

ORACLE INEQUALITIES, VARIABLE SELECTION AND UNIFORM INFERENCE IN HIGH-DIMENSIONAL CORRELATED RANDOM EFFECTS PANEL DATA MODELS

ANDERS BREDAHL KOCK
AARHUS UNIVERSITY AND CREATES

ABSTRACT. In this paper we study high-dimensional correlated random effects panel data models. Our setting is useful as it allows including time invariant covariates as under random effects yet allows for correlation between covariates and unobserved heterogeneity as under fixed effects. We use the Mundlak-Chamberlain device to model this correlation. Allowing for a flexible correlation structure naturally leads to a high-dimensional model in which least squares estimation easily becomes infeasible with even a moderate number of explanatory variables.

Imposing a combination of sparsity and weak sparsity on the parameters of the model we first establish an oracle inequality for the Lasso. This is valid even when the error terms are heteroskedastic and no structure is imposed on the time series dependence of the error terms.

Next, we provide upper bounds on the sup-norm estimation error of the Lasso. As opposed to the classical ℓ_1 - and ℓ_2 -bounds the sup-norm bounds do not directly depend on the unknown degree of sparsity and are thus well suited for thresholding the Lasso for variable selection. We provide sufficient conditions under which thresholding results in consistent model selection. Pointwise valid asymptotic inference is established for a post-thresholding estimator. Finally, we show how the Lasso can be desparsified in the correlated random effects setting and how this leads to uniformly valid inference even in the presence of heteroskedasticity and autocorrelated error terms.

Keywords: Panel data, Lasso, oracle inequality, sup-norm bounds, high-dimensional models, weak sparsity, correlated random effects, Mundlak-Chamberlain, variable selection, uniform inference.

JEL-codes: C01, C10, C23.

1. INTRODUCTION

In this paper we study panel data models under correlated random effects. As we will see, these models naturally become high-dimensional when the correlation between the covariates and the unobserved heterogeneity is to be modeled in a flexible way. The

I am grateful to Mehmet Caner and Peter Phillips for urging me to pursue the ideas of this paper. I would also like to thank seminar participants at PUC-Rio and participants at the European Meeting of Statisticians 2013 in Budapest for helpful comments and suggestions. The paper has also benefitted tremendously from excellent comments by two anonymous referees as well as the associate and co-editor Jianqing Fan. Financial support from the Danish National Research Foundation (DNRF78) is gratefully acknowledged.

e-mail: akock@creates.au.dk. Address: Aarhus University and CREATES, Fuglesangs Alle 4, 8210 Aarhus V.

baseline panel data model studied is

$$(1) \quad y_{i,t} = x'_{i,t} \beta^* + c_i^* + \epsilon_{i,t}, \quad i = 1, \dots, N, \quad t = 1, \dots, T$$

where $x_{i,t}$ is a $p_{N,T} \times 1$ vector of covariates and $p_{N,T}$ is indexed by N and T to indicate that the number of covariates can increase in the sample size. In the sequel we shall omit this indexation. The c_i^* s are the unobserved time invariant heterogeneities (such as intelligence, ability, motivation or perseverance of a person) while the $\epsilon_{i,t}$ are the error terms about which we shall be more specific later.

The two classical assumptions on the unobserved heterogeneities c_i^* are the fixed and random effects assumptions. Under fixed effects, no restrictions are imposed on the dependence between the covariates $x_{i,t}$ and the unobserved heterogeneity c_i^* . While the fixed effects assumption does not restrict the dependence between the covariates and the c_i^* , it rules out the inclusion of time invariant regressors. This is a serious drawback in applications if the primary interest is in the effect of a time invariant variable such as years of schooling in a wage equation. The random effects specification allows including time constant variables but imposes $x_{i,t}$ to be uncorrelated with c_i^* . This is often unreasonable. In a wage regression an observed explanatory variable, such as years of schooling, is very likely to be correlated with unobserved perseverance. The version of the correlated random effects approach studied in this paper strikes a middle ground between fixed and random effects as it allows for correlation between covariates and unobserved heterogeneity while still admitting the inclusion of time constant variables. In this sense, it unifies the fixed and random effects approaches by modeling the relationship between c_i^* and $x_{i,t}$. To distinguish between time invariant and time varying explanatory variables, write $x_{i,t} = (w'_i, v'_{i,t})'$ where w_i are the time invariant regressors and $v_{i,t}$ is a $p_v \times 1$ vector of time varying regressors. Inspired by Mundlak (1978), Chamberlain (1982, 1984) proposed the specification $E(c_i^* | X_i) = \alpha^* + \sum_{t=1}^T \phi_t^{*'} x_{i,t}$ where $X_i = (x_{i,1}, \dots, x_{i,T})'$. This, however, does not allow us to identify the coefficients of the time invariant covariates. Only the sum of the entries of β^* and $\sum_{t=1}^T \phi_t^*$ pertaining to w_i can be identified¹. For that reason, we shall work under a slightly more restrictive version of the Chamberlain device, namely

$$(2) \quad E(c_i^* | X_i) = \alpha^* + \sum_{t=1}^T \phi_t^{*'} v_{i,t}$$

where each of the ϕ_t^* are p_v -dimensional parameter vectors. This rules out the presence of the time invariant w_i on the right hand side of (2). However, the time invariant regressors can still be correlated with the unobserved heterogeneity through their correlation with the time varying $v_{i,t}$ in (2). The time varying covariates can clearly be correlated with the unobserved effects in a rather general way. Thus, the Mundlak-Chamberlain device models the dependence between the c_i^* and the covariates. It allows all elements of $x_{i,t} = (w'_i, v'_{i,t})'$ to be correlated with c_i^* and is much more flexible than the classical random effects assumption $E(c_i^* | X_i) = E(c_i^*)$.

Defining $a_i = c_i^* - E(c_i^* | X_i) = c_i^* - \alpha^* - \sum_{t=1}^T \phi_t^{*'} v_{i,t}$ and plugging into the baseline panel model (1) yields

$$(3) \quad y_{i,t} = x'_{i,t} \beta^* + \alpha^* + \sum_{t=1}^T \phi_t^{*'} v_{i,t} + a_i + \epsilon_{i,t}, \quad i = 1, \dots, N, \quad t = 1, \dots, T.$$

¹The first $p - p_v$ entries of β^* and $\sum_{t=1}^T \phi_t^*$ are the coefficients of the time invariant covariates.

(3) reveals that modeling the dependence between c_i^* and the covariates naturally leads to a high-dimensional model even for moderate p_v since on top of the p parameters in β^* , (3) contains $1 + Tp_v$ parameters pertaining to the Mundlak-Chamberlain specification (2). This will often render standard least squares estimation unstable – or even infeasible if $N < 1 + Tp_v$ ². Such a situation is not unreasonable as it can occur even for $N = 50$ and $T = 5$ if $p_v \geq 10$, a situation not uncommon in growth economics panel data where the number of countries (N) is often limited due to lack of data. This calls for alternative estimation methods if one wants to use the Mundlak-Chamberlain device. We shall show that if $\phi^* = (\alpha^*, \phi_1^*, \dots, \phi_T^*)$ is weakly sparse in the sense that its ℓ_1 -norm is not too large, then we can use the Lasso of Tibshirani (1996) to estimate all parameters of (3). It is important to stress that the c_i^* s themselves need not be weakly sparse. Their size is not restricted. Thus, our results also broaden the domain of applicability of the Mundlak-Chamberlain device. We contribute by

- (1) providing a nonasymptotic oracle inequality for the estimation error of the Lasso.
- (2) establishing a sup-norm upper bound on the estimation error of the Lasso. The techniques involved are quite different from the ones used for establishing the traditional ℓ_1 - or ℓ_2 -oracle inequalities.
- (3) Based on the established sup-norm bound we show that the thresholded Lasso can be used for consistent variable selection.
- (4) Establish pointwise valid inference for a post-thresholding estimator.
- (5) show how the pointwise inferential results can be improved to uniformly valid inference by desparsifying the Lasso as in Zhang and Zhang (2014); van de Geer et al. (2014). We provide an estimator of the asymptotic covariance matrix of the desparsified Lasso which is valid under heteroskedasticity and dependence.

1.1. Related literature. The last 10-15 years have witnessed a great deal of research into procedures that can handle high-dimensional data sets. In particular, a lot of attention has been given to penalized estimators. The Lasso of Tibshirani (1996) is the most prominent of these procedures and a lot of subsequent research has focussed on investigating the theoretical properties of Lasso type estimators, see Fan and Li (2001), Zhao and Yu (2006), Meinshausen and Bühlmann (2006), Candès and Tao (2007), Fan and Lv (2008), Bickel et al. (2009), Belloni and Chernozhukov (2011), Negahban et al. (2012), Fan et al. (2014) and Bühlmann and van de Geer (2011) to mention just a few. For recent reviews we refer to Bühlmann and van de Geer (2011), Belloni and Chernozhukov (2011) and Fan et al. (2011).

In the context of panel data models Kock (2013a) has considered oracle efficient inference in random and fixed effects panel data models with fewer covariates than observations. In Li et al. (2015) and Qian and Su (2016) panel data models with structural breaks are estimated by shrinkage estimators. Caner and Han (2014) propose a group bridge estimator in approximate factor panel data models while oracle inequalities in fixed effects panel data models have been established in Belloni et al. (2015) and Kock (2013b). Koenker (2004) introduced shrinkage of the individual effects to alleviate the incidental parameter problem in panel quantile regression. This was further studied and developed by Galvao and Montes-Rojas (2010) and Lamarche (2010). Manresa (2013) used the Lasso to structure

²To understand the infeasibility of least squares when $N < 1 + Tp_v$ it suffices to note that there is no variation over time in the covariates pertaining to ϕ_t^* , $t = 1, \dots, T$ which results in a Gram matrix of at most rank N .

spillovers between individuals in panels. For excellent general expositions of panel data models we refer to Baltagi (2008); Hsiao (2014); Wooldridge (2010).

The rest of the paper is organized as follows: Section 2 introduces relevant notation and discusses weak sparsity. Section 3 provides a non-asymptotic oracle inequalities for the Lasso under correlated random effects. Next, Section 4 establishes a sup-norm bound for the Lasso and shows how this can be used to conduct consistent variable selection via thresholding. Pointwise inferential results are established. Section 5 constructs uniformly valid confidence bands for the desparsified Lasso. Finally, Section 6 provides a simulation study and Section 7 concludes. All proofs are deferred to the appendix.

2. SETUP AND NOTATION

2.1. Notation. For any $x \in \mathbb{R}^n$, $\|x\| = \sqrt{\sum_{i=1}^n x_i^2}$, $\|x\|_{\ell_1} = \sum_{i=1}^n |x_i|$ and $\|x\|_{\ell_\infty} = \max_{1 \leq i \leq n} |x_i|$ denote ℓ_2 -, ℓ_1 - and ℓ_∞ -norms, respectively. We shall also make use of $\|x\|_{\ell_0} = \sum_{i=1}^n \mathbf{1}_{\{x_i \neq 0\}}$ which is simply the number of non-zero entries of x . For an $n \times n$ matrix M , $\|M\|_{\ell_\infty} = \max_{1 \leq i \leq n} \sum_{j=1}^n |M_{i,j}|$ denotes the induced ℓ_∞ matrix norm while $\|M\|_\infty = \max_{1 \leq i,j \leq n} |M_{i,j}|$. If M is also symmetric, $\phi_{\min}(M)$ and $\phi_{\max}(M)$ denote the minimal and maximal eigenvalues of M . For two deterministic sequences a_n and b_n we write $a_n \asymp b_n$ if there exist constants $0 < a_1 \leq a_2$ such that $a_1 b_n \leq a_n \leq a_2 b_n$ for all $n \geq 1$. For any set A , $|A|$ denotes its cardinality while A^c denotes its complement.

For any vector $x \in \mathbb{R}^n$ and subset A of $\{1, \dots, n\}$, x_A denotes the vector in $\mathbb{R}^{|A|}$ only consisting of the elements indexed by A . Next, for any two real numbers a and b , $a \wedge b = \min(a, b)$ and $a \vee b = \max(a, b)$. $\lfloor a \rfloor$ denotes the largest integer no greater than a . For any $x \in \mathbb{R}^n$, $\text{sign}(x)$ denotes the sign function applied to each component of x where, by convention, the sign of zero is zero. All asymptotic results are for $N \rightarrow \infty$ with T fixed.

2.2. Correlated random effects, weak sparsity and the Mundlak-Chamberlain device. In (1), let $J_1 = \{j : \beta_j^* \neq 0\} \subseteq \{1, \dots, p\}$ denote the set of active covariates with $s_1 = |J_1|$. We shall assume that β^* is sparse, i.e. $s_1 < p$, which is a standard assumption in the literature on high-dimensional models. We turn next to the weak sparsity of the Mundlak-Chamberlain projection. Specification (2) is a generalization of the Mundlak (1978) projection

$$(4) \quad E(c_i^* | X_i) = \alpha^* + \psi^{*'} \bar{v}_i$$

where $\bar{v}_i = 1/T \sum_{t=1}^T v_{i,t}$. (2) reduces to (4) (up to a scalar $1/T$) if $\phi_t^* = \psi^*$ for $t = 1, \dots, T$. The Chamberlain specification (2) allows for a richer correlation structure between $x_{i,t}$ and c_i^* than the Mundlak specification and we shall therefore adopt the former in the sequel calling it the Mundlak-Chamberlain specification/device. Note again that both settings allow only the time varying covariates $v_{i,t}$ to enter the specification of $E(c_i^* | X_i)$ such that the correlation between the time invariant w_i and the unobserved effects c_i^* must go through the $v_{i,t}$. This is necessary to identify the coefficients of the time invariant covariates and still allows for a rather flexible correlation structure between these and the unobserved effects.

To further motivate how high-dimensionality can arise naturally in the correlated random effects setting, notice that it is even more general than (2) to assume

$$E(c_i^* | X_i) = g(V_i)$$

for some Borel-measurable function $g : \mathbb{R}^{p_v T} \rightarrow \mathbb{R}$ and $V_i = (v_{i,1}, \dots, v_{i,T})'$. One can now seek to approximate g by a linear combination of basis functions (instead of only the plain $p_v T$ entries of X_i as in the Mundlak-Chamberlain device). For this approximation to work well, one would most likely need many such basis functions resulting in a very high-dimensional model. If, for example, g is additively separable such that $g(V_i) = \sum_{t=1}^T \sum_{j=1}^{p_v} g_{tj}(v_{i,t,j})$ and one wishes to approximate each of the $g_{tj}(v_{i,t,j})$ by a linear combination of B basis functions then this would lead to $T p_v B$ parameters to be estimated in addition to β^* upon plugging into (1). Least squares will be infeasible if $N < T p_v B$ which can easily occur. While the theory below remains valid if the basis functions can be chosen bounded we shall focus on assumption (2) in the sequel. This has been used in applied papers such as Papke and Wooldridge (2008), Christiansen et al. (2008) or Xu et al. (2009).

We assume that the coefficients ϕ^* are weakly sparse (a generalization of strict, or ℓ_0 , sparsity) in the sense that $R = \|\phi^*\|_{\ell_1}$ does not increase faster than \sqrt{N} . In particular, R is even bounded if the entries of ϕ^* sorted in decreasing order satisfy $|\phi_j^*| \leq A j^{-d}$, $j = 1, \dots, 1 + Tp$ for some $A \geq 0$ and $d > 1$ as this implies absolute summability of the entries of ϕ^* . The weak sparsity assumption does not require any of the entries of ϕ^* to equal zero but does not rule this out either.

