AARHUS UNIVERSITY
BUSINESS AND SOCIAL SCIENCES
DEPARTMENT OF ECONOMICS AND BUSINESS

CREATES
Center for Research in Econometric Analysis of Time Series

Inference in High-dimensional Dynamic Panel Data Models

Anders Bredahl Kock and Haihan Tang

CREATES Research Paper 2014-58

Department of Economics and Business
Aarhus University
Fuglesangs Allé 4
DK-8210 Aarhus V
Denmark

Email: oekonomi@au.dk
Tel: +45 8716 5515

Inference in High-dimensional Dynamic Panel Data Models

Anders Bredahl Kock* Haihan Tang†

December 30, 2014

Abstract

We establish oracle inequalities for a version of the Lasso in high-dimensional fixed effects dynamic panel data models. The inequalities are valid for the coefficients of the dynamic and exogenous regressors. Separate oracle inequalities are derived for the fixed effects. Next, we show how one can conduct simultaneous inference on the parameters of the model and construct a uniformly valid estimator of the asymptotic covariance matrix which is robust to conditional heteroskedasticity in the error terms. Allowing for conditional heteroskedasticity is important in dynamic models as the conditional error variance may be non-constant over time and depend on the covariates. Furthermore, our procedure allows for inference on high-dimensional subsets of the parameter vector of an increasing cardinality. We show that the confidence bands resulting from our procedure are asymptotically honest and contract at the optimal rate. This rate is different for the fixed effects than for the remaining parts of the parameter vector.

Keywords: Panel data Dynamic models, Lasso, Desparsification, High-dimensional data, Uniform inference, Honest inference, Oracle inequality, Confidence intervals, Tests.
JEL: C13, C23, C55.

1 Introduction

Dynamic panel data models are widely used in economics and the social sciences in particular. They are extremely popular as workers, firms, and countries often differ due to unobserved factors. Furthermore, these units are often sampled repeatedly over time in many modern applications thus allowing to model the dynamic development of these. However, so far no work has been done on how to conduct inference in the high-dimensional dynamic fixed effects model

$$y_{i,t} = \sum_{l=1}^{L_N} \alpha_l y_{i,t-l} + x_{i,t}'\beta + \eta_i + \varepsilon_{i,t}, \; i = 1, ..., N, \quad \text{and} \quad t = 1, ..., T \tag{1.1}$$

where the presence of L_N lags of $y_{i,t}$ allows for autoregressive dependence of $y_{i,t}$ on its own past. $x_{i,t}$ is a $p_{x,N} \times 1$ vector of exogenous variables and $\eta_i, i = 1, ..., N$ are the N individual specific fixed effects while $\varepsilon_{i,t}$ are idiosyncratic error terms. The fixed effects let us model the heterogeneity of the individuals in a flexible way not possible in the cross sectional model $y_i = x_i'\beta + \eta_i + \varepsilon_i$ where the η_i are unidentifiable and all individual specific variation must be pushed to the error term. It is common to think of the subscript t as time such that one observes N individuals in T time periods. Applications of panel data are widespread: ranging from wage regressions where one seeks to explain worker's salary, to models of economic growth where one seeks to determine the factors that impact growth over time of a panel of countries. In wage regressions the fixed effects are often interpreted as accounting for unobserved

*Aarhus University and CREATES. Email: akock@creates.au.dk. Financial support from the Center for Research in the Econometric Analysis of Time Series (grant DNRF78) is gratefully acknowledged.

†Cambridge University, Faculty of Economics. Email: hht23@cam.ac.uk.

characteristics such as ability, intelligence or perseverance of a worker while in growth models they are often understood as accounting for culture of a country or other unobserved factors remaining (approximately) constant over time. As an example of the latter, Islam (1995) used exactly a dynamic panel data model to study growth convergence of a panel of countries and reached conclusions vastly different from the ones obtained by a plain linear regression model which ignores the repeated sampling and individual specific heterogeneity.

Recent years have witnessed a great deal of research on high-dimensional models with particular emphasis on the linear regression model. Among the most popular procedures is the Lasso of Tibshirani (1996) which sparked a lot of research on estimators possessing the oracle property of Fan and Li (2001), such as e.g. the adaptive Lasso of Zou (2006) or the Bridge estimator of Huang et al. (2008) among others. Subsequently to the oracle property, which is an asymptotic one, a lot of focus has been devoted to establishing finite sample oracle inequalities. That is, upper bounds on the estimation and prediction error that are valid with a guaranteed probability for a fixed sample size, see Candes and Tao (2007); Bickel et al. (2009); Bunea et al. (2007); Zhang and Huang (2008); Meinshausen and Yu (2009); van de Geer (2008); Negahban et al. (2012) among many others. However, until recently, not much work had been done on inference in high-dimensional models for Lasso-type estimators as these possess a rather complicated distribution even in the low dimensional case, see Knight and Fu (2000). This problem has been cleverly approached by unpenalized estimation after double selection by Belloni et al. (2012, 2014) or by desparsification in Zhang and Zhang (2014); van de Geer et al. (2014); Javanmard and Montanari (2013); Caner and Kock (2014). See also Lockhart et al. (2014); Meinshausen (2013) and Nickl and van de Geer (2013).

The focus in the above mentioned work has been almost exclusively on independent data and often on the plain linear regression model while high-dimensional panel data has not been treated. Exceptions are Kock (2013) and Belloni et al. (2014) who have established oracle inequalities and asymptotically valid inference for a low-dimensional parameter in *static* panel data models, respectively. Caner and Zhang (2014) have studied the properties of penalized GMM, which can be used to estimate dynamic panel data models, in the case of fewer parameters than observations. To the best of our knowledge, no research has been conducted on inference in high-dimensional dynamic panel data models. Note that high-dimensionality may arise from three sources in the dynamic panel data model (1.1). These sources are the coefficients pertaining to the lagged left hand side variables (α_l), the exogenous variables (β), as well as the fixed effects (η_i). In particular, we shall see that (joint) inference involving an η_i behaves markedly different from inference only involving α_l's and β. Furthermore, panel data differs from the classical linear regression model in that one does not have independence across $t = 1, ..., T$ for any i as consecutive observations in time can be highly correlated for any given individual. Ignoring this dependence may lead to gravely misleading inference even in low-dimensional panel data models. For that reason we shall make *no* assumptions on this dependence structure across $t = 1, ..., T$ for the $x_{i,t}$. Static panel data models are a special case of (1.1) corresponding to $\alpha_l = 0,\ l = 1, ..., L_N$.

