119	0.142	0.13
$2CCE$	0.153	0.154	0.15
$Multi-way$	0.17	0.15	0.12
Panel SAR	0.061	0.065	0.061

Note: The table reports rejection rates for 5% level tests from a Monte Carlo simulation experiment for different sample sizes. The simulations are based on 1 000 simulation replications. Row labels indicate which standard errors is used. Columns indicates different values for the parameters of th DGP. Row labeled White uses heteroskedasticity robust standard errors. Rows labeled $CCE_{spatial}$ *and* CCE_{time} *use cluster standard errors, clustering in the cross-sectional and the time series dimension, respectively. Row labeled* $2CCE$ *uses the two-way cluster standard errors. Row labeled* $Multi-way$ *is an extension of the* $2CCE$ *that accounts for non-contemporanueos spatial dependence. Finally, row labeled Panel SAR uses standard errors that comes form the estimation of a Panel SAR model.*

Table 3: Simulation Results. T test rejection rates for a 5% level test

$$u_{it} = \rho u_{it-1} + \lambda \sum_{j=1}^{N} w_{ij}\epsilon_{jt}$$

	$N = 10, T = 10$	$N = 20, T = 20$	$N = 50, T = 50$
White	0.055	0.213	0.184
$CCE_{spatial}$	0.071	0.062	0.189
CCE_{time}	0.084	0.262	0.075
$2CCE$	0.091	0.070	0.082
$Multi - way$	0.012	0.010	0.080
Panel SAR	0.061	0.065	0.061
Quasi diff $+ CCE_t$	0.072	0.067	0.065

Note: The table reports rejection rates for 5% level tests from a Monte Carlo simulation experiment for different sample sizes. The simulations are based on 1 000 simulation replications. Row labels indicate which standard errors is used. Columns indicates different values for the parameters of th DGP. Row labeled White uses heteroskedasticity robust standard errors. Rows labeled $CCE_{spatial}$ *and* CCE_{time} *use cluster standard errors, clustering in the cross sectional and the time series dimension, respectively. Row labeled* $2CCE$ *uses the two-way cluster standard errors. Row labeled* $Multi - way$ *is an extension of the* $2CCE$ *that accounts for non-contemporanueos spatial dependence. Row labeled Panel SAR uses standard errors that comes from the estimation of a Panel SAR model. Row labeled* $Quasi diff + CCE_{time}$ *uses standard errors constructed by hybrid between the parametric model for the time series dependence and the cluster for the spatial dependence. It uses the estimation of an AR(1) model to perform a quasi difference and eliminate the time series dependence, then cluster by year the transformed errors to consider the spatial dependence in a non parametric way.*

Table 4: Simulation Results. T test rejection rates for a 5% level test

$$u_{it} = \alpha_i + \delta_t + \varepsilon_{it}$$

	$N = 10, T = 10$	$N = 20, T = 20$	$N = 50, T = 50$
White	0.295	0.385	0.685
$CCE_{spatial}$	0.222	0.188	0.179
CCE_{time}	0.225	0.173	0.177
2*CCE*	0.155	0.099	0.075

Note: The table reports rejection rates for 5% level tests from a Monte Carlo simulation experiment for different sample sizes. The simulations are based on 1 000 simulation replications. Row labels indicate which standard errors is used. Columns indicates different values for the parameters of th DGP. Row labeled White uses heteroskedasticity robust standard errors. Rows labeled $CCE_{spatial}$ and CCE_{time} use cluster standard errors, clustering in the cross sectional and the time series dimension, respectively. Row labeled 2CCE uses the two-way cluster standard errors.

Table 5: Spatial-Time Autoregressive Model of the OLS Transformed Residuals

$$U_t = \lambda W_N U_t + \rho_1 U_{t-1} + \rho_2 W_N U_{t-1} + \epsilon_t$$

	λ	ρ_1	ρ_2
QML	0.20***	0.50***	0.19***
	(0.02)	(0.01)	(0.03)
NLS	0.40***	0.49***	0.08
	(0.05)	(0.02)	(0.05)

Note: The table shows the QML and NLS estimators of the model $U_t = \lambda W_N U_t + \rho_1 U_{t-1} + \rho_2 W_N U_{t-1} + \epsilon_t$*.*
Standard errors in parenthesis are calculated using a bootstrap.
****Significant at 1%, **Significant at 5%, *Significant at 1%.*

Table 6: Spatial-Time Autoregressive Model of the 2SLS Transformed Residuals

$$U_t = \lambda W_N U_t + \rho_1 U_{t-1} + \rho_2 W_N U_{t-1} + \epsilon_t$$

	λ	ρ_1	ρ_2
QML	0.19***	0.49***	0.22***
	(0.02)	(0.01)	(0.02)
NLS	0.42***	0.50***	0.08
	(0.05)	(0.02)	(0.05)

Note: The table shows the QML and NLS estimators of the model $U_t = \lambda W_N U_t + \rho_1 U_{t-1} + \rho_2 W_N U_{t-1} + \epsilon_t$*.*
Standard errors in parenthesis are calculated using a bootstrap.
****Significant at 1%, **Significant at 5%, *Significant at 1%.*