3. AN ORACLE INEQUALITY

The model studied is the one in (3). Let $z_{i,t} = (1, v'_{i,1}, \dots, v'_{i,T}, x'_{i,t})'$ be the $(p + T p_v + 1) \times 1$ vector of explanatory variables for individual i at time t . Note, that only the last p_v entries of $x_{i,t}$ vary across $t = 1, \dots, T$. Define $Z_i = (z_{i,1}, \dots, z_{i,T})'$ and $Z = (Z'_1, \dots, Z'_N)'$ as well as $a = (a_1 \iota', \dots, a_N \iota')'$ where ι is a $T \times 1$ vector of ones. Next, $y_i = (y_{i,1}, \dots, y_{i,T})'$, $\epsilon_i = (\epsilon_{i,1}, \dots, \epsilon_{i,T})'$ for $i = 1, \dots, N$ and $y = (y'_1, \dots, y'_N)'$ as well as $\epsilon = (\epsilon'_1, \dots, \epsilon'_N)'$. Now we may write

$$y = Z\gamma^* + (a + \epsilon) = Z\gamma^* + u.$$

where $\gamma^* = (\phi^{*'}, \beta^{*'})'$ and $u = a + \epsilon$ is a composite error term. Define $\Gamma = E(\frac{1}{NT} Z' Z)$ whose properties will enter the oracle inequalities below. $\gamma^* = (\phi^{*'}, \beta^{*'})'$ is estimated by minimizing

$$(5) \quad L(\gamma) = \frac{1}{2} \sum_{i=1}^N \sum_{t=1}^T (y_{i,t} - z'_{i,t} \gamma)^2 + \lambda_{N,T} \sum_{k=1}^{p+T p_v+1} |\gamma_k| = \frac{1}{2} \|y - Z\gamma\|^2 + \lambda_{N,T} \|\gamma\|_{\ell_1}.$$

The Lasso estimator is denoted $\hat{\gamma} = (\hat{\phi}', \hat{\beta}')'$. In order to state the oracle inequality for the Lasso in the correlated random effects setting we first put forward the statistical assumptions of the panel data model:

- A1 a) $\{X_i, \epsilon_i\}_{i=1}^N$ are identically and independently distributed
- b) $E(\epsilon_{i,t} | X_i, c_i^*) = 0$ for $i = 1, \dots, N$

Assumption A1a) is standard in the panel data literature, see e.g. Wooldridge (2002) or Arellano (2003). We stress that the requirement that the data is identically distributed is not necessary but it makes the exposition slightly easier. Part b) is the standard strict exogeneity assumption on $\{x_{i,t}, t = 1, \dots, T\}$ (conditional on the unobserved effects). Note that we do not impose any restrictions on the temporal dependence of the error terms. Furthermore, $\{\epsilon_{1,t}\}_{t=1}^T$ do not have to be identically distributed and in particular they can be conditionally and unconditionally heteroskedastic.

- A2) The covariates $x_{1,t}$ are jointly uniformly subgaussian in the sense that $\sup_{\|b\| \leq 1} E \exp((x'_{1,t}b)^2/L^2) \leq 1$ for some $L > 0$ and $\epsilon_{1,t}$ are uniformly subgaussian, i.e. there exist constants C and K such that $P(|\epsilon_{1,t}| \geq t) \leq \frac{1}{2}Ke^{-Ct^2}$ for all $1 \leq t \leq T$.

Assumption A2) controls the tail behaviour of the covariates and the error terms. It is a standard assumption in the high-dimensional econometrics literature. Under assumptions A1 and A2 we have the following oracle inequality. Recall that $J_1 = \{j : \beta_j^* \neq 0\}$ and $R = \|\phi^*\|_{\ell_1}$.

Theorem 1. *Let assumptions A1 and A2 be satisfied and let $E(\epsilon_{i,t}) = 0$ for all i and t . Assume that $\phi_{\min}(\Gamma) > 0$. Choose $\lambda_{N,T} = \sqrt{16a_{N,T} \log(p + Tp_v + 1)NT}$ and assume $32\sqrt{\frac{a_{N,T} \log(p + Tp_v + 1)}{N}}|J_1| \leq \phi_{\min}(\Gamma)/2$ and $16\sqrt{\frac{a_{N,T} \log(p + Tp_v + 1)}{N}}R \leq 1$ for some $a_{N,T} \geq 1$. Then, one has*

$$(6) \quad \|\hat{\gamma} - \gamma^*\| \leq 24\sqrt{\frac{a_{N,T} \log(p + Tp_v + 1)}{\phi_{\min}^2(\Gamma)N}}|J_1| + 8\left(\frac{a_{N,T} \log(p + Tp_v + 1)}{\phi_{\min}^2(\Gamma)N}R^2\right)^{1/4}$$

with probability at least $1 - A(p + Tp_v + 1)^{1-Ba_{N,T}} - A(p + Tp_v + 1)^{2-Ba_{N,T}}$ for positive constants $A, B > 0$. Furthermore, the bound in (6) is valid uniformly over $\{\beta^* \in \mathbb{R}^p : \|\beta^*\|_{\ell_0} \leq s_1\} \times \{\phi^* \in \mathbb{R}^{p_v T+1} : \|\phi^*\|_{\ell_1} \leq R\}$.

The bound in (6) consists of two natural parts: i) the first summand which is related to the effective dimension (the number of non-zero entries) of β^* and ii) the second summand which is related to the "dimension" (the ℓ_1 -norm R) of ϕ^* .

The oracle inequalities in (6) is valid for any $a_{N,T} \geq 1$. The larger one chooses $a_{N,T}$, the larger will the probability with which the oracle inequality is valid be. However, the right hand side of (6) is increasing in $a_{N,T}$ and thus there is a tradeoff between sharpness of the bound and the probability with which it is valid. Later we shall see that the concrete choice $a_{N,T} = \log(p \vee N \vee T)$ works well for our purposes.

In the case where ϕ^* is sparse as well, with the number of non-zero entries being s_2 , the second summand in (6) can be shown to drop out at the price of replacing $|J_1|$ by the total number of non-zero entries in γ^* in the first summand.

Notice that all entries of $\hat{\gamma}$ converge at the same rate. We do not have to distinguish between the estimates of β^* and the ϕ^* pertaining to the Mundlak-Chamberlain assumption on the unobserved heterogeneity. Next, no gains are made from larger T – as the dependence over time of the covariates has not been restricted, one can not hope to gain precision from more time series observations. The mere application of the Mundlak-Chamberlain device implies that the first $1 + p_v T$ columns of Z_i by construction have no variation over time which leaves no room for imposing independence or mixing assumptions on them.

The challenging part of the proof of Theorem 1 consists in providing the lower bound on the probability with which the oracle inequality is valid. In particular, many oracle inequalities rely on independent sampling which is not satisfied in our panel data framework.

4. A SUP-NORM BOUND, VARIABLE SELECTION BY THRESHOLDING AND POINTWISE CONFIDENCE BANDS

So far we have considered oracle inequalities in the ℓ_2 -norm. We next turn to variable selection by means of thresholding. For this purpose we can threshold $\hat{\gamma}$ based on the

ℓ_2 -bounds on the estimation error in Theorem 1. However, first developing an upper bound on the sup-norm estimation error will allow us to make a finer distinction between the zero and non-zero coefficients. To derive an upper bound on the sup-norm estimation error we shall assume that also ϕ^* is strictly sparse and let $J = \{j : \gamma_j^* \neq 0\}$ denote the set of non-zero entries of $\gamma^* = (\phi^{*'} , \beta^{*'})'$. In this case $|J_1|$ in (6) must be replaced by $|J|$ while $R = 0$. Assuming that the coefficients in the Mundlak-Chamberlain projection are strictly sparse is not unreasonable, yet less general than the setting studied so far, and a weaker version of (7) below can be established even in the weakly sparse setting. However, this bound would depend on R and here we are seeking sup-norm bounds neither depending on R nor $|J|$.

Theorem 2. *Let the assumptions of Theorem 1 be satisfied with J_1 replaced by J and $R = 0$. Assume furthermore that $\sqrt{\frac{\log^2(N \vee T \vee p)}{\phi_{\min}^2(\Gamma)N}} |J| \rightarrow 0$. Setting $a_{N,T} = \log(N \vee T \vee p)$, one has*

$$(7) \quad \|\hat{\gamma} - \gamma^*\|_{\ell_\infty} \leq 9 \|\Gamma^{-1}\|_{\ell_\infty} \sqrt{\frac{\log^2(N \vee T \vee p)}{N}}$$

with probability tending to one.

$\sqrt{\frac{\log^2(N \vee T \vee p)}{\phi_{\min}^2(\Gamma)N}} |J| \rightarrow 0$ restricts the growth rate of $s = |J|$ and is rather standard. The main point is that opposed to (6) the upper bound in (7) does not depend on the underlying dimension s which will allow for sharper thresholding in the sequel and thus more precise variable selection under correlated random effects. Thus, the above bounds are sharper than the ones one could have obtained by simply using that $\|\cdot\|_{\ell_\infty} \leq \|\cdot\|$. As the upper bound in Theorem 2 increases in $\|\Gamma^{-1}\|_{\ell_\infty}$ it is useful to provide examples of settings where $\|\Gamma^{-1}\|_{\ell_\infty}$ is bounded.

Lemma 1. *Assume that $x_{i,t} = v_{i,t}$ is weakly stationary for $i = 1, \dots, N$ and define $\Gamma_0 = E(x_{1,1}x'_{1,1})$. Assume that $E(x_{1,t}x'_{1,s}) = \rho_1^{|t-s|}\Gamma_0$, $\rho_1 \in (-1, 1)$ with $\frac{1+|\rho_1|}{1-|\rho_1|} \frac{4}{T^2(1-\rho_1)^2} < 1$. Then, $\|\Gamma^{-1}\|_{\ell_\infty}$ is bounded if $\|\Gamma_0^{-1}\|_{\ell_\infty}$ is bounded.*

The stationarity assumption in Lemma 1 is rather innocent and it is not surprising that as temporal dependence, ρ_1 , between the covariates increases, Γ gets closer to being singular and $\|\Gamma^{-1}\|_{\ell_\infty}$ increases. The assumption that $x_{i,t} = v_{i,t}$ states that we only have time varying regressors and is merely made for technical convenience such that the covariance matrix Γ does not have to be split into more than two submatrices prior to inversion. Note that the contemporaneous correlation between the covariates, Γ_0 , can be rather general – for example $\|\Gamma_0^{-1}\|_{\ell_\infty}$ is bounded if Γ_0 is an equicorrelation matrix or if it is a Toeplitz matrix of the form $\Gamma_{0,l,k} = \rho_2^{|l-k|}$ for some $\rho_2 \in (-1, 1)$. Finally, while Lemma 1 provides an example of $\|\Gamma^{-1}\|_{\ell_\infty}$ being bounded, (7) remains useful for variable selection by means of thresholding as long as $\|\Gamma^{-1}\|_{\ell_\infty}$ does not increase too fast.

4.1. Variable selection by thresholding. Having established an upper bound on $\|\hat{\gamma} - \gamma^*\|_{\ell_\infty}$, we turn next to variable selection by thresholding. Define

$$(8) \quad \tilde{\gamma}_j = \begin{cases} \hat{\gamma}_j & \text{if } |\hat{\gamma}_j| \geq L \\ 0 & \text{if } |\hat{\gamma}_j| < L \end{cases}$$

for some $L > 0$. Setting $S_1 = 9 \|\Gamma^{-1}\|_{\ell_\infty} \frac{\log[N \vee p \vee T]}{\sqrt{N}}$ we have the following result.

Theorem 3. *Let the assumptions of Theorem 2 be satisfied and assume that $\min_{j \in J} |\gamma_j^*| \geq 4S_1$ and set $L = 2S_1$. Then,*

$$P(\text{sign}(\hat{\gamma}) = \text{sign}(\gamma^*)) \rightarrow 1.$$

Theorem 3 gives sufficient conditions under which the thresholded Lasso can detect the correct sparsity pattern of γ^* . The important thing to notice is that the absolute value of the smallest non-zero coefficient must be at least of the order of the ℓ_∞ -rate of convergence of $\hat{\gamma}$ to γ^* . As we have argued above there is a clear wedge between the larger ℓ_2 -estimation error bound from Theorem 1, which depends on $s = |J|$, and the smaller ℓ_∞ -estimation error bound from Theorem 2. Therefore, it is important to derive a sup-norm bound prior to thresholding as thresholding based on this allows consistent model selection even when the non-zero coefficients are much closer to zero than would be possible in the case of thresholding based on ℓ_2 -norm error bounds.

4.2. Pointwise valid confidence bands. Having selected variables by thresholding as justified by Theorem 3 one can reestimate the coefficients of the selected variables, i.e. those indexed by $\hat{J} = \{j : \hat{\gamma}_j \neq 0\}$, by a least squares regression only including these variables. Formally, let $\hat{\gamma}_{PostOLS}$ be the vector whose j th element equals the just explained least square estimate for all $j \in \hat{J}$ and zero otherwise. $\hat{\gamma}_{OLS,J}$ will denote oracle assisted least squares only including the relevant variables, i.e. those indexed by J . The following theorem shows that this indeed leads to pointwise valid confidence bands for the non-zero entries of γ^* , i.e. those indexed by J , as $\hat{J} = J$ asymptotically by Theorem 3.

Theorem 4. *Let the assumptions of Theorem 3 be satisfied and assume that $Z_J' Z_J$ is invertible for N sufficiently large. Then, for any vector ρ of length $|J|$,*

$$(9) \quad \left| \sqrt{N} \rho' (\hat{\gamma}_{PostOLS,J} - \gamma_J^*) - \sqrt{N} \rho' (\hat{\gamma}_{OLS,J} - \gamma_J^*) \right| = o_p(1).$$

Theorem 4 reveals that performing least squares after model selection leads to inference that is asymptotically equivalent to inference based on least squares only including the relevant variables. However, it is important to stress that such inference is of a pointwise nature. It is indeed not uniformly valid over *any* ℓ_0 -ball of radius greater than one as the result relies on $\min_{j \in J} |\gamma_j^*| \geq 4S_1$ whose complement has non-zero intersection with every such ℓ_0 ball. To be concrete, assume $\gamma_j^* = S_1$ resulting in $\hat{\gamma}_{PostOLS,j} = 0$ for N sufficiently large. In that case, choosing $\rho = e_j$ one gets that $|\sqrt{N}(\hat{\gamma}_{PostOLS,j} - \gamma_j^*) - \sqrt{N}(\hat{\gamma}_{OLS,j} - \gamma_j^*)| = |\log(N \vee p \vee T) + O_p(1)| \rightarrow \infty$. This non-uniformity of (9) manifests itself in confidence bands that occasionally undercover the true parameter as we shall see in section 6. This finding echoes the warning regarding pointwise inference of Leeb and Pötscher (2005) and we therefore turn next to the construction of uniformly valid confidence bands.