Traditional approaches to inference in low-dimensional static panel data models have considered the N fixed effects η_i as nuisance parameters which have been removed by taking either first differences or demeaning the data over time for each individual i, see e.g. Wooldridge (2010); Arellano (2003); Baltagi (2008). However, this approach does not work as straightforwardly in dynamic panel data models as first differences or demeaning results in a model where the error terms and covariates are correlated. Furthermore, in this paper we take the stand that the fixed effects may be of intrinsic interest, and thus one may also be interested in testing hypothesis on these instead of removing them.

By now it has become common practice to assume ℓ_0-sparsity in high-dimensional models. That is, only a subset of the many potential explanatory variables are truly relevant such that many coefficients are zero. However, classical ℓ_0-sparsity may be a severe restriction to impose in particular on the vector of fixed effects $(\eta_1, ..., \eta_N)$. For that reason, we only impose an assumption of ℓ_1-sparsity on the fixed effects which nests ℓ_0-sparsity as a special case. In the low-dimensional case Bonhomme and Manresa (2012) have assumed a different type of structure, namely grouping, on the fixed effects. However, in the high-dimensional setting we are considering ℓ_1-sparsity works well.

Our inferential procedure is closest in spirit to the one in van de Geer et al. (2014) who cleverly used nodewise regressions to *desparsify* the Lasso and to construct an approximate inverse of the non-invertible sample Gram matrix in the context of the linear regression model. In particular, we show how nodewise regressions can be used to construct one of the blocks of the approximate inverse of the empirical Gram matrix in dynamic panel data models. More precisely, we contribute by first establishing an oracle inequality for a version of the Lasso in dynamic panel data models for all groups of parameters. As can be expected, the fixed effects turn out to behave differently from the remaining parameters. Next, we show how joint asymptotically gaussian inference may be conducted on the three types of parameters in (1.1). In particular, we show that hypotheses involving an increasing number of parameters can be tested and provide a uniformly consistent estimator of the asymptotic covariance matrix which is robust to conditional heteroskedasticity. Thus, we introduce a feasible procedure for inference in high-dimensional heteroskedastic dynamic panel data models. Allowing for conditional heteroskedasticity is important in dynamic models like the one considered here as the conditional variance is known to often depend on the current state of the process, see e.g. Engle (1982). Thus, assuming the error terms to be independent of the covariates with a constant variance is not reasonable. Next, we show that confidence bands constructed by our procedure are asymptotically honest (uniform) in the sense of Li (1989) over a certain subset of the parameter space. Finally, we show that the confidence bands have uniformly the optimal rate of contraction for all types of parameters. Thus, the honesty is not bought at the price of wide confidence bands as is the case for sparse estimators, c.f. Pötscher (2009). Simulations reveal that our procedure performs well in terms of size, power, and coverage rate of the constructed intervals.

The rest of the paper is organized as follows. Section 2 introduces the estimator and provides an oracle inequality for all types of parameters. Next, Section 3 shows how limiting gaussian inference may be be conducted and provides a feasible estimator of the covariance matrix which is robust to heteroskedasticity even in the case where the number of parameter estimates we seek the limiting distribution for diverges with the sample size. Section 4 shows that confidence intervals constructed by our procedure are honest and contract at the optimal rate for all types of parameters. Section 5 studies our estimator in Monte Carlo experiments while Section 6 concludes. All the proofs of our results are deferred to Appendix A; Appendix B contains further auxiliary lemmas needed in Appendix A.

2 The Model

2.1 Notation

For $x \in \mathbb{R}^n$, let $\|x\|_0 = \sum_{i=1}^n 1(x_i \neq 0)$, $\|x\| = \sqrt{\sum_{i=1}^n x_i^2}$, $\|x\|_1 = \sum_{i=1}^n |x_i|$ and $\|x\|_\infty = \max_{1\leq i\leq n} |x_i|$ denote the ℓ_0, ℓ_2, ℓ_1 and ℓ_∞ norms, respectively. Let e_m denote the unit column vector with mth entry being 1 in some Euclidean space depending on the context. If the argument of $\|\cdot\|_\infty$ is a matrix, then $\|\cdot\|_\infty$ denotes the absolute elementwise maximum of the matrix. For some generic set $R \subseteq \{1, \dots, n\}$, let $x_R \in \mathbb{R}^{|R|}$ denote the vector obtained by extracting the elements of $x \in \mathbb{R}^n$ whose indices are in R, where $|R|$ denotes the cardinality of R; $R^c = \{1, \dots, n\} \setminus R$. For an $n \times n$ matrix A, A_R denotes the submatrix consisting of the rows and columns indexed by R. $\otimes$ is the Kronecker product. Let $a \vee b$ and $a \wedge b$ denote $\max(a, b)$ and $\min(a, b)$, respectively. For two real sequences (a_n) and (b_n), $a_n \lesssim b_n$ means that $a_n \leq C b_n$ for some fixed, finite and positive constant C for all $n \geq 1$. $\text{sgn}(\cdot)$ is the sign function. $\text{maxeval}(\cdot)$ and $\text{mineval}(\cdot)$ are the maximal and minimal eigenvalues of the argument, respectively. For some vector $x \in \mathbb{R}^n$, $\text{diag}(x)$ gives a $n \times n$ diagonal matrix with x supplying the diagonal entries.