Table 7: T-test rejection rates for 5% Level Tests

$\lambda = 0.4$, $\rho_1 = 0.5$
$N = 50,\ T = 31$

Variance Estimator	β_1	β_2
$White$	0.67	0.69
$CCE_{spatial}$	0.23	0.24
CCE_{time}	0.49	0.54
$2CCE$	0.16	0.19
$Multi-way$	0.16	0.18
$Panel\ SAR$	0.05	0.06

Note: The table reports rejection rates for 5% level tests from a Monte Carlo simulation experiment. The error term of the model is simulated using the spatial-time autoregressive model. The value of the parameters of the spatial autoregressive model used in the simulation comes from the estimation by nonlinear least squares of the spatial-time model of the transformed residuals using the real data. In each simulation, $\hat{\beta}_1$ and $\hat{\beta}_2$ are estimated from an OLS regression of the simulated residuals over the real covariates $logwage^{Min} - logwage(50)$, $[logwage^{Min} - logwage(50)]^2$ and a full set of state dummies and time dummies. Column 1 and 2 show the probability of reject the true null hypothesis of two individual test $\beta_1 = 0$ and $\beta_2 = 0$, respectively. The number of replications is 1000. Row labeled White uses heteroskedasticity robust standard errors. Rows labeled $CCE_{spatial}$ and CCE_{time} use cluster standard errors, clustering in the cross sectional and the time series dimension, respectively. Row labeled $2CCE$ uses the two-way cluster standard errors. Row labeled $Multi-way$ is an extension of the $2CCE$ that accounts for non-contemporanueos spatial dependence that takes into account dependence across different states in t and $t-1$. Row labeled Panel SAR uses standard errors that comes from the estimation of a Panel SAR model.

Table 8: OLS Estimation of the U.S. State-Year Panel Model of Minimum Wage and Wage Inequality

$logwage_{it}(10) - logwage_{it}(50) = \alpha_i + \delta_t + \beta_1[logwage_{it}^{Min} - logwage_{it}(50)] + \beta_2[logwage_{it}^{Min} - logwage_{it}(50)]^2 + u_{it}$

$\hat{\beta}_1$				$\hat{\beta}_2$			
OLS Estimator	**Standard Error Estimator**		**t-statistic**	**OLS Estimator**	**Standard Error Estimator**		**t-statistic**
0.48	*White*	0.06	7.92^{***}	0.2	*White*	0.04	5.63^{***}
	$CCE_{spatial}$	0.14	3.43^{***}		$CCE_{spatial}$	0.08	2.50^{**}
	CCE_{time}	0.15	3.20^{***}		CCE_{time}	0.09	2.23^{**}
	$2CCE$	0.22	2.14^{**}		$2CCE$	0.13	1.53
	$Multi-way$	0.22	2.14^{**}		$Multi-way$	0.13	1.53
	Panel SAR	0.28	1.69^{*}		*Panel SAR*	0.17	1.15

Note: The table shows the results of the OLS regression of the model $logwage_{it}(10) - logwage_{it}(50) = \alpha_i + \delta_t + \beta_1[logwage_{it}^{Min} - logwage_{it}(50)] + \beta_2[logwage_{it}^{Min} - logwage_{it}(50)]^2 + u_{it}$*. The data are the same as used in* David et al. (2016)*. Row labeled White uses heteroskedasticity robust standard errors. Rows labeled* $CCE_{spatial}$ *and* CCE_{time} *use cluster standard errors, clustering in the cross sectional and the time series dimension, respectively. Row labeled* $2CCE$ *uses the two-way cluster standard errors. Row labeled* $Multi-way$ *is an extension of the* $2CCE$ *that accounts for non-contemporanueos spatial dependence that takes into account dependence across different states in* t *and* $t-1$*. Row labeled Panel SAR uses standard errors that comes from the estimation of a Panel SAR model.*

****Significant at 1%, **Significant at 5%, *Significant at 10%.*

Table 9: Marginal Effect of Minimum Wage over Wage Inequality (2SLS estimation)

$$\hat{\beta}_1 + 2\hat{\beta}_2 \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} [logwage_{it}^{Min} - logwage_{it}(50)]$$

$\hat{\beta}_1 + 2\hat{\beta}_2 \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} [logwage_{it}^{Min} - logwage_{it}(50)]$			
OLS Estimator	**Standard Error Estimator**		**t-statistic**
0.27	*White*	0.03	8.70^{***}
	$CCE_{spatial}$	0.07	3.86^{***}
	CCE_{time}	0.08	3.37^{***}
	$2CCE$	0.11	2.24^{**}
	$Multi-way$	0.11	2.24^{**}
	Panel SAR	0.18	1.50