5. UNIFORMLY VALID CONFIDENCE BANDS

We now turn to constructing confidence bands for the elements of γ^* which are uniformly valid, or honest, over all γ^* in certain ℓ_0 -balls. To this end, we extend the desparsification idea of Zhang and Zhang (2014); van de Geer et al. (2014) to correlated random effects panel data models. As we neither impose independence nor stationarity across the time series observations the extension requires a careful analysis of the dependence structure

as well as a uniformly consistent estimator of the covariance matrix of the parameter estimates in this setting. Such an estimator must accommodate that we do not impose the $\epsilon_{i,t}$ to be independent across $t = 1, \dots, T$. First, defining $\mu_{N,T} = \lambda_{N,T}/(NT)$ and $\tilde{p} = p + Tp_v + 1$ the first order conditions of (5) may be written as

$$-Z'(y - Z\hat{\gamma})/(NT) + \mu_{N,T}\hat{\kappa} = 0,$$

$$\|\hat{\kappa}\|_\infty \leq 1,$$

and $\hat{\kappa}_j = \text{sign}(\hat{\gamma}_j)$ if $\hat{\gamma}_j \neq 0$ for $j = 1, \dots, \tilde{p}$. As $y = Z\gamma^* + u$ and defining $\hat{\Gamma} = Z'Z/(NT)$ the above display yields

$$\mu_{N,T}\hat{\kappa} + \hat{\Gamma}(\hat{\gamma} - \gamma^*) = Z'u/(NT).$$

If $\tilde{p} > N$ then $\hat{\Gamma}$ is not invertible and a closed form for $\hat{\gamma} - \gamma^*$ is not available by standard techniques. Assume instead that we have an approximate inverse $\hat{\Theta}$ of $\hat{\Gamma}$. Then, the above display can be rewritten as

$$(10) \quad \hat{\gamma} = \gamma^* - \hat{\Theta}\mu_{N,T}\hat{\kappa} + \hat{\Theta}Z'u/(NT) - \Delta/N^{1/2}, \quad \Delta = \sqrt{N}(\hat{\Theta}\hat{\Gamma} - I_{\tilde{p}})(\hat{\gamma} - \gamma^*)$$

where Δ is the error resulting from using an approximate inverse, $\hat{\Theta}$, of $\hat{\Gamma}$ as opposed to an exact inverse. We shall see that Δ is asymptotically negligible. Note also that the bias term $\hat{\Theta}\mu_{N,T}\hat{\kappa}$ resulting from the penalization of the parameters is known. This suggests removing it by adding it to both sides of (10), resulting in the following estimator:

$$(11) \quad \hat{b} = \hat{\gamma} + \hat{\Theta}\mu_{N,T}\hat{\kappa} = \gamma^* + \hat{\Theta}Z'u/(NT) - \Delta/N^{1/2}.$$

Hence, for any $\tilde{p} \times 1$ vector ρ with $\|\rho\|_2 = 1$ we can consider

$$(12) \quad \sqrt{N}\rho'(\hat{b} - \gamma^*) = \rho'\hat{\Theta}Z'u/(N^{1/2}T) - \rho'\Delta$$

such that a central limit theorem for $\rho'\hat{\Theta}Z'u/(N^{1/2}T)$ and a verification of asymptotic negligibility of $\rho'\Delta$ will yield asymptotic gaussian inference. A leading special case is $\rho = e_j$ where e_j is the j 'th unit vector for $\mathbb{R}^{\tilde{p}}$. Then, (12) reduces to

$$(13) \quad \sqrt{N}(\hat{b}_j - \gamma_j^*) = (\hat{\Theta}Z'u/(N^{1/2}T))_j - \Delta_j.$$

In general, let $H = \{j = 1, \dots, \tilde{p} : \rho \neq 0\}$ be a set of fixed cardinality. Thus, H contains the indices of the coefficients involved in the hypothesis being tested.

5.1. Constructing $\hat{\Theta}$. In this subsection we construct the approximate inverse $\hat{\Theta}$ of $\hat{\Sigma}$. This is done by panel nodewise regressions. The principle is as in van de Geer et al. (2014) but verification of desirable properties must take into account the correlated random effects structure. Let Z_j be the j 'th column in Z and Z_{-j} all columns of Z except for the j 'th one. First, define the nodewise regression estimates

$$(14) \quad \hat{\psi}_j = \underset{r \in \mathbb{R}^{\tilde{p}-1}}{\text{argmin}} \frac{1}{NT} \|Z_j - Z_{-j}r\|^2 + 2\lambda_j \|r\|_{\ell_1}$$

for each $j = 1, \dots, \tilde{p}$ and the λ_j will be made precise later. Using the notation $\hat{\psi}_j = \{\hat{\psi}_{j,k}; k = 1, \dots, \tilde{p}, k \neq j\}$ we define

$$\hat{C} = \begin{pmatrix} 1 & -\hat{\psi}_{1,2} & \cdots & -\hat{\psi}_{1,\tilde{p}} \\ -\hat{\gamma}_{2,1} & 1 & \cdots & -\hat{\psi}_{2,\tilde{p}} \\ \cdots & \cdots & \ddots & \cdots \\ -\hat{\psi}_{\tilde{p},1} & -\hat{\psi}_{\tilde{p},2} & \cdots & 1 \end{pmatrix}.$$

To define $\hat{\Theta}$ we introduce $\hat{T}^2 = \text{diag}(\hat{\tau}_1^2, \dots, \hat{\tau}_{\tilde{p}}^2)$ which is a $\tilde{p} \times \tilde{p}$ diagonal matrix with

$$(15) \quad \hat{\tau}_j^2 = \frac{1}{NT} \|Z_j - Z_{-j} \hat{\psi}_j\|^2 + \lambda_j \|\hat{\psi}_j\|_1,$$

for all $j = 1, \dots, \tilde{p}$. We now define

$$(16) \quad \hat{\Theta} = \hat{T}^{-2} \hat{C}.$$

³ It remains to be shown that this $\hat{\Theta}$ is close to being an inverse of $\hat{\Gamma}$. To this end, we define $\hat{\Theta}_j$ as the j 'th row of $\hat{\Theta}$ but understood as a $\tilde{p} \times 1$ vector and analogously for \hat{C}_j . Thus, $\hat{\Theta}_j = \hat{C}_j / \hat{\tau}_j^2$. Denoting by e_j the j 'th $\tilde{p} \times 1$ unit vector, arguments similar to the ones in van de Geer et al. (2014)⁴ relying on the first order conditions of (14) yield that

$$(17) \quad \|\hat{\Theta}_j \hat{\Gamma} - e_j'\|_\infty \leq \frac{\lambda_j}{\hat{\tau}_j^2}.$$

5.2. Asymptotic Properties of $\hat{\Theta}$. In order to show that $\rho' \hat{\Theta} Z' u / (N^{1/2} T)$ is asymptotically gaussian, one needs to understand the limiting behaviour of $\hat{\Theta}$ constructed above. We show that $\hat{\Theta}$ is close to $\Theta = \Gamma^{-1}$ in an appropriate sense. To this end, note that by Yuan (2010)

$$(18) \quad \Theta_{j,j} = \left[\Gamma_{j,j} - \Gamma_{j,-j} \Gamma_{-j,-j}^{-1} \Gamma_{-j,j} \right]^{-1} \quad \text{and} \quad \Theta_{j,-j} = -\Theta_{j,j} \Gamma_{j,-j} \Gamma_{-j,-j}^{-1},$$

where $\Theta_{j,j}$ is the j th diagonal entry of Θ , $\Theta_{j,-j}$ is the $1 \times (\tilde{p} - 1)$ vector obtained by removing the j th entry of the j th row of Θ , $\Gamma_{-j,-j}$ is the submatrix of Γ with the j th row and column removed, $\Gamma_{j,-j}$ is the j th row of Γ with its j th entry removed, $\Gamma_{-j,j}$ is the j th column of Γ with its j th entry removed. Next, let $z_{i,t,j}$ be the j th element of $z_{i,t}$ and $z_{i,t,-j}$ be all elements except the j th. Define the $(\tilde{p} - 1) \times 1$ vector

$$\psi_j := \underset{\delta}{\operatorname{argmin}} \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T E[z_{i,t,j} - z'_{i,t,-j} \delta]^2$$

such that

$$(19) \quad \psi_j = \left(\frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T E[z_{i,t,-j} z'_{i,t,-j}] \right)^{-1} \left(\frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T E[z_{i,t,-j} z_{i,t,j}] \right) = \Gamma_{-j,-j}^{-1} \Gamma_{-j,j}.$$

³A practical benefit is that the nodewise regressions actually only have to be run for $j \in H$ and not all $j = 1, \dots, \tilde{p}$ as we only need to estimate the covariance matrix of those parameters involved in the hypothesis being tested.

⁴The probabilistic analysis of the limiting properties of $\hat{\Theta}$ is, however, quite different from the one in van de Geer et al. (2014).

Therefore, $\Theta_{j,-j} = -\Theta_{j,j}\psi'_j$ showing that $\Theta_{Z,j,-j}$ and ψ'_j only differ by a multiplicative constant. In particular, the j th row of Θ is sparse if and only if ψ_j is sparse. Furthermore, defining $\eta_{j,i,t} := z_{i,t,j} - z'_{i,t,-j}\psi_j$ we may write

$$(20) \quad z_{i,t,j} = z'_{i,t,-j}\psi_j + \eta_{i,t,j}, \quad \text{for } i = 1, \dots, N, \quad t = 1, \dots, T.$$

where by the definition of ψ_j as an L^2 minimizer,

$$(21) \quad \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T E[z_{i,t,-j}\eta_{i,t,j}] = 0.$$

We shall sometimes write the nodewise regression equation (20) in stacked form as $Z_j = Z_{-j}\psi_j + \eta_j$. In light of Theorem 1, it is sensible that the Lasso estimator $\hat{\psi}_j$ defined in (14) is close to the population regression coefficients ψ_j (we shall make this formal in the Appendix). Next, defining

$$\tau_j^2 := E \left[\frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T (z_{i,t,j} - z'_{i,t,-j}\psi_j)^2 \right] = \Gamma_{j,j} - \Gamma_{j,-j}\Gamma_{-j,-j}^{-1}\Gamma_{-j,j} = \frac{1}{\Theta_{j,j}}$$

observe $\Theta_{j,-j} = -\psi'_j/\tau_j^2$. Thus, we can write $\Theta = T^{-1}C$ where $T = \text{diag}(\tau_1^2, \dots, \tau_{\tilde{p}}^2)$ and C is defined similarly to \hat{C} but with ψ_j replacing $\hat{\psi}_j$ for $j = 1, \dots, \tilde{p}$. Finally, let Θ_j denote the j th row of Θ written as a column vector. We shall see that $\hat{\psi}_j$ and $\hat{\tau}_j^2$ are close to ψ_j and τ_j^2 , respectively such that $\hat{\Theta}_j$ is close to Θ_j (see Lemma 7 in the appendix).

5.3. Confidence bands. To present the uniformly valid confidence bands in the correlated random effects setting define $\Gamma_{zu} = \frac{1}{T^2} E \left(\sum_{t=1}^T z_{1,t}u_{1,t} \right) \left(\sum_{t=1}^T z_{1,t}u_{1,t} \right)'$ and its estimator $\hat{\Gamma}_{zu} = \frac{1}{N} \sum_{i=1}^N \left[\frac{1}{T^2} \left(\sum_{t=1}^T z_{i,t}\hat{u}_{i,t} \right) \left(\sum_{t=1}^T z_{i,t}\hat{u}_{i,t} \right)' \right]$ where $\hat{u}_{i,t} = y_{i,t} - z'_{i,t}\hat{\gamma}$ are the feasible empirical residuals. Choose $\lambda_j \asymp \sqrt{\frac{\log(\tilde{p})}{N}}$ for each $j \in H$. Let s_j denote the number of non-zero elements of Θ_j , $s = |J|$, and recall $\tilde{p} = p + Tp_v + 1$. Then we impose

- A3 a) $\sqrt{\frac{\log(\tilde{p} \vee N)^5}{N}} s_j \rightarrow 0$
 b) $\frac{\log(\tilde{p} \vee N)}{N^{1/2}} s^{1/2} s_j \rightarrow 0$
 c) $\frac{\log(\tilde{p})}{N^{1/2}} s \rightarrow 0$
 d) $\phi_{\min}(\Gamma)$ and $\phi_{\min}(\Gamma_{zy})$ are bounded away from zero. $\phi_{\max}(\Gamma)$ and $\phi_{\max}(\Gamma_{zy})$ are bounded from above. $|H|$ is bounded.

Assumption A3 limits the dimension of the models that can be handled. As \tilde{p} only enters logarithmically, A3 can be valid even when \tilde{p} is much larger than N . Note, however, that neither s nor s_j can grow faster than $N^{1/2}$. This requirement is similar to what is needed in the plain cross section model in van de Geer et al. (2014). Finally, we remark that we have only assumed that the number of variable involved in the hypothesis to be tested ($|H|$) to be bounded to keep expressions simple. Our theory can go through even when $|H|$ tends to infinity slower than $N^{1/2}$. Define $\mathcal{B}_{\ell_0}(s) = \{\|\gamma^*\|_{\ell_0} \leq s\}$ and let $\Phi(t)$ be the cdf of the standard normal distribution. Then,

Theorem 5. *Let Assumptions 1-3 be satisfied. Then, for all $\rho \in \mathbb{R}^{\tilde{p}}$ with $\|\rho\|_2 = 1$,*

$$(22) \quad \sup_{t \in \mathbb{R}} \sup_{\gamma^* \in \mathcal{B}_{\ell_0}(s)} \left| P \left(\frac{N^{1/2} \rho'(\hat{b} - \gamma^*)}{\sqrt{\rho' \hat{\Theta} \hat{\Gamma}_{zu} \hat{\Theta} \rho}} \leq t \right) - \Phi(t) \right| \rightarrow 0.$$

Next, $\gamma' \hat{\Theta} \hat{\Gamma}_{zu} \hat{\Theta}' \gamma$ converges uniformly to the asymptotic variance of $N^{1/2} \rho'(\hat{b} - \gamma^*)$, i.e.

$$(23) \quad \sup_{\gamma^* \in \mathcal{B}_{\ell_0}(s)} |\rho' \hat{\Theta} \hat{\Gamma}_{zu} \hat{\Theta}' \rho - \rho' \Theta \Gamma_{zu} \Theta' \rho| = o_p(1).$$

Furthermore, letting $\hat{\sigma}_j = \sqrt{e_j' \hat{\Theta} \hat{\Gamma}_{zu} \hat{\Theta}' e_j}$ (corresponding to $\rho = e_j$ in (22)) and $z_{1-\delta/2}$ be the $1 - \delta/2$ percentile of the standard normal distribution, one has for all $j = 1, \dots, \tilde{p}$

$$(24) \quad \lim_{N \rightarrow \infty} \inf_{\gamma^* \in \mathcal{B}_{\ell_0}(s)} P \left(\gamma_j^* \in \left[\hat{b}_j - z_{1-\delta/2} \frac{\hat{\sigma}_j}{\sqrt{N}}, \hat{b}_j + z_{1-\delta/2} \frac{\hat{\sigma}_j}{\sqrt{N}} \right] \right) = 1 - \delta.$$

Finally, letting $\text{diam}([a, b]) = b - a$ be the length of an interval $[a, b]$ in the real line, we have that

$$(25) \quad \sup_{\gamma^* \in \mathcal{B}_{\ell_0}(s)} \text{diam} \left(\left[\hat{b}_j - z_{1-\delta/2} \frac{\hat{\sigma}_j}{\sqrt{N}}, \hat{b}_j + z_{1-\delta/2} \frac{\hat{\sigma}_j}{\sqrt{N}} \right] \right) = O_p \left(\frac{1}{\sqrt{N}} \right).$$

First, (22) reveals that $\frac{N^{1/2} \rho'(\hat{b} - \gamma^*)}{\sqrt{\rho' \hat{\Theta} \hat{\Gamma}_{zu} \hat{\Theta}' \rho}}$ converges to the standard normal distribution uniformly over the ℓ_0 -ball of radius at most s . We stress again that we have not restricted the dependence structure over time of neither the covariates nor the error terms and the error terms are also allowed to be heteroskedastic. The joint asymptotic normality in (22) also allows one to construct Wald tests. To be precise, for any $H \subseteq \{1, \dots, \tilde{p}\}$,

$$(26) \quad \left\| (\hat{\Theta} \hat{\Gamma}_{zu} \hat{\Theta}')_H^{-1/2} \sqrt{N} (\hat{b}_H - \gamma_H^*) \right\|_2^2 \xrightarrow{d} \chi^2(h),$$

where $h = |H|$. Wald tests test of general smooth hypotheses can be constructed in the usual way by means of the delta method.