The model in (1.1) can be rewritten as

$$y_{i,t} = z_{i,t}'\alpha + \eta_i + \varepsilon_{i,t}, \quad i = 1, ..., N, \ t = 1, ..., T \tag{2.1}$$

where $z_{i,t} := (y_{i,t-1}, \ldots, y_{i,t-L_N}, x_{i,t}')'$ and $\alpha := (\alpha_1, \ldots, \alpha_{L_N}, \beta')'$ are $p_N \times 1$ vectors ($p_N = p_{x,N} + L_N$). The subscript N indicates that dimensions of T_N, L_N, $p_{x,N}$ and p_N vary with the cross-sectional dimension N but we suppress the subscript N whenever no confusion arises. Note however, that in asymptotic arguments where N as well as T tend to infinity, we shall often think of T being a function of N. We assume that initial observations $y_{i,0}, y_{i,-1}, \ldots, y_{i,1-L}$ are available for $i = 1, ...N$. Next, (2.1) may be written more compactly as

$$y_i = Z_i'\alpha + \eta_i\iota + \varepsilon_i,$$

where $Z_i := (z_{i,1}, \ldots, z_{i,T})$ is a $p \times T$ matrix, $y_i := (y_{i,1}, \ldots, y_{i,T})'$, $\varepsilon_i := (\varepsilon_{i,1}, \ldots, \varepsilon_{i,T})'$, and ι is a $T \times 1$ vector of ones. Then, one can write

$$y = (Z \quad D)\begin{pmatrix} \alpha \\ \eta \end{pmatrix} + \varepsilon =: \Pi\gamma + \varepsilon,$$

where $Z := (Z_1, \ldots, Z_N)'$, $y := (y_1', \ldots, y_N')'$ and $\varepsilon := (\varepsilon_1', \ldots, \varepsilon_N')'$. $\eta := (\eta_1, \ldots, \eta_N)'$ contains the fixed effects, $D := I_N \otimes \iota$, and $\Pi := (Z, D)$. Finally, $\gamma := (\alpha', \eta')'$ contains all $p + N$ parameters of the model. Thus the dynamic panel model (1.1) can be written more compactly as something resembling a linear regression model. There are several differences, however. First, blocks of rows in the data matrix Π may be heavily dependent. Second, we shall see that α and η have markedly different properties as a result of the fact that the probabilistic properties of the blocks of a properly scaled version of the Gram matrix pertaining to Π are very different. Third, imposing ℓ_1-sparsity only on η implies that the oracle inequalities which we use as a stepping stone towards inference do not follow directly from the technique in, e.g., Bickel et al. (2009). In fact, we do not get explicit expressions for the upper bounds but instead characterize them as solutions to certain quadratic equations in two variables.

2.2 ℓ_1-sparsity and the Panel Lasso

Let $J_1 = \{j : \alpha_j \neq 0, j = 1, \ldots, p\}$ denote the active set of lagged left hand side variables and $x_{i,t}$ with $s_1 = |J_1|$. In this sense α is said to be ℓ_0-sparse with the sparsity index s_1. ℓ_0-sparsity is by now a standard assumption in the high-dimensional statistics. On the other hand, and as already mentioned in the introduction, the fixed effects η might not be ℓ_0-sparse as unobserved heterogeneity might have an effect for all individuals/countries. For example, the effect of the often unmeasured intelligence of a worker may not have a zero impact for any worker in a wage model. However, we believe it is much more reasonable to assume that there exists a group of individuals either well below or well above average intelligence for which η_i is either very negative or very positive while a big group of the populations may be considered "average" in the sense that their intelligence only marginally impacts their salary after controlling for a high-dimensinal vector of $x_{i,t}$ and lagged $y_{i,t}$. This is certainly the case when modeling $y_{i,t}$ in deviations from its population mean. Similarly, there are many other examples where a large group may be conjectured to have very small individual specific fixed effects. This motivates ℓ_1-sparsity.

Let J_2 be the indices of the s_2 largest elements in $\{|\eta_i|, i = 1, ..., N\}$, i.e. the s_2 largest fixed effects in absolute value. Assume that $\|\eta_{J_2^c}\|_1 \leq c_N$ for some $c_N \geq 0$. In words, the s_2 largest fixed effects are unrestricted while the ℓ_1-norm of the remaining fixed effects can not be larger than c_N. Note that $c_N = 0$ is equivalent to ℓ_0-sparsity. ℓ_1-sparsity seems to have been used first in the high-dimensional statistics literature by Zhang and Huang (2008) and is useful in dynamic panel data models where the matrix D has a natural scale such that J_2 is uniquely identified (J_2 does not depend on any specific normalization). Finally, we remark that ignoring the absence of a natural scale of the random and temporally dependent $y_{i,t}$ and $x_{i,t}$ one could also impose ℓ_1-sparsity on α. However, we refrain from this generalization here.

Our starting point for inference is the minimiser $\hat{\gamma} = (\hat{\alpha}', \hat{\eta}')'$ of the following panel Lasso objective function

$$L(\gamma) = \|y - \Pi\gamma\|^2 + 2\lambda_N\|\alpha\|_1 + 2\frac{\lambda_N}{\sqrt{N}}\|\eta\|_1. \tag{2.2}$$

As usual λ_N is a positive regularization sequence. Note that we penalize α and η differently to reflect the fact that we have NT observations to estimate α_j for $j = 1, ..., p$ while only T observations are available to estimate each η_i. The minimization problem can be solved easily as it simply corresponds to a weighted Lasso with known weights. However, the probabilistic analysis of the properly scaled Gram matrix is different from the one for the standard Lasso as it must be broken into several steps. We now turn to the assumptions needed for our inferential procedure.

Assumption 1.
$\{(x_{i,1}', \dots, x_{i,T}', \varepsilon_i')\}_{i=1}^N$ *is an independent sequence and*

$$\mathbb{E}[\varepsilon_{i,t}|y_{i,t-1}, ..., y_{i,1-L}, x_{i,t}, ..., x_{i,1}] = 0 \qquad \textit{for} \qquad i = 1, ..., N,\ t = 1, ..., T.$$

Assumption 1 imposes independence across $i = 1, ..., N$ which is standard in the panel data literature, see e.g. Wooldridge (2010) or Arellano (2003). Note however, that we do not assume the data to be identically distributed across $i = 1, ..., N$. Assumption 1 also implies, by iterated expectations, that the error terms form a martingale difference sequence with respect to the filtration generated by the variables in the above conditioning and thus restricts the degree of dependence in the error terms across t.[1] However, it still allows for considerable dependence over time, as higher moments than the first are not restricted. Furthermore, the error terms need not be identically distributed over time for any individual. We also note that Assumption 1 does not rule out that the error terms are conditionally heteroskedastic. In particular, they may be *autoregressively* conditionally heteroskedastic (ARCH). In panel terminology, both lags of $y_{i,t}$ and $x_{i,t}$ are called *predetermined* or *weakly exogenous*.