Note: Column one shows the marginal effect of the minimum wage on wage inequality: $\hat{\beta}_1 + 2\hat{\beta}_2 \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} [logwage_{it}^{Min} - logwage_{it}(50)]$. *Column 2 shows different variance estimators of the estimator of the marginal effect. Column 3 shows the t-statistic of the marginal effect using different variance estimators. Results come from an OLS estimation of* $logwage_{it}(10) - logwage_{it}(50) = \alpha_i + \delta_t + \beta_1[logwage_{it}^{Min} - logwage_{it}(50)] + \beta_2[logwage_{it}^{Min} - logwage_{it}(50)]^2 + u_{it}$. *The data are the same as used in* David et al. (2016). *Row labeled White uses heteroskedasticity robust standard errors. Rows labeled* $CCE_{spatial}$ *and* CCE_{time} *use cluster standard errors, clustering in the cross sectional and the time series dimension, respectively. Row labeled* $2CCE$ *uses the two-way cluster standard errors. Row labeled* $Multi-way$ *is an extension of the* $2CCE$ *that accounts for non-contemporanueos spatial dependence that takes into account dependence across different states in* t *and* $t-1$. *Row labeled Panel SAR uses standard errors that comes from the estimation of a Panel SAR model.*

****Significant at 1%, **Significant at 5%, *Significant at 10%.*

Table 10: 2SLS Estimation of the U.S. State-Year Panel Model of Minimum Wage and Wage Inequality

$logwage_{it}(10) - logwage_{it}(50) = \alpha_i + \delta_t + \beta_1[logwage_{it}^{Min} - logwage_{it}(50)] + \beta_2[logwage_{it}^{Min} - logwage_{it}(50)]^2 + u_{it}$

$\hat{\beta}_1$				$\hat{\beta}_2$			
OLS Estimator	**Standard Error Estimator**		**t-statistic**	**OLS Estimator**	**Standard Error Estimator**		**t-statistic**
0.41	*White*	0.24	1.70*	0.2	*White*	0.14	1.43
	$CCE_{spatial}$	0.45	0.91		$CCE_{spatial}$	0.27	0.74
	CCE_{time}	0.34	1.21		CCE_{time}	0.19	1.05
	$2CCE$	0.54	0.75		$2CCE$	0.32	0.63
	$Multi-way$	0.54	0.75		$Multi-way$	0.32	0.63
	Panel SAR	0.67	0.61		*Panel SAR*	0.40	0.50

Note: The table shows the results of the 2SLS regression of the model $logwage_{it}(10) - logwage_{it}(50) = \alpha_i + \delta_t + \beta_1[logwage_{it}^{Min} - logwage_{it}(50)] + \beta_2[logwage_{it}^{Min} - logwage_{it}(50)]^2 + u_{it}$. *The data are the same as in* David et al. (2016). *Row labeled White uses heteroskedasticity robust standard errors. Rows labeled* $CCE_{spatial}$ *and* CCE_{time} *use cluster standard errors, clustering in the cross sectional and the time series dimension, respectively. Row labeled* $2CCE$ *uses the two-way cluster standard errors. Row labeled* $Multi-way$ *is an extension of the* $2CCE$ *that accounts for non-contemporanueos spatial dependence that takes into account dependence across different states in* t *and* $t-1$. *Row labeled Panel SAR uses standard errors that comes from the estimation of a Panel SAR model.*

****Significant at 1%, **Significant at 5%, *Significant at 1%.*

Table 11: Marginal Effect of Minimum Wage over Wage Inequality (2SLS estimation)

$$\hat{\beta}_1 + 2\hat{\beta}_2 \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} [logwage_{it}^{Min} - logwage_{it}(50)]$$

$\hat{\beta}_1 + 2\hat{\beta}_2 \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} [logwage_{it}^{Min} - logwage_{it}(50)]$			
OLS Estimator	**Standard Error Estimator**		**t-statistic**
0.20	*White*	0.12	1.67*
	$CCE_{spatial}$	0.27	0.74
	CCE_{time}	0.15	1.37
	$2CCE$	0.30	0.67
	$Multi - way$	0.30	0.67
	$Panel\ SAR$	0.37	0.54

Note: Column one shows the marginal effect of minimum wage over wage inequality: $\hat{\beta}_1 + 2\hat{\beta}_2 \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} [logwage_{it}^{Min} - logwage_{it}(50)]$*. Column 2 shows different variance estimators of the estimator of the marginal effect. Column 3 shows the t-statistic of the marginal effect using different variance estimators. Results come from a 2SLS estimation of* $logwage_{it}(10) - logwage_{it}(50) = \alpha_i + \delta_t + \beta_1[logwage_{it}^{Min} - logwage_{it}(50)] + \beta_2[logwage_{it}^{Min} - logwage_{it}(50)]^2 + u_{it}$*. The data are the same used in* David et al. (2016)*. Row labeled White uses heteroskedasticity robust standard errors. Rows labeled* $CCE_{spatial}$ *and* CCE_{time} *use cluster standard errors, clustering in the cross sectional and the time series dimension, respectively. Row labeled* $2CCE$ *uses the two-way cluster standard errors. Row labeled* $Multi - way$ *is an extension of the* $2CCE$ *that accounts for non-contemporanueos spatial dependence that takes into account dependence across different states in* t *and* $t - 1$*. Row labeled Panel SAR uses standard errors that comes from the estimation of a Panel SAR model.*

****Significant at 1%, **Significant at 5%, *Significant at 1%.*