(24) is a consequence of (22) and entails that the confidence band $[\hat{b}_j - z_{1-\delta/2} \frac{\hat{\sigma}_j}{\sqrt{N}}, \hat{b}_j + z_{1-\delta/2} \frac{\hat{\sigma}_j}{\sqrt{N}}]$ is *uniformly valid* for γ_j^* over $\mathcal{B}_{\ell_0}(s)$. Uniform validity is important to produce practically useful confidence sets as it ensures that there is a known time N , *not depending on* γ^* , after which the coverage rate of the confidence set is not much smaller than $1 - \delta$. Thus, pointwise confidence bands that do not satisfy (24) but

$$\inf_{\gamma^* \in \mathcal{B}_{\ell_0}(s)} \lim_{N \rightarrow \infty} P \left(\gamma_j^* \in \left[\hat{b}_j - z_{1-\delta/2} \frac{\hat{\sigma}_j}{\sqrt{N}}, \hat{b}_j + z_{1-\delta/2} \frac{\hat{\sigma}_j}{\sqrt{N}} \right] \right) = 1 - \delta,$$

are of less practical use since the N from which point and onwards the coverage is close to $1 - \delta$ is allowed to depend on the unknown γ^* . Of course, a uniformly valid confidence set S_N could also easily be produced by setting $S_N = \mathbb{R}$ for all $N \geq 1$. Such a confidence set is clearly of little practical use. Thus, (25) is important as it reveals that the confidence band $[\hat{b}_j - z_{1-\delta/2} \frac{\hat{\sigma}_j}{\sqrt{N}}, \hat{b}_j + z_{1-\delta/2} \frac{\hat{\sigma}_j}{\sqrt{N}}]$ uniformly has the optimal rate of contraction $1/\sqrt{N}$. Recall that the asymptotic result in Theorem 4 was of a pointwise nature. We shall illustrate in the simulations that the difference between these and the above uniformly valid ones are not just theoretical as the pointwise confidence bands can sometimes seriously undercover the true parameter.

6. MONTE CARLO

In this section we investigate the finite sample properties of the proposed procedures by means of Monte Carlo experiments. The Lasso is implemented using the publicly available `glmnet` package for R.

$\lambda_{N,T}$ and λ_j are chosen by GIC as proposed by Fan and Tang (2013). To provide a benchmark for the Lasso and the thresholded Lasso, least squares including all variables is also implemented whenever feasible. This procedure is denoted OLSA and corresponds to the classical implementation of the Mundlak-Chamberlain device as in, e.g., Wooldridge (2002). At the other extreme, least squares *only* including the relevant variables is applied to provide an infeasible target which we are ideally aiming at. This procedure is called the OLS Oracle (OLSO). We measure the performance of the proposed estimators along the following dimensions

- (1) The average ℓ_2 -estimation error of the parameter estimates of β^* and ϕ^* .
- (2) The average ℓ_∞ -estimation error of the parameter estimates of β^* and ϕ^* .
- (3) How often is the true model included in the model chosen. This measure is reported for β^* as well as ϕ^* .
- (4) How often is the correct sparsity pattern uncovered, i.e. how often is exactly the correct model chosen. This measure is reported for β^* as well as ϕ^* .
- (5) What is the mean number of non-zero parameters in the estimated model. This is reported for β^* as well as ϕ^* .
- (6) Size: We evaluate the size of the χ^2 -test in (26) for a hypothesis involving more than one parameter.
- (7) Power: We evaluate the power of a χ^2 -test in (26) for a hypothesis involving more than one parameter.
- (8) Coverage rate: We calculate the coverage rate of a gaussian confidence interval constructed as in (24). This is done for the coefficient for a time invariant and a time varying covariate.
- (9) Length of confidence interval: We calculate the length of the two confidence intervals considered in point (8) above.

The data is generated from (1) and the error terms are $N(0,1)$ unless mentioned otherwise. One time invariant covariate is included, thus $p_v = p - 1$. To be precise, we generate the covariates as follows. Let $r_{i,t} = (\zeta_{i,t}, \xi_{i,t}')' \sim N_p(0, \Omega)$ be $p \times 1$ and uncorrelated over time. $\xi_{i,t}$ is $p_v \times 1$ and $\Omega_{l,k} = 0.5^{|l-k|}$. In all experiments $v_{i,t} = av_{i,t-1} + \xi_{i,t}$, $a \in (-1, 1)$ while $w_i = \zeta_{i,1}$. This construction ensures that the time invariant and time varying covariates are correlated.

All tests are carried out at a 5% significance level and all confidence intervals are at the 95% level. The χ^2 -tests always involve the two first parameters in γ^* of which we deliberately make sure that first one is non-zero (equaling one) and the second one is zero. Thus, the χ^2 -test involves the coefficient of the time invariant regressor and of the first time varying regressor. For measuring the power of the χ^2 -test, we test the false hypothesis $H_0 : (\gamma_1^*, \gamma_2^*) = (1, 0.25)$. The hypothesis is only false on the second entry of γ^* . Similarly, we construct confidence intervals for the first two parameters of γ^* such that the coverage rate can be compared between the coefficients of the time invariant and time varying regressors.

Remark: Conventions and definitions. All results regarding χ^2 -tests and confidence intervals for the Lasso are based on the desparsified Lasso. For the OLS oracle the joint χ^2 -test is not carried out as it by construction leaves out all zero coefficients from the outset. We use the convention that the confidence band for the zero coefficient is a point mass at zero. All results relating to tests and confidence intervals of the thresholded Lasso are based on post selection least squares estimation as detailed in Section 4.2. Again the

joint χ^2 -test is not carried out as thresholding may have eliminated one of the coefficients involved in the hypothesis. If a variable is excluded by thresholding we use the convention that the confidence band for its coefficient is a point mass at zero. We stress that these are just conventions and regarding tests and confidence bands one can safely focus only on the results for the desparsified Lasso and OLSA if one disagrees on these conventions.

The following experiments are carried out to gauge the performance along the above dimensions (the number of Monte Carlo replications is always 1000).

Recall that the effective sample size is N and the dimension of the parameter vector should be compared to this. ϕ^* has Tp_v entries where $p_v = p - 1$ in all experiments. Recall also that the Mundlak-Chamberlain device is infeasible in general when $p_v T > N$. Thus, in order to be able to compare to the classical least squares implementation of the Mundlak-Chamberlain device, we start with some settings where $p_v T \leq N$.

- Experiment A: $N = 100$ and $T = 5$ with β^* having two equidistant entries of 1 and 8 of zero (thus, $p = 10$, $p_v = 9$). ϕ^* has the first 5 entries equal to 1 and the last $Tp_v - 5 = 40$ equal to zero. $a = 0.5$
- Experiment B: As experiment A but with the last 40 entries of ϕ^* equal to 0.05.
- Experiment C: As experiment B but with $a = 0.9$.
- Experiment D: As experiment C but with $\epsilon_{i,t}$ and $r_{i,t}$ being $t(3)$ -distributed.

The above experiments become gradually harder and we have deliberately chosen p moderately small since otherwise the classical Mundlak-Chamberlain device (OLSA) is not even feasible. The following experiments increase the dimension of β^* . Note that even the oracle easily becomes infeasible as is the case in experiment G.

- Experiment E: $N = 100, T = 5$. β^* has 4 equidistant nonzero entries equaling one and 36 entries equaling 0. The first 5 entries of ϕ^* equal 1. The last $Tp_v - 5 = 190$ entries equal 0. $a = 0.5$. Note that OLSA is not feasible.
- Experiment F: As experiment E but with the last 190 entries of ϕ^* equal to 0.01. Note that now not even the oracle is feasible as ϕ^* is too dense.
- Experiment G: $N = 400, T = 5$. β^* has 4 equidistant nonzero entries equaling one and 36 entries equaling 0. The first 5 entries of ϕ^* equal 1. The remaining $Tp_v - 5 = 190$ entries equal 0.01. The reason for choosing $N = 400$ is to make OLSA feasible again.
- Experiment H: As experiment G but with all non-zero coefficients equaling 0.01 instead of 1. The point is to illustrate the difference between pointwise and uniformly valid confidence bands by considering small, yet nonzero coefficients.

6.1. Results. Experiment A shows that as expected the estimation error of the Lasso is between the one of the oracle and full least squares. The thresholded Lasso selects the correct sparsity pattern for β^* as well as ϕ^* almost all the time. The power of the χ^2 -test based on the desparsified Lasso is high and its size is not much higher than the nominal one (and certainly lower than the one from the classical implementation of the Mundlak-Chamberlain device). The coverage rate of the confidence intervals is close to 95% as desired – perhaps with a slight undercoverage for the coefficient of the time invariant covariate (see TI column). Note that the confidence interval of the time varying coefficient is around twice as wide as the one of the time varying regressor; a finding which will reappear in the remaining simulations. The reason for this is that even though the asymptotic theory only utilizes the variation over N , in practice the time varying

regressors show more variation than the time-invariant ones as the former vary over time as well.

Experiment B changes ϕ^* such that it has no zero entries. The thresholded Lasso now estimates β^* almost as precisely as the infeasible oracle and both Lasso based estimators estimate ϕ^* more precisely than the oracle. The thresholded Lasso still selects the correct sparsity pattern of β^* every time but no longer detects the sparsity pattern of ϕ^* due to the presence of many small albeit non-zero coefficients. The size distortion of the χ^2 -test has increased for the desparsified conservative Lasso but it remains below the one for OLSA.

Experiment C increases the dependence over time of the covariates. This results in slightly less precise parameter estimates. Furthermore, the sparsity pattern of β^* is detected less often by the thresholded Lasso. Overall, the results are robust to increased time series dependence. The reason for this is that all variables in Z pertaining to the Mundlak-Chamberlain projection are already identical over time. The confidence bands of the desparsified Lasso become slightly wider resulting in superior coverage of the coefficient of the time invariant regressor compared to all other methods.

Experiment D adds heavy tails to experiment C. The Lasso based procedures are not affected much by this while the oracle and full OLS deteriorate in particular when it comes to estimating ϕ^* . They are now much less precise in this respect than the Lasso and the thresholded Lasso. Furthermore, the variable selection capability of the latter is not affected by heavy tails. Note that the power of the χ^2 -test is reduced as the χ^2 -distribution is no longer a good approximation with $N = 100$ observations and heavy tailed variables. However, the performance remains above the one of OLSA. Finally, there is only slight undercoverage for the confidence intervals produced by the desparsified Lasso while the length of all bands increases in the presence of heavy tails.

Experiment E is a very ill-conditioned one in which $N > p_v T$ such that OLSA is not feasible. However, the model is exactly sparse such that the oracle can exclude many irrelevant variables. This explains its lower estimation error compared to the Lasso based methods. The thresholded Lasso always selects the correct sparsity pattern as the zero and non-zero coefficients are well-separated. For the same reason, the non-uniformity of the asymptotics of the thresholded Lasso do not result in confidence bands with undercoverage even though the bands are as narrow as those of the oracle.

In experiment F not even OLSO is feasible. The thresholded Lasso still does a good job in detecting the sparsity pattern of β^* . All numbers regarding size, power and confidence bands are reasonable. The confidence band for the time invariant variable is still wider than that for the time varying variable.

In experiment G $N = 400$ to make all procedures feasible. The thresholded Lasso detects the correct sparsity pattern of β^* almost all the time but does a poor job on ϕ^* as its non-zero coefficients are very small. The χ^2 -tests based on the desparsified Lasso perform well in terms of size and power and the confidence intervals have coverage close to the nominal rate.

In experiment H all the non-zero coefficients are very close to zero. This reduces the estimation error of the Lasso of β^* as well as ϕ^* which is now lower than the one of the oracle. However, it is detrimental to the screening and variable selection properties of the Lasso and the thresholded Lasso. None of these is able to even retain the relevant variables. This directly materializes in zero coverage of the pointwise post-thresholding least squares based confidence bands. On the other hand, uniformly valid confidence bands based on

the desparsified Lasso are not affected by the poor variable selection properties of the first step Lasso – they only utilize the good first step estimation precision. Therefore, the confidence bands based on the desparsified Lasso have excellent coverage and are as narrow as those of the oracle.

7. CONCLUSION

This paper has established ℓ_2 oracle inequalities for the Lasso in high-dimensional panel data models under correlated random effects. Importantly, we allowed for the inclusion of time invariant regressors and correlation between all regressors and the unobserved heterogeneity. This strikes a middle ground between fixed and random effects. A weak sparsity assumption on the coefficients of the Mundlak-Chamberlain projection and an ℓ_∞ oracle inequality not depending on the dimension of the model was given. This allowed for sharp thresholding results and was derived in a manner quite different from the one used to establish classical ℓ_1 and ℓ_2 oracle inequalities. Pointwise asymptotic inference results were established for post thresholding least squares. Finally, we showed how the Lasso may be desparsified in the correlated random effects setting and how this can be used for uniformly valid inference under heteroskedasticity and dependence.

Avenues for future research include extending our results to settings with dependence across individuals as often encountered in, e.g., finance. Furthermore, it is of interest to consider non-linear panel data models.

	Estimation and variable selection														χ^2		Coverage rate		Length			
	$\ell_2(\beta)$				$\ell_\infty(\beta)$				Sub(β)		Spar(β)		Spar(ϕ)		# β	# ϕ	Size	Power	TI	TV	TI	TV
	$\ell_2(\phi)$	$\ell_2(\beta)$	$\ell_\infty(\phi)$	$\ell_\infty(\beta)$	$\ell_\infty(\phi)$	Sub(β)	Sub(ϕ)	Spar(β)	Spar(ϕ)													
Exp A	Lasso	0.20	0.37	0.18	0.23	1.00	1.00	1.00	0.65	0.04	2.43	11.08	0.08	0.99	0.90	0.96	0.49	0.22				
	TLasso	0.20	0.31	0.18	0.22	1.00	1.00	1.00	1.00	0.94	2.00	5.06		0.93	1.00	1.00	0.47	0.00				
	OLSO	0.12	0.26	0.11	0.19	1.00	1.00	1.00	1.00	1.00	2.00	5.00		0.93	1.00	1.00	0.47	0.00				
	OLSA	0.25	1.35	0.17	0.50	1.00	1.00	1.00	0.00	0.00	10.00	45.00	0.15	0.98	0.84	0.94	0.47	0.23				
Exp B	Lasso	0.17	0.58	0.15	0.27	1.00	1.00	0.00	0.17	0.00	3.67	22.31	0.12	0.99	0.86	0.96	0.45	0.22				
	TLasso	0.16	0.54	0.15	0.27	1.00	1.00	0.00	1.00	0.00	2.00	7.26		0.89	1.00	1.00	0.50	0.00				
	OLSO	0.15	1.34	0.14	0.50	1.00	1.00	1.00	1.00	1.00	2.00	45.00		0.84	1.00	1.00	0.47	0.00				
	OLSA	0.25	1.35	0.17	0.50	1.00	1.00	1.00	0.00	1.00	10.00	45.00	0.15	0.98	0.84	0.94	0.47	0.23				
Exp C	Lasso	0.32	0.57	0.30	0.27	0.90	0.90	0.00	0.10	0.00	3.73	23.67	0.09	0.99	0.91	0.95	0.51	0.20				
	TLasso	0.31	0.58	0.30	0.27	0.89	0.90	0.00	0.81	0.00	1.98	14.13		0.77	0.99	1.00	0.37	0.00				
	OLSO	0.13	1.52	0.13	0.57	1.00	1.00	1.00	1.00	1.00	2.00	45.00		0.84	1.00	1.00	0.43	0.00				
	OLSA	0.24	1.52	0.16	0.57	1.00	1.00	1.00	0.00	1.00	10.00	45.00	0.13	0.98	0.84	0.94	0.43	0.23				
Exp D	Lasso	0.37	0.79	0.33	0.39	0.95	0.95	0.00	0.19	0.00	3.53	20.72	0.10	0.78	0.87	0.95	0.79	0.34				
	TLasso	0.36	0.75	0.33	0.39	0.93	0.93	0.00	0.85	0.00	2.02	10.71		0.80	1.00	1.00	0.66	0.00				
	OLSO	0.24	2.78	0.23	1.05	1.00	1.00	1.00	1.00	1.00	2.00	45.00		0.83	1.00	1.00	0.76	0.00				
	OLSA	0.43	2.79	0.29	1.05	1.00	1.00	1.00	0.00	1.00	10.00	45.00	0.15	0.71	0.83	0.94	0.76	0.39				
Exp E	Lasso	0.42	0.42	0.29	0.27	1.00	1.00	1.00	0.64	0.00	4.47	14.73	0.07	0.98	0.92	0.96	0.71	0.23				
	TLasso	0.42	0.36	0.29	0.27	1.00	1.00	1.00	1.00	0.98	4.00	5.02		0.92	1.00	1.00	0.46	0.00				
	OLSO	0.16	0.26	0.13	0.19	1.00	1.00	1.00	1.00	1.00	4.00	5.00		0.92	1.00	1.00	0.47	0.00				
Exp F	Lasso	0.38	0.49	0.28	0.28	1.00	1.00	0.00	0.41	0.00	4.99	20.07	0.08	0.98	0.90	0.97	0.70	0.22				
	TLasso	0.38	0.40	0.28	0.27	1.00	1.00	0.00	1.00	0.00	4.00	5.09		0.91	1.00	1.00	0.49	0.00				
Exp G	Lasso	0.18	0.26	0.13	0.12	1.00	1.00	0.00	0.32	0.00	5.25	25.80	0.06	1.00	0.93	0.96	0.23	0.11				
	TLasso	0.18	0.22	0.13	0.12	1.00	1.00	0.00	1.00	0.00	4.00	5.13		0.95	1.00	1.00	0.25	0.00				
	OLSO	0.08	1.48	0.07	0.31	1.00	1.00	1.00	1.00	1.00	4.00	195.00		0.84	1.00	1.00	0.24	0.00				
	OLSA	0.23	1.48	0.10	0.31	1.00	1.00	1.00	0.00	1.00	40.00	195.00	0.14	1.00	0.84	0.94	0.24	0.12				
Exp H	Lasso	0.02	0.15	0.01	0.03	0.00	0.00	0.00	0.00	0.00	0.03	3.26	0.04	1.00	0.96	0.96	0.24	0.11				
	TLasso	0.02	0.14	0.01	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.02		0.00	1.00	1.00	0.00	0.00				
	OLSO	0.08	1.48	0.07	0.31	1.00	1.00	1.00	1.00	1.00	4.00	195.00		0.84	1.00	1.00	0.24	0.00				
	OLSA	0.23	1.48	0.10	0.31	1.00	1.00	1.00	0.00	1.00	40.00	195.00	0.14	1.00	0.84	0.94	0.24	0.12				