In order to introduce the next assumption define the scaled empirical Gram matrix

$$\Psi_N = S^{-1}\Pi'\Pi S^{-1} = \begin{pmatrix} \frac{1}{NT}Z'Z & \frac{1}{T\sqrt{N}}Z'D \\ \frac{1}{T\sqrt{N}}D'Z & \mathrm{I}_N \end{pmatrix} \quad \text{where} \quad S = \begin{pmatrix} \sqrt{NT}\mathrm{I}_p & 0 \\ 0 & \sqrt{T}\mathrm{I}_N \end{pmatrix}$$

When $p + N > NT$, Ψ_N is singular. However, it suffices that a compatibility type condition tailored to the panel data structure is satisfied. To be precise we define

$$\kappa^2(A, r_1, r_2) := \min_{\substack{R_1 \subseteq \{1,...,p\}, |R_1| \le r_1 \\ R_2 \subseteq \{1,...,N\}, |R_2| \le r_2 \\ R := R_1 \cup R_2}} \min_{\substack{\delta \neq 0, \\ \|\delta_{R^c}\|_1 \le 4\|\delta_R\|_1}} \frac{\delta' A \delta}{\frac{1}{r_1 + r_2}\|\delta_R\|_1^2}$$

[1]It can also be verified that $\{\varepsilon_{i,t}\}_{t=1}^T$ forms a martingale difference sequence with respect to the natural filtration for all $i = 1, \dots, N$. This is because the $\varepsilon_{i,t}$ are (linear) functions of the variables in the conditioning set in Assumption 1.

[2]and we shall later see that $\kappa^2(\Psi_N, r_1, r_2)$ is bounded away from zero provided $\kappa^2(\Psi, r_1, r_2)$ is bounded away from zero, where

$$\Psi = \begin{pmatrix} \Psi_Z & 0 \\ 0 & \mathrm{I}_N \end{pmatrix} := \begin{pmatrix} \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\mathbb{E}[z_{i,t}z_{i,t}'] & 0 \\ 0 & \mathrm{I}_N \end{pmatrix}.$$

[3] It is of interest that the compatibility constant $\kappa^2(\Psi_N, r_1, r_2)$ does not depend on the level c_N of ℓ_1-sparsity. The reasons become clear from the proof of Theorem 1 below.

Assumption 2. $\kappa^2 = \kappa^2(\Psi, s_1, s_2)$ *is uniformly bounded away from zero.*

Assumption 2 is rather innocent as it is trivially satisfied when the Ψ is positive definite as is often imposed. Compatibility type conditions are standard in the literature and various versions and their interrelationship have been investigated in van de Geer et al. (2009).

Assumption 3. *There exist positive constants C and K such that*

(a) $\varepsilon_{i,t}$ are uniformly subgaussian; that is, $\mathbb{P}(|\varepsilon_{i,t}| \geq \epsilon) \leq \frac{1}{2}Ke^{-C\epsilon^2}$ for every $\epsilon \geq 0$, $i = 1, \dots, N$ and $t = 1, \dots, T$.

(b) $z_{i,t,l}$ are uniformly subgaussian; that is, $\mathbb{P}(|z_{i,t,l}| \geq \epsilon) \leq \frac{1}{2}Ke^{-C\epsilon^2}$ for every $\epsilon \geq 0$, $i = 1, \dots, N$, $t = 1, \dots, T$ and $l = 1, \dots, p$.

In the context of the plain static regression model it is common practice to assume the error terms as well as the covariates to be subgaussian. However, this assumption is not quite innocent in the context of the dynamic panel data model (1.1) as $y_{i,t}$ is generated by the model and its properties are thus completely determined by those of $x_{i,t}, \varepsilon_{i,t}$ as well as the parameters of the model. Lemma 2 in Appendix A shows that $y_{i,t}$ is subgaussian if $x_{i,t}$ and $\varepsilon_{i,t}$ satisfy this property and the parameters are well-behaved. In particular, a wide class of (causal) stationary processes are included.

2.3 The Oracle Inequalities

With the above assumptions in place we are ready to state our first result. Defining $\mathcal{F}(s_1, s_2, c_N) := \{\alpha \in \mathbb{R}^p : \|\alpha\|_0 \leq s_1\} \times \{\eta \in \mathbb{R}^N : \exists\, I \subseteq \{1, ..., N\} \text{ with } |I| \leq s_2 \text{ and } \|\eta_{I^c}\|_1 \leq c_N\}$, one has

Theorem 1 (Oracle inequalities). *Let Assumptions 1 - 3 hold. Then, choosing $\lambda_N = \sqrt{4MNT(\log(p \vee N))^3}$ for some $M > 0$, the following inequalities are valid with probability at least*

$$1 - Ap^{1-BM^{1/3}} - AN^{1-BM^{1/3}} - A(p^2 + pN)e^{-B\left(\frac{N}{(s_1+s_2)^2}\right)^{1/3}}$$

for positive constants A and B and $s_1 + s_2 \leq \sqrt{N}$.

$$\frac{1}{NT}\left\|\Pi(\hat{\gamma} - \gamma)\right\|^2 \leq \frac{120\lambda_N^2(s_1 + s_2)}{\kappa^2(NT)^2} + \frac{20\lambda_N}{NT}\frac{c_N}{\sqrt{N}}$$

$$\|\hat{\alpha} - \alpha\|_1 \leq \frac{120\lambda_N(s_1 + s_2)}{\kappa^2 NT} + 20\frac{c_N}{\sqrt{N}}$$

$$\|\hat{\eta} - \eta\|_1 \leq \frac{120\lambda_N(s_1 + s_2)}{\kappa^2\sqrt{NT}} + 20c_N.$$

Moreover, the above bounds are valid uniformly over $\mathcal{F}(s_1, s_2, c_N)$.

---

[2]Here $R_1 \cup R_2$ is understood as $R_1 \cup (R_2 + p)$ where the addition is elementwise.
[3]Ψ actually also depends on N and T but for brevity we are silent about this.