TABLE 1. $\ell_2(\beta)$ and $\ell_2(\phi)$ are the average ℓ_2 errors of the parameter estimates. $\ell_\infty(\beta)$ and $\ell_\infty(\phi)$ are the sup-norm estimation errors. Sub(β) and Sub(ϕ): fraction of times the estimated model contains all the relevant variables pertaining to these coefficients. Spar(β) and Spar(ϕ) show how often exactly the correct subset of variables is chosen. # β and # ϕ give the average number of non-zero β s and ϕ s, respectively. Size and Power of the χ^2 test are reported as well as the coverage rate of confidence intervals of the time invariant covariate (TI) (whose coefficient equals 1) and of a time varying covariate (TV) (whose coefficient equals 0). Finally, the lengths of the same confidence intervals are given.

8. APPENDIX

8.1. Oracle inequalities and variables selection by thresholding. As is customary when establishing oracle inequalities we start with a deterministic bound which is valid on a certain set. This bound will follow from a result in Negahban et al. (2012) upon reparameterizing the objective function. After stating the deterministic bound, we use the structure of the panel data to provide a lower bound on the probability of this set. This is where the real work lies.

Deterministic bound: Note that $\hat{\gamma}$ is a minimizer of

$$(27) \quad L(\gamma) = \frac{1}{2NT} \|y - Z\gamma\|^2 + \mu_{N,T} \|\gamma\|_{\ell_1}$$

where $\mu_{N,T} = \lambda_{N,T}/(NT)$. As $\|\gamma_J + \gamma_{J^c}\|_{\ell_1} = \|\gamma_J\|_{\ell_1} + \|\gamma_{J^c}\|_{\ell_1}$ for any $\gamma \in \mathbb{R}^{p+Tp_v+1}$ and $J \subseteq \{1, \dots, p + Tp_v + 1\}$, the ℓ_1 -norm is decomposable with respect to J in the terminology of Negahban et al. (2012). Therefore, on

$$\mathcal{A} = \left\{ \frac{1}{NT} \|Z'u\|_{\ell_\infty} \leq \mu_{N,T}/2 \right\}$$

it follows by Lemma 1 in Negahban et al. (2012) that for any $J \subseteq \{1, \dots, p + Tp_v + 1\}$ one has that $\hat{\gamma} - \gamma^*$ belongs to the set

$$\mathcal{C}(J, \gamma^*) = \left\{ \Delta \in \mathbb{R}^{p+Tp_v+1} : \|\Delta_{J^c}\|_{\ell_1} \leq 3\|\Delta_J\|_{\ell_1} + 4\|\gamma_{J^c}^*\|_{\ell_1} \right\}.$$

If, furthermore,

$$(28) \quad \frac{1}{NT} \Delta' Z' Z \Delta \geq \kappa \|\Delta\|^2 - \tau^2(\gamma^*) \quad \text{for all } \Delta \in \mathcal{C}(J, \gamma^*)$$

for $\kappa > 0$ and τ a function of γ^* , then Theorem 1 of Negahban et al. (2012) yields that

$$(29) \quad \|\hat{\gamma} - \gamma^*\|^2 \leq 9 \frac{\mu_{N,T}^2}{\kappa^2} |J| + \frac{\mu_{N,T}}{\kappa} (2\tau^2(\gamma^*) + 4\|\gamma_{J^c}^*\|_{\ell_1}).$$

The hard part now consists of providing good values of $\mu_{N,T}, \kappa$ and $\tau(\gamma^*)$. We shall also utilize that we are free to choose the set J . We first provide a lower bound on the probability of \mathcal{A} .

Lemma 2. Assume that assumptions A1 and A2 are satisfied and set $\mu_{N,T} = \sqrt{\frac{16a_{N,T} \log[p+Tp_v+1]}{N}}$. Then, $P(\mathcal{A}) \geq 1 - A[p + Tp_v + 1]^{1-Ba_{N,T}}$.

Proof. First, note that $\frac{1}{NT} \|Z'u\|_{\ell_\infty} \leq \frac{1}{NT} \|Z'\epsilon\|_{\ell_\infty} + \frac{1}{NT} \|Z'a\|_{\ell_\infty}$. Consider one of the last p entries entry of $\frac{1}{NT} Z'a$. Such an entry is of the form $\frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T a_i x_{i,t,k}$ for some $k = 1, \dots, p$. Since a_i and $x_{i,t,k}$ are uniformly subgaussian for all i, t and k it follows that $a_i x_{i,t,k}$ is uniformly subexponential with mean zero which, in turn, implies that the same is the case for $\frac{1}{T} \sum_{t=1}^T a_i x_{i,t,k}$. A similar argument applies to the first $1 + Tp_v$ entries of $\frac{1}{NT} Z'a$. Thus, by a union bound and Corollary 5.17 in Vershynin (2012), it follows that there exists positive constants A, B such that

$$P\left(\frac{1}{NT} \|Z'a\|_{\ell_\infty} \geq \mu_{N,T}/4\right) \leq A[p + Tp_v + 1] e^{-Ba_{N,T} \frac{\log[p+Tp_v+1]}{N} N} = A[p + Tp_v + 1]^{1-Ba_{N,T}}.$$

$P\left(\frac{1}{NT} \|Z'\epsilon\|_{\ell_\infty} \geq \mu_{N,T}/4\right)$ can be bounded by the same quantity using the same technique.

Thus, $P\left(\frac{1}{NT} \|Z'u\|_{\ell_\infty} \geq \mu_{N,T}/2\right) \leq A[p + Tp_v + 1]^{1-Ba_{N,T}}$. \square

We next turn to condition (28), called the restricted strong convexity condition. The following Lemma will be useful.

Lemma 3. *For all $\Delta \in \mathcal{C}(J, \gamma^*)$ one has*

$$\frac{1}{NT} \Delta' Z' Z \Delta \geq \Delta' \Gamma \Delta - 32 \left\| \frac{1}{NT} Z' Z - \Gamma \right\|_{\infty} (|J| \|\Delta\|^2 + \|\gamma_{J^c}^*\|_{\ell_1}^2).$$

Proof. Define $\Gamma_{N,T} = \frac{1}{NT} Z' Z$ and observe that for all $\Delta \in \mathcal{C}(J, \gamma^*)$

$$\begin{aligned} \|\Delta\|_{\ell_1} &= \|\Delta_J\|_{\ell_1} + \|\Delta_{J^c}\|_{\ell_1} \leq 4 \|\Delta_J\|_{\ell_1} + 4 \|\gamma_{J^c}^*\|_{\ell_1} \leq 4\sqrt{|J|} \|\Delta_J\| + 4 \|\gamma_{J^c}^*\|_{\ell_1} \\ &\leq 4\sqrt{|J|} \|\Delta\| + 4 \|\gamma_{J^c}^*\|_{\ell_1} \end{aligned}$$

such that

$$\Delta' \Gamma_{N,T} \Delta \geq \Delta' \Gamma \Delta - \|\Gamma_{N,T} - \Gamma\|_{\infty} \|\Delta\|_{\ell_1}^2 \geq \Delta' \Gamma \Delta - 32 \|\Gamma_{N,T} - \Gamma\|_{\infty} (|J| \|\Delta\|^2 + \|\gamma_{J^c}^*\|_{\ell_1}^2).$$

□

Define

$$\mathcal{B} = \left\{ \left\| \frac{1}{NT} Z' Z - \Gamma \right\|_{\infty} \leq \sqrt{\frac{a_{N,T} \log[p + Tp_v + 1]}{N}} \right\}$$

Lemma 4. *Under assumptions A1 and A2) one has that*

$$P(\mathcal{B}) \geq 1 - A[p + Tp_v + 1]^{2-Ba_{N,T}}$$

for positive constants A and B .

Proof. Consider an element of the lower right hand $p \times p$ block of $\frac{1}{NT} Z' Z - \Gamma$ (a similar argument applies to the remaining entries with slightly different notation). Such an element is on the form $\frac{1}{N} \sum_{i=1}^N \left(\frac{1}{T} \sum_{t=1}^T [x_{i,t,k} x_{i,t,l} - E(x_{i,t,k} x_{i,t,l})] \right)$ for some $k, l \in \{1, \dots, p\}$. A small calculation shows that $\frac{1}{T} \sum_{t=1}^T [x_{i,t,k} x_{i,t,l} - E(x_{i,t,k} x_{i,t,l})]$ is subexponential for all $1 \leq i \leq N$ and $1 \leq k, l \leq p$. By the independence across $i = 1, \dots, N$ we may apply Corollary 5.17 in Vershynin (2012) to conclude that there exist constants A and B such that for $\epsilon = \sqrt{\frac{a_{N,T} \log[p + Tp_v + 1]}{N}}$,

$$\begin{aligned} &P \left(\left| \frac{1}{N} \sum_{i=1}^N \left(\frac{1}{T} \sum_{t=1}^T [x_{i,t,k} x_{i,t,l} - E(x_{i,t,k} x_{i,t,l})] \right) \right| \geq \epsilon \right) \\ (30) \quad &\leq A e^{-B(\epsilon^2 N)} = A e^{-Ba_{N,T} \log[p + Tp_v + 1]}. \end{aligned}$$

Next, via a union bound over $[p + Tp_v + 1]^2$ terms

$$P \left(\|\Gamma_{N,T} - \Gamma\|_{\infty} > \epsilon \right) \leq A[p + Tp_v + 1]^2 e^{-Ba_{N,T} \log(p^2 + Np)} = A[p + Tp_v + 1]^{2-Ba_{N,T}}.$$

□

Proof of Theorem 1. Set $J = J_1$. Then we conclude from Lemma 3 that for all $\Delta \in \mathcal{C}(J, \gamma^*)$ one has on \mathcal{B}

$$\begin{aligned}
 \frac{1}{NT} \Delta' Z' Z \Delta &\geq \Delta' \Gamma \Delta - 32 \left\| \frac{1}{NT} Z' Z - \Gamma \right\|_{\infty} (|J_1| \|\Delta\|^2 + \|\gamma_{J_1}^*\|_{\ell_1}^2) \\
 &\geq \phi_{\min}(\Gamma) \|\Delta\|^2 - 32 \sqrt{\frac{a_{N,T} \log(p + Tp_v + 1)}{N}} (|J_1| \|\Delta\|^2 + R^2) \\
 (31) \quad &\geq \frac{\phi_{\min}(\Gamma)}{2} \|\Delta\|^2 - 32 \sqrt{\frac{a_{N,T} \log(p + Tp_v + 1)}{N}} R^2
 \end{aligned}$$

where the last estimate follows from $32 \sqrt{\frac{a_{N,T} \log(p + Tp_v + 1)}{N}} |J_1| \leq \phi_{\min}(\Gamma)/2$. Thus, in (29) we can set $\kappa = \phi_{\min}(\Gamma)/2$, $\tau^2 = 32 \sqrt{\frac{a_{N,T} \log(p + Tp_v + 1)}{N}} R^2$, $\|\gamma_{J_1}^*\|_{\ell_1} = R$, $\mu_{N,T} = \sqrt{\frac{16a_{N,T} \log[p + Tp_v + 1]}{N}}$ and conclude that

$$\begin{aligned}
 \|\hat{\gamma} - \gamma^*\|^2 &\leq 576 \frac{a_{N,T} \log[p + Tp_v + 1]}{\phi_{\min}^2(\Gamma) N} |J_1| \\
 &\quad + \sqrt{\frac{64a_{N,T} \log[p + Tp_v + 1]}{\phi_{\min}^2(\Gamma) N}} \left(64 \sqrt{\frac{a_{N,T} \log(p + Tp_v + 1)}{N}} R^2 + 4R \right) \\
 (32) \quad &\leq 576 \frac{a_{N,T} \log[p + Tp_v + 1]}{N} |J_1| + 8 \sqrt{\frac{64a_{N,T} \log(p + Tp_v + 1)}{\phi_{\min}^2(\Gamma) N}} R.
 \end{aligned}$$

where the second inequality used that $16 \sqrt{\frac{a_{N,T} \log(p + Tp_v + 1)}{N}} R \leq 1$. Thus, the subadditivity of $x \mapsto \sqrt{x}$ yields

$$(33) \quad \|\hat{\gamma} - \gamma^*\| \leq 24 \sqrt{\frac{a_{N,T} \log(p + Tp_v + 1)}{\phi_{\min}^2(\Gamma) N}} |J_1| + 8 \left(\frac{a_{N,T} \log(p + Tp_v + 1)}{\phi_{\min}^2(\Gamma) N} R^2 \right)^{1/4}$$

The conclusion of the Theorem now follows upon noting that (33) is valid on $\mathcal{A} \cap \mathcal{B}$ whose probability has been bounded from below by Lemmas 2 and 4 upon synchronizing the constants in these lemmas. \square

Proof of Theorem 2. Note first that for all $\Delta \in \mathcal{C}(J, \gamma^*)$ one has that

$$\|\Delta\|_{\ell_1} = \|\Delta_J\|_{\ell_1} + \|\Delta_{J^c}\|_{\ell_1} \leq 4 \|\Delta_J\|_{\ell_1} + 4 \|\gamma_{J^c}^*\|_{\ell_1} \leq 4\sqrt{|J|} \|\Delta\|$$

where we used $\|\gamma_{J^c}^*\|_{\ell_1} = 0$. Therefore, as $\hat{\gamma} - \gamma^* \in \mathcal{C}(J, \gamma^*)$, using the bound in (6) with $R = 0$ yields

$$(34) \quad \|\hat{\gamma} - \gamma^*\|_{\ell_1} \leq 96 \sqrt{\frac{a_{N,T} \log[p + Tp_v + 1]}{\phi_{\min}^2(\Gamma) NT}} |J|$$

on $\mathcal{A} \cap \mathcal{B}$. Note that this set has probability tending to one as $a_{N,T} = \log(p \vee N \vee T)$. Next, the Karush-Kuhn-Tucker condition for the problem (27) read

$$-\frac{1}{NT} Z' (y - Z\hat{\gamma}) + \mu_{N,T} \hat{z} = 0$$

where $\|\hat{z}\|_{\ell_\infty} \leq 1$ and $\hat{z}_j = \text{sign}(\hat{\gamma}_j)$ if $\hat{\gamma}_j \neq 0$. This can be rewritten as

$$\frac{1}{NT} Z' Z (\hat{\gamma} - \gamma^*) = \frac{1}{NT} Z' u - \mu_{N,T} \hat{z}.$$

This is, in turn, equivalent to

$$\Gamma(\hat{\gamma} - \gamma^*) = \left(\Gamma - \frac{1}{NT} Z' Z \right) (\hat{\gamma} - \gamma^*) + \frac{1}{NT} Z' u - \mu_{N,T} \hat{z}$$

such that

$$\|\hat{\gamma} - \gamma^*\|_{\ell_\infty} \leq \|\Gamma^{-1}\|_{\ell_\infty} \left\| \Gamma - \frac{1}{NT} Z' Z \right\|_{\ell_\infty} \|\hat{\gamma} - \gamma^*\|_{\ell_1} + \|\Gamma^{-1}\|_{\ell_\infty} \left\| \frac{1}{NT} Z' u \right\|_{\ell_\infty} + \|\Gamma^{-1}\|_{\ell_\infty} \mu_{N,T}$$

where we used $\|\hat{z}\|_{\ell_\infty} \leq 1$. Next, consider one term at a time in the above display. First, Lemma 4 yields that on \mathcal{B} , $\left\| \Gamma - \frac{1}{NT} Z' Z \right\|_{\ell_\infty} \leq \sqrt{\frac{\log[p+Tp_v+1]a_{N,T}}{N}} \leq C_1 \sqrt{\frac{\log^2[(p \vee N \vee T)]}{N}}$ for some $C_1 > 0$ ⁵. Furthermore, the right hand side of (34) tends to zero. Thus, $\|\hat{\gamma} - \gamma^*\|_{\ell_1} \leq 1/(C_1)$ for N sufficiently large. In total,

$$\|\Gamma^{-1}\|_{\ell_\infty} \left\| \Gamma - \frac{1}{NT} Z' Z \right\|_{\ell_\infty} \|\hat{\gamma} - \gamma^*\|_{\ell_1} \leq \|\Gamma^{-1}\|_{\ell_\infty} \sqrt{\frac{\log^2[(p \vee N \vee T)]}{N}}.$$

Furthermore, on \mathcal{A} , with $\mu_{N,T} = \sqrt{\frac{16a_{N,T} \log(p \vee N \vee T)}{N}}$ a small modification of the proof of Lemma 2 yields that $\left\| \frac{1}{NT} Z' u \right\|_{\ell_\infty} \leq \sqrt{\frac{16a_{N,T} \log(p \vee N \vee T)}{N}} \leq 4\sqrt{\frac{\log^2(p \vee N \vee T)}{N}}$. Thus,

$$\|\Gamma^{-1}\|_{\ell_\infty} \left\| \frac{1}{NT} Z' u \right\|_{\ell_\infty} \leq 4\|\Gamma^{-1}\|_{\ell_\infty} \sqrt{\frac{\log^2(p \vee N \vee T)}{N}}.$$

Therefore, using that $\mu_{N,T} = \sqrt{\frac{16a_{N,T} \log(p \vee N \vee T)}{N}} = 4\sqrt{\frac{\log^2(p \vee N \vee T)}{N}}$ we conclude that

$$\|\Gamma^{-1}\|_{\ell_\infty} \mu_{N,T} \leq 4\|\Gamma^{-1}\|_{\ell_\infty} \sqrt{\frac{\log^2(p \vee N \vee T)}{N}}$$

In total,

$$\|\hat{\gamma} - \gamma^*\|_{\ell_\infty} \leq 9\|\Gamma^{-1}\|_{\ell_\infty} \sqrt{\frac{\log^2(p \vee N \vee T)}{N}}$$

where. Finally, $\mathcal{A} \cap \mathcal{B}$ has probability tending to one. \square

Lemma 5 (Theorem 1.1 in Gil (2003) adapted to our setting). *Let*

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}$$

and define $v^{up} = \|A_{12}A_{22}^{-1}\|_{\ell_\infty}$ and $v^{low} = \|A_{21}A_{11}^{-1}\|_{\ell_\infty}$. Then, if $v^{low}v^{up} < 1$, A is invertible and

$$\|A^{-1}\|_{\ell_\infty} \leq \frac{(\|A_{11}^{-1}\|_{\ell_\infty} \vee \|A_{22}^{-1}\|_{\ell_\infty})(1 + v^{low})(1 + v^{up})}{1 - v^{low}v^{up}}.$$

⁵Recall that $a_{N,T} = \log(p \vee N \vee T)$ and use that $\log[p + Tp_v + 1] \lesssim \log(p \vee N \vee T)$.

Proof of Lemma 1. Let V be a $T \times T$ matrix whose (i, j) th entry is $\rho_1^{|i-j|}$. By the stationarity assumption on the $x_{i,t}$, one has that

$$\Gamma = \begin{bmatrix} V \otimes \Gamma_0 & v \otimes \Gamma_0 \\ v' \otimes \Gamma_0' & \Gamma_0 \end{bmatrix}$$

where v is a $T \times 1$ vector whose s th element equals $\frac{1}{T} \sum_{t=1}^T \rho_1^{|t-s|}$. Note that in the terminology of Lemma 5

$$v^{up} := \|(v \otimes \Gamma_0) \Gamma_0^{-1}\|_{\ell_\infty} = \|v \otimes \mathbf{I}_p\|_{\ell_\infty} = \|v\|_{\ell_\infty} \leq \frac{2}{T(1-\rho_1)}$$

and

$$\begin{aligned} v^{low} &:= \|(v' \otimes \Gamma_0')(V^{-1} \otimes \Gamma_0^{-1})\|_{\ell_\infty} = \|v' V^{-1} \otimes \mathbf{I}_p\|_{\ell_\infty} = \|v' V^{-1}\|_{\ell_\infty} \\ &= \|V^{-1} v\|_{\ell_\infty} \leq \|V^{-1}\|_{\ell_\infty} \|v\|_{\ell_\infty} \leq \frac{(1+|\rho_1|)^2}{1-\rho_1^2} \frac{2}{T(1-\rho_1)} = \frac{1+|\rho_1|}{1-|\rho_1|} \frac{2}{T(1-\rho_1)} \end{aligned}$$

where we have used that $\|C \otimes D\|_{\ell_\infty} = \|C\|_{\ell_\infty} \|D\|_{\ell_\infty}$ for arbitrary matrices C, D as well as that V^{-1} is a banded matrix with only the diagonal and its two adjacent bands non-zero⁶ such that $\|V^{-1}\|_{\ell_\infty} = \frac{(1+|\rho_1|)^2}{1-\rho_1^2}$. By assumption $v^{low} v^{up} < 1$ such that Lemma 5 yields the boundedness of $\|\Gamma^{-1}\|_{\ell_\infty}$ as $\|\Gamma_0^{-1}\|_{\ell_\infty}$ is assumed to be bounded. \square

Proof of Theorem 3. First, for all $j \in J$, with probability tending to one,

$$|\hat{\gamma}_j| \geq \min_{j \in J} |\gamma_j^*| - \|\hat{\gamma} - \gamma^*\|_{\ell_\infty} \geq 4S_1 - S_1 = 3S_1 > L$$

such that $\tilde{\gamma}_j = \hat{\gamma}_j \neq 0$ for all $j \in J$. A similar argument shows that the sign is actually correct. On the other hand, for all $j \in J^c$,

$$|\hat{\gamma}_j| \leq \|\hat{\gamma} - \gamma^*\|_{\ell_\infty} \leq S_1 < L$$

with probability tending to one. \square

Proof of Theorem 4. When $\hat{J} = J$ one has $\hat{\gamma}_{PostOLS,J} = \hat{\gamma}_{OLS,J}$. Thus, they can only differ when $\hat{J} \neq J$; an event which is asymptotically negligible by Theorem 3. \square

Lemmas and proof related to uniform inference. In the remainder of the appendix we shall repeatedly make use of the following bounds where for brevity we write $\tilde{p} = (p + Tp_v + 1)$.

⁶To be precise, V^{-1} equal $\frac{1}{1-\rho_1^2}$ times a banded matrix whose diagonal equals $(1, 1 + \rho_1^2, \dots, 1 + \rho_1^2, 1)$ while the two bands on either side of the diagonal have all elements equal to $-\rho_1$.

Lemma 6. *Let assumptions A1-A3 be satisfied. Then,*

$$(35) \quad \frac{1}{NT} \|Z(\hat{\gamma} - \gamma^*)\|^2 = O_p\left(\frac{\log(\tilde{p})}{N} s\right)$$

$$(36) \quad \|\hat{\gamma} - \gamma^*\|_1 = O_p\left(\sqrt{\frac{\log(\tilde{p})}{N}} s\right)$$

$$(37) \quad \frac{1}{NT} \|Z_{-j}(\hat{\psi}_j - \psi_j)\|^2 = O_p\left(\frac{\log(\tilde{p})}{N} s_j\right)$$

$$(38) \quad \|\hat{\psi}_j - \psi_j\|_1 = O_p\left(\sqrt{\frac{\log(\tilde{p})}{N}} s_j\right)$$

$$(39) \quad \|\eta'_j Z_{-j}/NT\|_\infty = O_p\left(\sqrt{\frac{\log(\tilde{p})}{N}}\right)$$

where (37)-(39) hold for all $j = 1, \dots, \tilde{p}$.

Proof. (36) follows directly from Theorem 1 with $R = 0$ and $a_{N,T}$ sufficiently large upon using that $\hat{\gamma} - \gamma^* \in \mathcal{C}(J, \gamma^*)$ implies $\|\hat{\gamma} - \gamma^*\|_{\ell_1} \leq 4\sqrt{s} \|\hat{\gamma} - \gamma^*\|$. To see why (35) is valid note that

$$\begin{aligned} \frac{1}{NT} \|Z(\hat{\gamma} - \gamma^*)\|^2 &\leq (\hat{\gamma} - \gamma^*)' \Gamma (\hat{\gamma} - \gamma^*) + |(\hat{\gamma} - \gamma^*)' \frac{Z'Z}{NT} (\hat{\gamma} - \gamma^*) - (\hat{\gamma} - \gamma^*)' \Gamma (\hat{\gamma} - \gamma^*)| \\ &\leq \phi_{\max}(\Gamma) \|\hat{\gamma} - \gamma^*\|^2 + \left\| \frac{1}{NT} Z'Z - \Gamma \right\|_\infty \|\hat{\gamma} - \gamma^*\|_{\ell_1}^2 \\ &= O_p\left(\frac{\log(\tilde{p})}{N} s\right) + O_p\left(\sqrt{\frac{\log(\tilde{p})}{N}}\right) O_p\left(\frac{\log(\tilde{p})}{N} s^2\right) \\ &= O_p\left(\frac{\log(\tilde{p})}{N} s\right) \end{aligned}$$

where we used Lemma 4 and $\sqrt{\frac{\log(\tilde{p})}{N}} s \rightarrow 0$.

Next, note that the arguments leading to Theorem 1 also apply to the nodewise regression $Z_j = Z_{-j}\psi_j + \eta_j$ for $j = 1, \dots, \tilde{p}$. Thus, (37) and (38) follow from the same arguments as above with $\lambda_j \asymp \sqrt{\frac{\log(\tilde{p})}{N}}$. Finally, (39) follows by the same technique as in Lemma 2. \square

Lemma 7. *Let assumptions A1, A2 and A3 be satisfied. Then,*

$$(40) \quad \|\hat{\Theta}_j - \Theta_j\|_1 = O_p\left(s_j \sqrt{\frac{\log(\tilde{p})}{N}}\right).$$

$$(41) \quad \|\hat{\Theta}_j - \Theta_j\|_2 = O_p\left(s_j^{1/2} \sqrt{\frac{\log(\tilde{p})}{N}}\right).$$

$$(42) \quad \|\Theta_j\|_1 = O(s_j^{1/2})$$

$$(43) \quad \|\hat{\Theta}_j\|_1 = O_p(s_j^{1/2})$$

Proof. First, consider $|\hat{\tau}_j^2 - \tau_j^2|$. To this end, we note that the first order conditions for the nodewise regressions can be manipulated to get

$$\hat{\tau}_j^2 = \frac{(Z_j - Z_{-j}\hat{\psi}_j)' Z_j}{NT} = \frac{\eta'_j \eta_j}{NT} + \frac{\eta'_j Z_{-j} \psi_j}{NT} - \frac{(\hat{\psi}_j - \psi_j)' Z_{-j}' Z_{-j} \psi_j}{NT} - \frac{(\hat{\psi}_j - \psi_j)' Z_{-j}' \eta_j}{NT}.$$

where the second equality used $Z_j = Z_{-j}\psi_j + \eta_j$. Using the above expression one gets for all $j = 1, \dots, \tilde{p}$

(44)

$$|\hat{\tau}_j^2 - \tau_j^2| \leq \left| \frac{\eta'_j \eta_j}{NT} - \tau_j^2 \right| + |\eta'_j Z_{-j}(\hat{\psi}_j - \psi_j)/NT| + |\eta'_j Z_{-j}\psi_j/NT| + \left| \frac{\psi'_j Z'_{-j} Z_{-j}(\hat{\psi}_j - \psi_j)}{NT} \right|.$$

Since $\frac{1}{T} \sum_{t=1}^T (\eta_{j,i,t}^2 - E(\eta_{i,t,j}^2))$ is mean zero and subexponential for all $i = 1, \dots, N$ it follows from the independence across i that $\frac{\eta'_j \eta_j}{NT} - \tau_j^2 = \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T (\eta_{i,t,j}^2 - E(\eta_{i,t,j}^2)) = O_p(N^{-1/2})$ (alternatively, the order of magnitude follows by the classical CLT). Next, consider the second term in (44). By (38) and (39) it follows that

$$\begin{aligned} |\eta'_j Z_{-j}(\hat{\psi}_j - \psi_j)/NT| &\leq \|\eta'_j Z_{-j}/NT\|_\infty \|\hat{\psi}_j - \psi_j\|_1 = O_p\left(\sqrt{\frac{\log(\tilde{p})}{N}}\right) O_p\left(s_j \sqrt{\frac{\log(\tilde{p})}{N}}\right) \\ &= O_p\left(s_j \frac{\log(\tilde{p})}{N}\right) \\ (45) \qquad \qquad \qquad &= O_p\left(\sqrt{s_j \frac{\log(\tilde{p})}{N}}\right) \end{aligned}$$

using $\sqrt{s_j \frac{\log(\tilde{p})}{N}} \rightarrow 0$. Furthermore, using the variational characterization of eigenvalues and $\phi_{\min}(\Gamma)$ bounded away from 0 we can arrive at $\|\psi_j\|_2$ uniformly bounded and $\|\psi_j\|_1 = O(\sqrt{s_j})$. Proceeding to the third term of (44),

$$(46) \qquad |\eta'_j Z_{-j}\psi_j/NT| \leq \|\eta'_j Z_{-j}/NT\|_\infty \|\psi_j\|_1 = O_p\left(\sqrt{s_j} \sqrt{\frac{\log(\tilde{p})}{N}}\right),$$

where we have also used (39). It remains to bound the fourth summand in (44). By the Karush-Kuhn-Tucker conditions for the nodewise regression one has

$$\lambda_j \hat{\kappa}_j + \frac{Z'_{-j} Z_{-j} \hat{\psi}_j}{NT} - \frac{Z'_{-j} Z_j}{NT} = 0,$$

which upon using $Z_j = Z_{-j}\psi_j + \eta_j$ yields

$$\left\| \frac{Z'_{-j} Z_{-j}}{NT} (\hat{\psi}_j - \psi_j) \right\|_\infty \leq \left\| \frac{Z'_{-j} \eta_j}{NT} \right\|_\infty + \|\lambda_j \hat{\kappa}_j\|_\infty = O_p\left(\sqrt{\frac{\log(\tilde{p})}{N}}\right).$$

where we have used $\|\hat{\kappa}_j\|_\infty \leq 1$ as well as (39) and $\lambda_j \asymp \sqrt{\frac{\log(\tilde{p})}{N}}$. This means, using $\|\psi_j\|_1 = O(s_j^{1/2})$,

$$(47) \qquad \left| \psi'_j \frac{Z'_{-j} Z_{-j}}{NT} (\hat{\psi}_j - \psi_j) \right| = O_p\left(s_j^{1/2} \sqrt{\frac{\log(\tilde{p})}{N}}\right).$$