Theorem 1 provides oracle inequalities for the prediction error as well as the estimation error of the parameter vectors. While these bounds are of independent interest we shall primarily use them as means towards our ultimate end of conducting (joint) inference on α and η. We stress that the bounds in Theorem 1 are finite sample bounds; they hold for any fixed values of N and T. The special case of ℓ_0-sparsity corresponds to $c_N = 0$ and simplifies the above bounds. However, in the case where one is only interested in $\|\hat{\alpha} - \alpha\|_1$, it suffices that $c_N = o(\sqrt{N})$ in order to make the contribution from c_N vanish asymptotically. c_N of this order is not unreasonable, as the number of lags of $y_{i,t}$ and covariates in $x_{i,t}$ can increase fast in N, thus leaving less unexplained heterogeneity for the fixed effects as more variation in $y_{i,t}$ is explained by the other covariates, resulting in $c_N \to 0$.

We also note that the oracle inequalities are not obtained in an entirely standard manner as the mixture of ℓ_0- and ℓ_1-sparsity in dynamic panel data models calls for a different proof technique which yields the upper bounds as solutions to certain quadratic equations. Finally, we remark that in analogy to oracle inequalities in the plain linear regression model the number of covariates in $x_{i,t}$ (p_x) may increase at an exponential rate in NT without hindering the right hand sides of the oracle inequalities in being small. Furthermore, we do not assume independence across $t = 1, ..., T$ for any individual thus altering the standard probabilistic analysis as well. Instead we use concentration inequalities for martingales to obtain bounds almost as sharp as in the completely independent case.

3 Inference

Following the idea of van de Geer et al. (2014) we next desparsify the Lasso estimator in order to conduct inference. To this end, we construct an approximate inverse of Ψ_N. One sub-block of the approximate inverse will be constructed by nodewise regressions, while we use the structure of the lower right $N \times N$ block of Ψ_N to directly construct an inverse for that part.

3.1 The Desparsified Lasso Estimator $\tilde{\gamma}$

First, observe that $L(\gamma)$ in (2.2) is convex in γ and that $\hat{\gamma}$ satisfies the Karush-Kuhn-Tucker (KKT) conditions

$$0 \in \partial L(\hat{\gamma}) = \begin{pmatrix} -2Z'(y - \Pi\hat{\gamma}) + 2\lambda_N \hat{\kappa}_1 \\ -2D'(y - \Pi\hat{\gamma}) + 2\frac{\lambda_N}{\sqrt{N}}\hat{\kappa}_2 \end{pmatrix}$$

where $\hat{\kappa}_1$ and $\hat{\kappa}_2$ are $p \times 1$ and $N \times 1$ vectors, respectively, such that $\hat{\kappa}_{1j} \in [-1, 1]$ with $\hat{\kappa}_{1j} = \text{sgn}(\hat{\alpha}_j)$ if $\hat{\alpha}_j \neq 0$ for $j = 1, \dots, p$. Similarly, $\hat{\kappa}_{2i} \in [-1, 1]$ with $\hat{\kappa}_{2i} = \text{sgn}(\hat{\eta}_i)$ if $\hat{\eta}_i \neq 0$ for $i = 1, \dots, N$. Hence,

$$-\Pi'\left(y - \Pi\hat{\gamma}\right) + \begin{pmatrix} \lambda_N \hat{\kappa}_1 \\ \frac{\lambda_N}{\sqrt{N}}\hat{\kappa}_2 \end{pmatrix} = 0. \tag{3.1}$$

Using that $y = \Pi\gamma + \varepsilon$ and multiplying by S^{-1} yields

$$\Psi_N S\left(\hat{\gamma} - \gamma\right) + S^{-1}\begin{pmatrix} \lambda_N \hat{\kappa}_1 \\ \frac{\lambda_N}{\sqrt{N}}\hat{\kappa}_2 \end{pmatrix} = S^{-1}\Pi'\varepsilon.$$

Usually one would proceed by isolating $S(\hat{\gamma} - \gamma)$ which implies inverting Ψ_N. However, when $p + N > NT$, Ψ_N is not invertible. The idea of van de Geer et al. (2014) and Javanmard and Montanari (2013) is to instead use an approximate inverse of Ψ_N. Suppose that $\hat{\Theta}$ is a reasonable approximation to such an inverse and rewrite the above display as

$$S\left(\tilde{\gamma} - \gamma\right) = \hat{\Theta}S^{-1}\Pi'\varepsilon - \Delta, \tag{3.2}$$

where

$$\tilde{\gamma} := \hat{\gamma} + S^{-1}\hat{\Theta}S^{-1}\begin{pmatrix} \lambda_N \hat{\kappa}_1 \\ \frac{\lambda_N}{\sqrt{N}}\hat{\kappa}_2 \end{pmatrix}, \qquad \Delta := \left(\hat{\Theta}\Psi_N - \mathrm{I}\right) S\left(\hat{\gamma} - \gamma\right).$$

Thus, $\tilde{\gamma}$ is the *desparsified* Lasso estimator in the dynamic panel context. Clearly, it is non-sparse as it adds a bias correction term to the sparse $\hat{\gamma}$. Δ is the error resulting from using an approximate inverse $\hat{\Theta}$ as opposed to an exact inverse. Note also that by (3.1) $\tilde{\gamma}$ can be easily calculated in practice. Thus, for any $(p+N)\times 1$ vector ρ with $\|\rho\| = 1$ we shall study the asymptotic behavior of

$$\rho' S\left(\tilde{\gamma} - \gamma\right) = \rho'\hat{\Theta}S^{-1}\Pi'\varepsilon - \rho'\Delta. \tag{3.3}$$

In order to conduct asymptotically gaussian inference for $\tilde{\gamma}$ it thus suffices to establish a central limit theorem for $\rho'\hat{\Theta}S^{-1}\Pi'\varepsilon$ as well as to show that $\rho'\Delta$ is asymptotically negligible. Furthermore, we shall provide a feasible estimator of the asymptotic variance of $\rho'\hat{\Theta}S^{-1}\Pi'\varepsilon$ even in the presence of conditional heteroskedasicity. A leading special case of (3.3) is when one is only interested in the asymptotic distribution of $\tilde{\gamma}_j$ corresponding to $\rho = e_j$ being the j'th basis vector of $\mathbb{R}^{p+N}$. In general, we will be interested in the asymptotic distribution of a subset $H \subseteq \{1, ..., p+N\}$ of the indices of γ with cardinality h and shall show that asymptotically honest gaussian inference is possible even for $h \to \infty$, H simultaneously involving elements of α and η, in the presence of heteroskedasticity with feasible covariance matrix estimation.