Thus,

$$|\hat{\tau}_j^2 - \tau_j^2| = O_p\left(s_j^{1/2} \sqrt{\frac{\log(\tilde{p})}{N}}\right).$$

Next, note that $\tau_j^2 = 1/\Theta_{j,j} \geq 1/\phi_{\max}(\Theta) = \phi_{\min}(\Gamma)$ for all $j = 1, \dots, p$ with $\phi_{\min}(\Gamma)$ bounded away from zero by assumption. Thus, τ_j^2 is bounded away from zero, and so

$$\hat{\tau}_j^2 = [\hat{\tau}_j^2 - \tau_j^2 + \tau_j^2] \geq \tau_j^2 - |\hat{\tau}_j^2 - \tau_j^2|$$

is bounded away from zero with probability tending to one using $|\hat{\tau}_j^2 - \tau_j^2| = O_p\left(s_j^{1/2} \sqrt{\frac{\log(\tilde{p})}{N}}\right) = o_p(1)$. This implies

$$(48) \quad \left| \frac{1}{\hat{\tau}_j^2} - \frac{1}{\tau_j^2} \right| = \frac{|\tau_j^2 - \hat{\tau}_j^2|}{\hat{\tau}_j^2 \tau_j^2} = O_p\left(s_j^{1/2} \sqrt{\frac{\log(\tilde{p})}{N}}\right).$$

We are now ready to bound $\|\hat{\Theta}_j - \Theta_j\|_1$. Recall that $\hat{\Theta}_j$ is formed by dividing \hat{C}_j by $\hat{\tau}_j^2$. Let Θ_j denote the j 'th row of Θ written as a column vector. Θ_j is formed by dividing C_j (j 'th row of C written as a column vector) by τ_j^2 . Therefore, using $\|\psi_j\|_1 = O(s_j^{1/2})$, (38), and (48)

$$\begin{aligned} \|\hat{\Theta}_j - \Theta_j\|_1 &= \left\| \frac{\hat{C}_j}{\hat{\tau}_j^2} - \frac{C_j}{\tau_j^2} \right\|_1 \leq \left| \frac{1}{\hat{\tau}_j^2} - \frac{1}{\tau_j^2} \right| + \left\| \frac{\hat{\psi}_j}{\hat{\tau}_j^2} - \frac{\psi_j}{\tau_j^2} \right\|_1 = \left| \frac{1}{\hat{\tau}_j^2} - \frac{1}{\tau_j^2} \right| + \left\| \frac{\hat{\psi}_j}{\hat{\tau}_j^2} - \frac{\psi_j}{\hat{\tau}_j^2} + \frac{\psi_j}{\hat{\tau}_j^2} - \frac{\psi_j}{\tau_j^2} \right\|_1 \\ &\leq \left| \frac{1}{\hat{\tau}_j^2} - \frac{1}{\tau_j^2} \right| + \frac{\|\hat{\psi}_j - \psi_j\|_1}{\hat{\tau}_j^2} + \|\psi_j\|_1 \left(\left| \frac{1}{\hat{\tau}_j^2} - \frac{1}{\tau_j^2} \right| \right) \\ &= O_p\left(s_j^{1/2} \sqrt{\frac{\log(\tilde{p})}{N}}\right) + O_p\left(s_j \sqrt{\frac{\log(\tilde{p})}{N}}\right) + O_p\left(s_j \sqrt{\frac{\log(\tilde{p})}{N}}\right) \\ (49) \quad &= O_p\left(s_j \sqrt{\frac{\log(\tilde{p})}{N}}\right). \end{aligned}$$

Next, for later purposes, we also bound $\|\hat{\Theta}_j - \Theta_j\|_2$. Using $\|\psi_j\|_2^2$ uniformly bounded

$$\begin{aligned} \|\hat{\Theta}_j - \Theta_j\|_2 &\leq \left| \frac{1}{\hat{\tau}_j^2} - \frac{1}{\tau_j^2} \right| + \frac{\|\hat{\psi}_j - \psi_j\|_2}{\hat{\tau}_j^2} + \|\psi_j\|_2 \left(\left| \frac{1}{\hat{\tau}_j^2} - \frac{1}{\tau_j^2} \right| \right) \\ &= O_p\left(s_j^{1/2} \sqrt{\frac{\log(\tilde{p})}{N}}\right) + O_p\left(s_j^{1/2} \sqrt{\frac{\log(\tilde{p})}{N}}\right) + O_p\left(s_j^{1/2} \sqrt{\frac{\log(\tilde{p})}{N}}\right), \\ (50) \quad &= O_p\left(s_j^{1/2} \sqrt{\frac{\log(\tilde{p})}{N}}\right). \end{aligned}$$

Finally, we show that $\|\hat{\Theta}_j\|_1 = O_p(\sqrt{s_j})$. To this end, recall $\|\psi_j\|_1 = O(\sqrt{s_j})$ such that

$$(51) \quad \|\Theta_j\|_1 \leq \frac{1}{\tau_j^2} + \|\psi_j/\tau_j^2\|_1 = O(s_j^{1/2})$$

(as τ_j^2 is uniformly bounded away from zero). Then, using $s_j^{1/2} \sqrt{\frac{\log(\tilde{p})}{N}} \rightarrow 0$,

$$(52) \quad \|\hat{\Theta}_j\|_1 \leq \|\hat{\Theta}_j - \Theta_j\|_1 + \|\Theta_j\|_1 = O_p\left(s_j \sqrt{\frac{\log(\tilde{p})}{N}}\right) + O(\sqrt{s_j}) = O_p(\sqrt{s_j}).$$

□

Proof of Theorem 5. We show that the ratio

$$(53) \quad t = \frac{N^{1/2} \rho'(\hat{b} - \gamma^*)}{\sqrt{\rho' \hat{\Theta} \Gamma_{zu} \hat{\Theta}' \rho}},$$

is asymptotically standard normal. First, note that by (12) one can write

$$t = t_1 + t_2,$$

where

$$t_1 = \frac{\rho' \hat{\Theta} Z' u / (N^{1/2} T)}{\sqrt{\rho' \hat{\Theta} \Gamma_{zu} \hat{\Theta}' \rho}} \text{ and } t_2 = -\frac{\rho' \Delta}{\sqrt{\rho' \hat{\Theta} \Gamma_{zu} \hat{\Theta}' \rho}}.$$

It suffices to show that t_1 is asymptotically standard normal and $t_2 = o_p(1)$.

Step 1. We first show that t_1 is asymptotically standard normal.

a) To show that t_1 is asymptotically standard normal we first show that

$$t'_1 = \frac{\rho' \Theta Z' u / (N^{1/2} T)}{\sqrt{\rho' \Theta \Gamma_{zu} \Theta' \rho}} = \frac{\rho' \Theta \sum_{i=1}^N \sum_{t=1}^T z_{i,t} u_{i,t} / (N^{1/2} T)}{\sqrt{\rho' \Theta \Gamma_{zu} \Theta' \rho}}$$

converges in distribution to a standard normal where $\Gamma_{zu} = \frac{1}{T^2} E \left(\sum_{t=1}^T z_{1,t} u_{1,t} \right) \left(\sum_{t=1}^T z_{1,t} u_{1,t} \right)' = \frac{1}{T^2} E \left(\sum_{t=1}^T z_{1,t} (\epsilon_{1,t} + a_1) \right) \left(\sum_{t=1}^T z_{1,t} (\epsilon_{1,t} + a_1) \right)'$. Then we show that t'_1 and t_1 are asymptotically equivalent. Note that

$$(54) \quad E \left[\frac{\rho' \Theta Z' u / (N^{1/2} T)}{\sqrt{\rho' \Theta \Gamma_{zu} \Theta' \rho}} \right] = E \left[\frac{\rho' \Theta \sum_{i=1}^N \sum_{t=1}^T z_{i,t} u_{i,t} / (N^{1/2} T)}{\sqrt{\rho' \Theta \Gamma_{zu} \Theta' \rho}} \right] = 0,$$

and

$$E \left[\frac{\rho' \Theta Z' u / (N^{1/2} T)}{\sqrt{\rho' \Theta \Gamma_{zu} \Theta' \rho}} \right]^2 = E \left[\frac{\rho' \Theta \sum_{i=1}^N \sum_{t=1}^T z_{i,t} u_{i,t} / (N^{1/2} T)}{\sqrt{\rho' \Theta \Gamma_{zu} \Theta' \rho}} \right]^2 = 1.$$

Therefore, using the independence and identical distributedness across $i = 1, \dots, N$ the classical central limit theorem yields t'_1 converges in distribution to a standard normal. Next, we remark that $\rho' \Theta \Gamma_{zu} \Theta' \rho$ is asymptotically bounded away from zero. Clearly,

$$(55) \quad \rho' \Theta \Gamma_{zu} \Theta' \rho \geq \phi_{\min}(\Gamma_{zu}) \|\Theta' \rho\|_2^2 \geq \phi_{\min}(\Gamma_{zu}) \phi_{\min}^2(\Theta) \|\rho\|_2^2 \geq \phi_{\min}(\Gamma_{zu}) \frac{1}{\phi_{\max}^2(\Gamma)},$$

which is bounded away from zero since $\phi_{\min}(\Gamma_{zu})$ is bounded away from zero and $\phi_{\max}(\Gamma)$ is bounded from above.

b) We now show that $t'_1 - t_1 = o_p(1)$. To do so it suffices that the numerators as well as the denominators of t'_1 and t_1 are asymptotically equivalent since $\rho' \Theta \Gamma_{zu} \Theta' \rho$ is bounded away from zero by (55). We first show that the denominators of t'_1 and t_1 are asymptotically equivalent, i.e.

$$(56) \quad |\rho' \hat{\Theta} \Gamma_{zu} \hat{\Theta}' \rho - \rho' \Theta \Gamma_{zu} \Theta' \rho| = o_p(1).$$

Set $\tilde{\Gamma}_{zu} = \frac{1}{N} \sum_{i=1}^N \left[\frac{1}{T^2} \left(\sum_{t=1}^T z_{i,t} u_{i,t} \right) \left(\sum_{t=1}^T z_{i,t} u_{i,t} \right)' \right]$. To establish (56) it suffices to show the following relations:

$$(57) \quad |\rho' \hat{\Theta} \Gamma_{zu} \hat{\Theta}' \rho - \rho' \hat{\Theta} \tilde{\Gamma}_{zu} \hat{\Theta}' \rho| = o_p(1).$$

$$(58) \quad |\rho' \hat{\Theta} \tilde{\Gamma}_{zu} \hat{\Theta}' \rho - \rho' \hat{\Theta} \Gamma_{zu} \hat{\Theta}' \rho| = o_p(1).$$

$$(59) \quad |\rho' \hat{\Theta} \Gamma_{zu} \hat{\Theta}' \rho - \rho' \Theta \Gamma_{zu} \Theta' \rho| = o_p(1).$$

We first prove (57). To this end, note that

$$(60) \quad |\rho' \hat{\Theta} \hat{\Gamma}_{zu} \hat{\Theta}' \rho - \rho' \hat{\Theta} \tilde{\Gamma}_{zu} \hat{\Theta}' \rho| \leq \|\hat{\Gamma}_{zu} - \tilde{\Gamma}_{zu}\|_{\infty} \|\hat{\Theta}' \rho\|_1^2.$$

But by (43) and $\|\rho\|_2 = 1$

$$(61) \quad \|\hat{\Theta}' \rho\|_1 = \left\| \sum_{j \in H} \hat{\Theta}_j \rho_j \right\|_1 \leq \sum_{j \in H} |\rho_j| \|\hat{\Theta}_j\|_1 = O_p(\sqrt{s_j}).$$

To proceed, we bound $\|\hat{\Gamma}_{zu} - \tilde{\Gamma}_{zu}\|_{\infty}$. Using $\hat{u}_{i,t} = u_{i,t} - z'_{i,t}(\hat{\gamma} - \gamma^*)$ in the definition of $\hat{\Gamma}_{zu}$ we get

$$(62) \quad \begin{aligned} \hat{\Gamma}_{zu} - \tilde{\Gamma}_{zu} &= -\frac{1}{N} \sum_{i=1}^N \frac{1}{T^2} \left(\sum_{t=1}^T z_{i,t} u_{i,t} \right) \left(\sum_{t=1}^T z_{i,t} z'_{i,t} (\hat{\gamma} - \gamma^*) \right)' - \frac{1}{N} \sum_{i=1}^N \frac{1}{T^2} \left(\sum_{t=1}^T z_{i,t} z'_{i,t} (\hat{\gamma} - \gamma^*) \right) \left(\sum_{t=1}^T z_{i,t} u_{i,t} \right)' \\ &\quad + \frac{1}{N} \sum_{i=1}^N \frac{1}{T^2} \left(\sum_{t=1}^T z_{i,t} z'_{i,t} (\hat{\gamma} - \gamma^*) \right) \left(\sum_{t=1}^T z_{i,t} z'_{i,t} (\hat{\gamma} - \gamma^*) \right)'. \end{aligned}$$

We bound each sum separately. By Cauchy-Schwarz's and Jensen's inequality

$$\begin{aligned} &\max_{1 \leq k, l \leq \tilde{p}} \left| \frac{1}{N} \sum_{i=1}^N \frac{1}{T^2} \left(\sum_{t=1}^T z_{i,t,k} u_{i,t} \right) \left(\sum_{t=1}^T z_{i,t,l} z'_{i,t} (\hat{\gamma} - \gamma^*) \right) \right| \\ &\leq \max_{1 \leq k, l \leq \tilde{p}} \frac{1}{N} \frac{1}{T^2} \left[\sum_{i=1}^N \left(\sum_{t=1}^T z_{i,t,k} u_{i,t} \right)^2 \right]^{1/2} \left[\sum_{i=1}^N \left(\sum_{t=1}^T z_{i,t,l} z'_{i,t} (\hat{\gamma} - \gamma^*) \right)^2 \right]^{1/2} \\ &\leq \max_{1 \leq k, l \leq \tilde{p}} \frac{1}{N} \frac{1}{T} \left[\sum_{i=1}^N \sum_{t=1}^T z_{i,t,k}^2 u_{i,t}^2 \right]^{1/2} \left[\sum_{i=1}^N \sum_{t=1}^T z_{i,t,l}^2 [z'_{i,t} (\hat{\gamma} - \gamma^*)]^2 \right]^{1/2} \\ &\leq \max_{1 \leq k \leq \tilde{p}} \left[\frac{1}{N} \frac{1}{T} \sum_{i=1}^N \sum_{t=1}^T z_{i,t,k}^2 u_{i,t}^2 \right]^{1/2} \left[\max_{i,t,l} z_{i,t,l}^2 \frac{1}{NT} \|Z(\hat{\gamma} - \gamma^*)\|^2 \right]^{1/2} \end{aligned}$$

where the final maximum is over the obvious indices of i, t, l . Tedious calculations using the subgaussianity of the $z_{i,t,k}$ and $u_{i,t}$ yield that $\max_{1 \leq k \leq \tilde{p}} \left[\frac{1}{N} \frac{1}{T} \sum_{i=1}^N \sum_{t=1}^T z_{i,t,k}^2 u_{i,t}^2 \right]^{1/2} = O_p(1)$ if $\log(\tilde{p})^5/N \rightarrow 0$ ⁷. Furthermore, $\max_{i,t,l} z_{i,t,l}^2 = O_p(\log(\tilde{p} \vee N))$. Thus, combining this with (35) we get that the first term of (62)

$$(63) \quad \left\| \frac{1}{N} \sum_{i=1}^N \frac{1}{T^2} \left(\sum_{t=1}^T z_{i,t} u_{i,t} \right) \left(\sum_{t=1}^T z_{i,t} z'_{i,t} (\hat{\gamma} - \gamma^*) \right)' \right\|_{\infty} = O_p\left(\frac{\log(\tilde{p} \vee N)}{N^{1/2}} s^{1/2}\right).$$