3.2 Construction of $\hat{\Theta}$

As is clear from the discussion above we need a good choice for $\hat{\Theta}$. In particular we shall show that

$$\hat{\Theta} = \begin{pmatrix} \hat{\Theta}_Z & 0 \\ 0 & \mathrm{I}_N \end{pmatrix}$$

works well. Here $\hat{\Theta}_Z$ will be constructed using nodewise regressions as in van de Geer et al. (2014) and we show that this is possible even when the rows of Z are not independent and identically distributed. The construction of $\hat{\Theta}_Z$ parallels the one in van de Geer et al. (2014) to a high extend but we shall sketch it here in our context as some of the definitions are needed again later. First, define

$$\hat{\phi}_j = \underset{\delta\in\mathbb{R}^{p-1}}{\operatorname{argmin}}\left\{\frac{1}{NT}\|z_j - Z_{-j}\delta\|^2 + 2\lambda_{node}\|\delta\|_1\right\}, \qquad j = 1, ..., p, \tag{3.4}$$

where z_j is the jth column of Z, Z_{-j} is the $NT\times(p-1)$ submatrix of Z with Z's jth column removed, and the $(p-1)\times 1$ vector $\hat{\phi}_j = \{\hat{\phi}_{j,k} : k = 1, \ldots, p, k \neq j\}$. Next, define

$$\hat{C} = \begin{pmatrix} 1 & -\hat{\phi}_{1,2} & \cdots & -\hat{\phi}_{1,p} \\ -\hat{\phi}_{2,1} & 1 & \cdots & -\hat{\phi}_{2,p} \\ \vdots & \vdots & \ddots & \vdots \\ -\hat{\phi}_{p,1} & -\hat{\phi}_{p,2} & \cdots & 1 \end{pmatrix}$$

and $\hat{\tau}_j^2 = \frac{1}{NT}\|z_j - Z_{-j}\hat{\phi}_j\|^2 + \lambda_{node}\|\hat{\phi}_j\|_1$ as well as $\hat{T}^2 = \operatorname{diag}(\hat{\tau}_1^2, \ldots, \hat{\tau}_p^2)$. Finally, we set $\hat{\Theta}_Z = \hat{T}^{-2}\hat{C}$. Let $\hat{C}_j$ denote the jth row of $\hat{C}$ and let $\hat{\Theta}_{Z,j}$ denote the jth row of $\hat{\Theta}_Z$ but both written as a $p\times 1$ vectors. Then, $\hat{\Theta}_{Z,j} = \hat{C}_j/\hat{\tau}_j^2$. For any $j = 1, ..., p$, the KKT condition for a minimum in (3.4) are

$$-\frac{1}{NT}Z_{-j}'(z_j - Z_{-j}\hat{\phi}_j) + \lambda_{node}w_j = 0, \tag{3.5}$$

where w_j is the subdifferential of the function $f(x) = \|x\|$ evaluated at $\hat{\phi}_j$. Using this, the definition of $\hat{\tau}_j$, and $\hat{\phi}_j' w_j = \|\hat{\phi}_j\|_1$ yields

$$\hat{\tau}_j^2 = \frac{1}{NT}(z_j - Z_{-j}\hat{\phi}_j)'(z_j - Z_{-j}\hat{\phi}_j) + \lambda_{node} \|\hat{\phi}_j\|_1 = \frac{1}{NT}(z_j - Z_{-j}\hat{\phi}_j)' z_j. \tag{3.6}$$

Thus, by the definition of $\hat{\Theta}_{Z,j}$, and as $\hat{\tau}_j^2$ is bounded away from zero (we shall later argue rigorously for this)

$$\frac{1}{NT} z_j' Z \hat{\Theta}_{Z,j} = 1. \tag{3.7}$$

Furthermore, the KKT conditions (3.5) can also be written as

$$\frac{1}{NT} Z_{-j}'(z_j - Z_{-j}\hat{\phi}_j) = \lambda_{node} w_j, \tag{3.8}$$

which implies $\frac{1}{NT} Z_{-j}' Z \hat{\Theta}_{Z,j} = \lambda_{node} w_j / \hat{\tau}_j^2$. Combining with (3.7) yields

$$\left\| \frac{1}{NT} Z' Z \hat{\Theta}_{Z,j} - e_j \right\|_\infty \leq \frac{\lambda_{node}}{\hat{\tau}_j^2}, \tag{3.9}$$

thus giving an estimate on how close $\hat{\Theta}_Z$ is to being an inverse of the upper left $p \times p$ block of Ψ_N. This result will be used as an important tool in Appendix A when we show that $\hat{\Theta}$ is a good approximate inverse for Ψ_N as a stepping stone towards showing the asymptotic negligibility of $\rho' \Delta$.

3.3 Asymptotic Properties of the Approximate Inverse

In order to show that $\rho' \hat{\Theta} S^{-1} \Pi' \varepsilon$ is asymptotically gaussian one needs to understand the limiting behaviour of $\hat{\Theta}$. We shall show that $\hat{\Theta}$ is close to

$$\Theta = \begin{pmatrix} \Theta_Z & 0 \\ 0 & \mathrm{I}_N \end{pmatrix} := \begin{pmatrix} \Psi_Z^{-1} & 0 \\ 0 & \mathrm{I}_N \end{pmatrix}$$

in an appropriate sense (Ψ_Z is invertible under Assumption 4 below). To that end, note that by Yuan (2010)

$$\Theta_{Z,j,j} = \left[\Psi_{Z,j,j} - \Psi_{Z,j,-j} \Psi_{Z,-j,-j}^{-1} \Psi_{Z,-j,j} \right]^{-1} \quad \text{and} \quad \Theta_{Z,j,-j} = -\Theta_{Z,j,j} \Psi_{Z,j,-j} \Psi_{Z,-j,-j}^{-1}, \tag{3.10}$$