As the second term in (62) has identical entries to the first (it is its transpose) they have identical $\|\cdot\|_{\infty}$ -norms. Regarding the third term in (62) note that using Cauchy-Schwarz

⁷See Proposition 2, Appendix B of Kock and Tang (2015).

and Jensen's inequality as above

$$\begin{aligned} \max_{1 \leq k, l \leq \tilde{p}} \left| \frac{1}{N} \sum_{i=1}^N \frac{1}{T^2} \left(\sum_{t=1}^T z_{i,t,k} z'_{i,t} (\hat{\gamma} - \gamma^*) \right) \left(\sum_{t=1}^T z_{i,t,l} z'_{i,t} (\hat{\gamma} - \gamma^*) \right) \right| &\leq \max_{i,t,k} z_{i,t,k}^2 \frac{1}{NT} \|Z(\hat{\gamma} - \gamma^*)\|^2 \\ (64) \qquad \qquad \qquad &= O_p \left(\frac{\log(\tilde{p} \vee N)^2}{N} s \right) \end{aligned}$$

Then, combining (63) and (64) implies that

$$\|\hat{\Gamma}_{zu} - \tilde{\Gamma}_{zu}\|_{\infty} = O_p \left(\frac{\log(\tilde{p} \vee N)}{N^{1/2}} s^{1/2} \right) + O_p \left(\frac{\log(\tilde{p} \vee N)^2}{N} s \right).$$

Combining with (61) yields

$$|\rho' \hat{\Theta} \hat{\Gamma}_{zu} \hat{\Theta}' \rho - \rho' \hat{\Theta} \tilde{\Gamma}_{zu} \hat{\Theta}' \rho| = O_p \left(\frac{\log(\tilde{p} \vee N)}{N^{1/2}} s^{1/2} s_j \right) + O_p \left(\frac{\log(\tilde{p} \vee N)^2}{N} s s_j \right) = o_p(1),$$

using $\frac{\log(\tilde{p} \vee N)}{N^{1/2}} s^{1/2} s_j \rightarrow 0$. This establishes (57). Next, we turn to (58). First, note that

$$(65) \qquad |\rho' \hat{\Theta} \tilde{\Gamma}_{zu} \hat{\Theta}' \rho - \rho' \hat{\Theta} \Gamma_{zu} \hat{\Theta}' \rho| \leq \|\tilde{\Gamma}_{zu} - \Gamma_{zu}\|_{\infty} \|\hat{\Theta}' \rho\|_1^2.$$

Using the subgaussianity of the $z_{i,t,k}$ and $u_{i,t}$ it can be shown that

$$\|\tilde{\Gamma}_{zu} - \Gamma_{zu}\|_{\infty} = O_p \left(\sqrt{\frac{\log(\tilde{p} \vee N)^5}{N}} \right).$$

By (65) and (61)

$$|\rho' \hat{\Theta} \tilde{\Gamma}_{zu} \hat{\Theta}' \rho - \rho' \hat{\Theta} \Gamma_{zu} \hat{\Theta}' \rho| = O_p \left(\sqrt{\frac{\log(\tilde{p} \vee N)^5}{N}} s_j \right) = o_p(1),$$

which establishes (58). Finally, we establish (59) to conclude (56). By Lemma 6.1 in van de Geer et al. (2014)

$$\begin{aligned} |\rho' \hat{\Theta} \Gamma_{zu} \hat{\Theta}' \rho - \rho' \Theta \Gamma_{zu} \Theta' \rho| &\leq \|\Gamma_{zu}\|_{\infty} \|\hat{\Theta}' \rho - \Theta' \rho\|_1^2 + 2 \|\Gamma_{zu} \Theta' \rho\|_2 \|\hat{\Theta}' \rho - \Theta' \rho\|_2 \\ &\leq \|\Gamma_{zu}\|_{\infty} \|(\hat{\Theta}' - \Theta') \rho\|_1^2 + 2 \phi_{\max}(\Gamma_{zu}) \|\Theta' \rho\|_2 \|(\hat{\Theta}' - \Theta') \rho\|_2. \end{aligned}$$

Note that

$$(66) \quad \|(\hat{\Theta}' - \Theta') \rho\|_1 = \left\| \sum_{j \in H} (\hat{\Theta}_j - \Theta_j) \rho_j \right\|_1 \leq \sum_{j \in H} \|\hat{\Theta}_j - \Theta_j\|_1 |\rho_j| = O_p \left(\sqrt{\frac{\log(\tilde{p})}{N}} s_j \right)$$

by (40) and $\|\rho\|_2 = 1$. Furthermore, using the symmetry of Θ ,

$$\|\Theta' \rho\|_2 \leq \phi_{\max}(\Theta) \|\rho\|_2 = \frac{1}{\phi_{\min}(\Gamma)},$$

which is bounded by assumption. Finally,

$$\|(\hat{\Theta}' - \Theta') \rho\|_2 = \left\| \sum_{j \in H} (\hat{\Theta}_j - \Theta_j) \rho_j \right\|_2 \leq \sum_{j \in H} \|\hat{\Theta}_j - \Theta_j\|_2 |\rho_j| = O_p \left(\sqrt{\frac{\log(\tilde{p})}{N}} s_j^{1/2} \right),$$

by (41) and $\|\rho\|_2 = 1$. Therefore, by $\|\Gamma_{zu}\|_{\infty} \leq \phi_{\max}(\Gamma_{zu})$ with the latter assumed bounded,

$$|\rho' \hat{\Theta} \Gamma_{zu} \hat{\Theta}' \rho - \rho' \Theta \Gamma_{zu} \Theta' \rho| = O_p \left(\frac{\log(\tilde{p})}{N} s_j^2 \right) + O_p \left(\sqrt{\frac{\log(\tilde{p})}{N}} s_j^{1/2} \right) = o_p(1).$$

The uniformity of (56) over $\mathcal{B}_{\ell_0}(s)$ follows from simply observing that (63) and (64) above are actually valid uniformly over this set and that this is the only place in which γ^* enters in the above arguments thus establishing (23).

We now turn to showing that the numerators of t'_1 and t_1 are asymptotically equivalent, i.e.

$$|\rho' \hat{\Theta} Z' u / (N^{1/2} T) - \rho' \Theta Z' u / (N^{1/2} T)| = o_p(1).$$

By Lemma 2 and (66)

$$\begin{aligned} N^{1/2} |\rho' \hat{\Theta} Z' u / (NT) - \rho' \Theta Z' u / (NT)| &\leq N^{1/2} \left\| \frac{Z' u}{NT} \right\|_{\infty} \|\rho'(\hat{\Theta} - \Theta)\|_1 \\ &= N^{1/2} O_p \left(\sqrt{\frac{\log(\tilde{p})}{N}} \right) O_p \left(\sqrt{\frac{\log(\tilde{p})}{N}} s_j \right) \\ &= O_p \left(\frac{\log(\tilde{p})}{N^{1/2}} s_j \right) \\ (67) \quad &= o_p(1). \end{aligned}$$

Step 2. It remains to be shown that $t_2 = o_p(1)$. The denominators of t_1 and t_2 are identical. Hence, the denominator of t_2 is asymptotically bounded away from zero with probability approaching one by (55) and (56). Thus, it suffices to show that the numerator of t_2 vanishes in probability. Note that, by the definition of Δ and $\|\rho\|_2 = 1$,

$$\begin{aligned} |\rho' \Delta| &\leq \max_{j \in H} |\Delta_j| \sum_{j \in H} |\rho_j| \leq \max_{j \in H} |(\hat{\Theta}'_j \hat{\Gamma} - e_j) (\sqrt{N}(\hat{\gamma} - \gamma^*))| \sum_{j \in H} |\rho_j| \\ (68) \quad &\leq \max_{j \in H} \left\| (\hat{\Theta}'_j \hat{\Gamma} - e_j) \right\|_{\infty} \left\| \sqrt{N}(\hat{\gamma} - \gamma^*) \right\|_1. \end{aligned}$$

First, it follows from (36) that $N^{1/2} \|\hat{\gamma} - \gamma^*\|_1 = O_p(\sqrt{\log(\tilde{p})} s)$. Next, we consider

$$\max_{j \in H} \left\| (\hat{\Theta}'_j \hat{\Gamma} - e_j) \right\|_{\infty} \leq \frac{\lambda_j}{\hat{\tau}_j^2} = O_p \left(\sqrt{\frac{\log(\tilde{p})}{N}} \right),$$

where we have used the definition of λ_j and $1/\hat{\tau}_j^2 = O_p(1)$ by (48). In total we have

$$|\rho' \Delta| = O_p \left(\sqrt{\frac{\log(\tilde{p})}{N}} \right) O_p(\sqrt{\log(\tilde{p})} s) = O_p \left(\frac{\log(\tilde{p})}{N^{1/2}} s \right) = o_p(1).$$

The fact that $\sup_{\gamma^* \in \mathcal{B}_{\ell_0}(s)} |\rho' \Delta| = o_p(1)$ follows from the observation that (36) actually yields that $\sup_{\gamma^* \in \mathcal{B}_{\ell_0}(s)} N^{1/2} \|\hat{\gamma} - \gamma^*\|_1 = O_p(\sqrt{\log(\tilde{p})} s)$ in the above argument and that this is the only place in which γ^* enters these arguments. Thus, for later reference,

$$(69) \quad \sup_{\gamma^* \in \mathcal{B}_{\ell_0}(s)} |\rho' \Delta| = o_p(1).$$

The uniformity in in (22) and (24)-(25) is established using arguments similar to the ones in Caner and Kock (2014). Thus full details are provided in the version of the present paper on the author's homepage at <https://sites.google.com/site/andersbkock/>. \square

REFERENCES

- Arellano, M. (2003). *Panel Data Econometrics*, Volume 1. Oxford University Press, Oxford.
- Baltagi, B. (2008). *Econometric analysis of panel data*. John Wiley & Sons.

- Belloni, A. and V. Chernozhukov (2011). High dimensional sparse econometric models: An introduction. *Inverse Problems and High-Dimensional Estimation*, 121–156.
- Belloni, A., V. Chernozhukov, C. Hansen, and D. Kozbur (2015). Inference in high dimensional panel models with an application to gun control. *Journal of Business & Economic Statistics* (just-accepted), 1–33.
- Bickel, P., Y. Ritov, and A. Tsybakov (2009). Simultaneous analysis of lasso and dantzig selector. *The Annals of Statistics* 37(4), 1705–1732.
- Bühlmann, P. and S. van de Geer (2011). *Statistics for High-Dimensional Data: Methods, Theory and Applications*. Springer-Verlag, New York.
- Candes, E. and T. Tao (2007). The dantzig selector: statistical estimation when p is much larger than n . *The Annals of Statistics*, 2313–2351.
- Caner, M. and X. Han (2014). Selecting the correct number of factors in approximate factor models: The large panel case with group bridge estimators. *Journal of Business & Economic Statistics* 32(3), 359–374.
- Caner, M. and A. B. Kock (2014). Asymptotically honest confidence regions for high dimensional parameters by the desparsified conservative lasso. *arXiv preprint arXiv:1410.4208*.
- Chamberlain, G. (1982). Multivariate regression models for panel data. *Journal of Econometrics* 18(1), 5–46.
- Chamberlain, G. (1984). *Panel Data (in Handbook of Econometrics)*. edited by Griliches, Z and Intriligator, M.D., Elsevier, Amsterdam: North Holland, 1247–1318.
- Christiansen, C., J. S. Joensen, and J. Rangvid (2008). Are economists more likely to hold stocks? *Review of Finance* 12(3), 465–496.
- Fan, J. and R. Li (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American statistical Association* 96(456), 1348–1360.
- Fan, J. and J. Lv (2008). Sure independence screening for ultrahigh dimensional feature space. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 70(5), 849–911.
- Fan, J., J. Lv, and L. Qi (2011). Sparse high dimensional models in economics. *Annual Review of Economics* 3, 291.
- Fan, J., L. Xue, and H. Zou (2014). Strong oracle optimality of folded concave penalized estimation. *Annals of Statistics* 42(3), 819–849.
- Fan, Y. and C. Y. Tang (2013). Tuning parameter selection in high dimensional penalized likelihood. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 75(3), 531–552.
- Galvao, A. F. and G. V. Montes-Rojas (2010). Penalized quantile regression for dynamic panel data. *Journal of Statistical Planning and Inference* 140(11), 3476–3497.
- Gil, M. (2003). Invertibility conditions for block matrices and estimates for norms of inverse matrices. *The Rocky Mountain Journal of Mathematics* 33(4), 1323–1335.
- Hsiao, C. (2014). *Analysis of panel data*. Number 54. Cambridge University Press.
- Kock, A. B. (2013a). Oracle efficient variable selection in random and fixed effects panel data models. *Econometric Theory* 29(01), 115–152.
- Kock, A. B. (2013b). Oracle inequalities for high-dimensional panel data models. *arXiv preprint arXiv:1310.8207*.
- Kock, A. B. and H. Tang (2015). Uniform inference in high-dimensional dynamic panel data models. *arXiv preprint arXiv:1501.00478*.
- Koenker, R. (2004). Quantile regression for longitudinal data. *Journal of Multivariate Analysis* 91(1), 74–89.

- Lamarche, C. (2010). Robust penalized quantile regression estimation for panel data. *Journal of Econometrics* 157(2), 396–408.
- Leeb, H. and B. M. Pötscher (2005). Model selection and inference: Facts and fiction. *Econometric Theory* 21(01), 21–59.
- Li, D., J. Qian, and L. Su (2015). Panel data models with interactive fixed effects and multiple structural breaks. *Journal of the American Statistical Association* (just-accepted), 1–42.
- Manresa, E. (2013). Estimating the structure of social interactions using panel data. *Unpublished Manuscript. CEMFI, Madrid*.
- Meinshausen, N. and P. Bühlmann (2006). High-dimensional graphs and variable selection with the lasso. *The Annals of Statistics* 34, 1436–1462.
- Mundlak, Y. (1978). On the pooling of time series and cross section data. *Econometrica: Journal of the Econometric Society*, 69–85.
- Negahban, S., P. Ravikumar, M. Wainwright, and B. Yu (2012). A unified framework for high-dimensional analysis of m-estimators with decomposable regularizers. *Statistical Science* 27(4), 538–557.
- Papke, L. E. and J. M. Wooldridge (2008). Panel data methods for fractional response variables with an application to test pass rates. *Journal of Econometrics* 145(1), 121–133.
- Qian, J. and L. Su (2016). Shrinkage estimation of common breaks in panel data models via adaptive group fused lasso. *Journal of Econometrics* 191(1), 86–109.
- Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, 267–288.
- van de Geer, S., P. Bühlmann, Y. Ritov, R. Dezeure, et al. (2014). On asymptotically optimal confidence regions and tests for high-dimensional models. *The Annals of Statistics* 42(3), 1166–1202.
- Vershynin, R. (2012). *Introduction to the non-asymptotic analysis of random matrices*. In "Compressed Sensing: Theory and Applications". Edited by Yonina C. Eldar and Gitta Kutyniok. Cambridge University Press.
- Wooldridge, J. (2002). *Econometric analysis of cross section and panel data*. The MIT Press.
- Wooldridge, J. M. (2010). *Econometric analysis of cross section and panel data*. MIT Press.
- Xu, Z., Z. Guan, T. S. Jayne, and R. Black (2009). Factors influencing the profitability of fertilizer use on maize in zambia. *Agricultural Economics* 40(4), 437–446.
- Yuan, M. (2010). High dimensional inverse covariance matrix estimation via linear programming. *The Journal of Machine Learning Research* 11, 2261–2286.
- Zhang, C.-H. and S. S. Zhang (2014). Confidence intervals for low dimensional parameters in high dimensional linear models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 76(1), 217–242.
- Zhao, P. and B. Yu (2006). On model selection consistency of lasso. *The Journal of Machine Learning Research* 7, 2541–2563.