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

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

such that

$$\phi_j = \left(\frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \mathbb{E}[z_{i,t,-j} z_{i,t,-j}'] \right)^{-1} \left(\frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \mathbb{E}[z_{i,t,-j} z_{i,t,j}] \right) = \Psi_{Z,-j,-j}^{-1} \Psi_{Z,-j,j}, \tag{3.11}$$

where $\Psi_{Z,-j,j}$ is the jth column of Ψ_Z with its jth entry removed. Therefore, $\Theta_{Z,j,-j} = -\Theta_{Z,j,j} \phi_j'$ and the jth row of Θ_Z is sparse if and only if ϕ_j is sparse. Furthermore, defining $\zeta_{j,i,t} := z_{i,t,j} - z_{i,t,-j}' \phi_j$ we may write

$$z_{i,t,j} = z_{i,t,-j}' \phi_j + \zeta_{j,i,t}, \qquad \text{for } i = 1, ..., N, \quad t = 1, ..., T.$$

where by the definition of ϕ_j

$$\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\mathbb{E}[z_{i,t,-j}\zeta_{j,i,t}]=0. \tag{3.12}$$

Thus, in light of Theorem 1 it is sensible that $\hat{\phi}_j$ defined in (3.4) is close to ϕ_j (we shall make this more formal in Appendix A). Next, defining

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

observe $\Theta_{Z,j,-j}=-\phi_j'/\tau_j^2$. Thus, we can write $\Theta_Z=T^{-1}C$ where $T=\text{diag}(\tau_1^2,...,\tau_p^2)$ and C is defined similarly to $\hat{C}$ but with ϕ_j replacing $\hat{\phi}_j$ for $j=1,...,p$. Finally, let $\Theta_{Z,j}$ denote the jth row of Θ_Z written as a column vector. In Lemma 1 we will see that $\hat{\phi}_j$ and $\hat{\tau}_j^2$ are close to ϕ_j and τ_j^2, respectively such that $\hat{\Theta}_{Z,j}$ is close to $\Theta_{Z,j}$. Write $\rho=(\rho_1',\rho_2')'$ with $\|\rho\|=1$, where $\rho_1\in\mathbb{R}^p$ and $\rho_2\in\mathbb{R}^N$. Hence define

$$H=H_1\cup\big(H_2+p\big):=\{j:\rho_{1j}\neq 0\}\cup\big(\{i:\rho_{2i}\neq 0\}+p\big),$$

where $|H_1|=h_{1,N}=h_1$, $|H_2|=h_{2,N}=h_2$, $|H|=h=h_1+h_2$, $j=1,\dots,p$ and $i=1,\dots,N$. Next, let

$$s_{node,j}=|S_{node,j}|:=|\{\Theta_{Z,j,k}\neq 0: k=1,\dots,p,k\neq j\}|,$$

and define $\bar{s}:=\max_{j\in H_1}s_{node,j}$.

Assumption 4. *(a) $mineval(\Psi_Z)$ is uniformly bounded away from zero and $maxeval(\Psi_Z)$ is uniformly bounded from above.*

(b) $\frac{(\log p)^3\bar{s}^2}{N}=o(1)$

(c) There exist positive constants C and K such that $\zeta_{j,i,t}$ are uniformly subgaussian; that is, $\mathbb{P}(|\zeta_{j,i,t}|\geq\epsilon)\leq\frac{1}{2}Ke^{-C\epsilon^2}$ for every $\epsilon>0$, $i=1,\dots,N$, $t=1,\dots,T$ and $j=1,\dots,p$.

Assumption 4(a) is standard and strengthens Assumption 2 slightly. Note that it implies that τ_j^2 is uniformly bounded away from zero as $\tau_j^2=1/\Theta_{Z,j,j}\geq 1/\text{maxeval}(\Theta_Z)=\text{mineval}(\Psi_Z)$. Part (b) restricts the rate of growth of $\bar{s}$ and implies in particular that $\bar{s}=o(\sqrt{N})$. It is used in verifying the compatibility condition for the nodewise regressions. Part (c) imposes subgaussianity on the error terms from the nodewise regressions.

Lemma 1. *Let Assumptions 1, 3 and 4 hold. Define $\lambda_{node}=\sqrt{4M(\log p)^3/N}$ for some $M>0$. Then, for M sufficiently large,*

$$\max_{j\in H_1}|\hat{\tau}_j^2-\tau_j^2|=O_p\left(\sqrt{\frac{\bar{s}(\log p)^3}{N}}\right) \tag{3.13}$$

$$\max_{j\in H_1}\frac{1}{\hat{\tau}_j^2}=O_p(1) \tag{3.14}$$

$$\max_{j\in H_1}\left|\frac{1}{\hat{\tau}_j^2}-\frac{1}{\tau_j^2}\right|=O_p\left(\sqrt{\frac{\bar{s}(\log p)^3}{N}}\right) \tag{3.15}$$

$$\max_{j\in H_1}\left\|\hat{\Theta}_{Z,j}-\Theta_{Z,j}\right\|_1=O_p\left(\bar{s}\sqrt{\frac{(\log p)^3}{N}}\right) \tag{3.16}$$

$$\max_{j\in H_1}\left\|\hat{\Theta}_{Z,j}-\Theta_{Z,j}\right\|=O_p\left(\sqrt{\frac{\bar{s}(\log p)^3}{N}}\right) \tag{3.17}$$

$$\max_{j\in H_1}\left\|\hat{\Theta}_{Z,j}\right\|_1=O_p(\sqrt{\bar{s}}) \tag{3.18}$$

Lemma 1 is used as a stepping stone towards the asymptotically gaussian inference as it indicates at what rate $\hat{\Theta}_Z$ approaches Θ_Z uniformly over H_1. Note that for $H_1 = \{1, ..., p\}$, (3.16) provides an upper bound on the induced ℓ_∞-distance between $\hat{\Theta}_Z$ and Θ_Z. However, we only need to control this distance for those indices corresponding to the parameters we seek the joint limiting distribution of. On the other hand, it should be stressed that the uniformity over H_1 of the above results is crucial in establishing the limiting gaussian inference and providing a feasible estimator of the covariance matrix of the parameter estimates.

3.4 The Asymptotic Distribution of $\tilde{\gamma}$

In this section we formalise the discussion in Section 3.1 as Theorem 2. To this end, define

$$\Sigma_{\Pi\varepsilon} = E(S^{-1}\Pi'\varepsilon\varepsilon'\Pi S^{-1}) = \begin{pmatrix} \mathbb{E}\left[Z'\varepsilon\varepsilon'Z/(NT)\right] & \mathbb{E}\left[Z'\varepsilon\varepsilon'D/(\sqrt{N}T)\right] \\ \mathbb{E}\left[D'\varepsilon\varepsilon'Z/(\sqrt{N}T)\right] & \mathbb{E}\left[D'\varepsilon\varepsilon'D/T\right] \end{pmatrix} = \begin{pmatrix} \Sigma_{1,N} & \Sigma_{2,N} \\ \Sigma_{2,N}' & \Sigma_{3,N} \end{pmatrix}.$$

and note that

$$\Sigma_{1,N} = \mathbb{E}\left[\frac{1}{NT}\sum_{i=1}^N\sum_{j=1}^N Z_i\varepsilon_i\varepsilon_j'Z_j'\right] = \frac{1}{NT}\sum_{i=1}^N \mathbb{E}\left[Z_i\varepsilon_i\varepsilon_i'Z_i'\right] = \frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T \mathbb{E}\left[\varepsilon_{i,t}^2 z_{i,t}z_{i,t}'\right],$$

where the second and third equality both follow from Assumption 1. Likewise, $\Sigma_{3,N} = \frac{1}{T}\sum_{i=1}^N \mathbb{E}\left[d_i\varepsilon_i\varepsilon_i'd_i'\right] = \frac{1}{T}\sum_{i=1}^N\sum_{t=1}^T \mathbb{E}\left[\varepsilon_{i,t}^2 d_{i,t}d_{i,t}'\right] = \text{diag}(\frac{1}{T}\sum_{t=1}^T \mathbb{E}[\varepsilon_{1,t}^2], ..., \frac{1}{T}\sum_{t=1}^T \mathbb{E}[\varepsilon_{N,t}^2])$, where d_i' is the ith $T \times N$ block of D, and $d_{i,t}$ is a $N \times 1$ zero vector with the ith entry replaced by 1. In the same manner, $\Sigma_{2,N} = \frac{1}{\sqrt{N}T}\sum_{i=1}^N \mathbb{E}\left[z_i\varepsilon_i\varepsilon_i'd_i'\right] = \frac{1}{\sqrt{N}T}\sum_{i=1}^N\sum_{t=1}^T \mathbb{E}\left[\varepsilon_{i,t}^2 z_{i,t}d_{i,t}'\right]$. In words, $\Sigma_{2,N}$ is a $p \times N$ matrix with its ith column being $\frac{1}{\sqrt{N}T}\sum_{t=1}^T \mathbb{E}[z_{i,t}\varepsilon_{i,t}^2]$. Finally, define the feasible sample counterpart of $\Sigma_{\Pi\varepsilon}$ as

$$\hat{\Sigma}_{\Pi\varepsilon} = \begin{pmatrix} \hat{\Sigma}_{1,N} & \hat{\Sigma}_{2,N} \\ \hat{\Sigma}_{2,N}' & \hat{\Sigma}_{3,N} \end{pmatrix} := \begin{pmatrix} \frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T \hat{\varepsilon}_{i,t}^2 z_{i,t}z_{i,t}' & \frac{1}{\sqrt{N}T}\sum_{i=1}^N\sum_{t=1}^T \hat{\varepsilon}_{i,t}^2 z_{i,t}d_{i,t}' \\ \frac{1}{\sqrt{N}T}\sum_{i=1}^N\sum_{t=1}^T \hat{\varepsilon}_{i,t}^2 d_{i,t}z_{i,t}' & \frac{1}{T}\sum_{i=1}^N\sum_{t=1}^T \hat{\varepsilon}_{i,t}^2 d_{i,t}d_{i,t}' \end{pmatrix},$$

where $\hat{\varepsilon}_{i,t} := y_{i,t} - z_{i,t}'\hat{\alpha} - \hat{\eta}_i$. The following assumptions are needed to establish the validity of asymptotically gaussian inference of our procedure.

Assumption 5. *Let $\tilde{p} := p \vee N \vee T$ and assume*

(a)

$$\frac{(h_1 \vee h_2 1\{h_1 \neq 0\})^2 \bar{s}^2 (\log\tilde{p})^7}{N} = o(1), \qquad \frac{(\log(N \vee T))^3 1\{h_2 \neq 0\}}{T} = o(1).$$

(b)

$$\frac{(h_1^2\bar{s}^2 \vee Nh_2^2)\left[(s_1+s_2) \vee \frac{\sqrt{T}c_N}{\sqrt{(\log(p\vee N))^3}}\right](\log\tilde{p})^5}{NT} = o(1).$$

(c)

$$\frac{(h_1 \vee h_2)\left[\left(\bar{s} \vee (\log\tilde{p})^2\right)1\{h_1 \neq 0\} \vee 1\{h_2 \neq 0\}\right]\left[(s_1+s_2)^2 \vee \frac{Tc_N^2}{(\log\tilde{p})^3}\right](\log\tilde{p})^4}{N} = o(1).$$

(d) mineval($\Sigma_{\Pi\varepsilon}$) is uniformly bounded away from zero and maxeval($\Sigma_{1,N}$) is uniformly bounded from above.

Assumption 5 is slightly stronger than what we actually need in order to prove Theorem 2 but it is less cluttered in terms of notation. Assumption 5 restricts the rate at which p, T, s_1, s_2, $\bar{s}$, h_1 and h_2 are allowed to increase as none of these are assumed to be bounded. First, note that $p = L + p_x$ only enters through its logarithm. Thus, we can allow for very high-dimensional models. Furthermore, h_1 as well as h_2 are allowed to increase with the sample size such that hypotheses of an increasing dimension involving α and η *simultaneously* can be tested. In the classical setting where one is only interested in testing hypotheses on α one has that $h_2 = 0$ such that assumption 5 simplifies. The case of hypotheses only involving the fixed effects η corresponds to $h_1 = 0$ and again the assumptions simplify. We also note that Assumption 5 requires that $\bar{s}$, h_1 and $s_1 + s_2$ to be $o(N^{1/2})$ while h_2 must be $o(T^{1/2})$.

Theorem 2. *Let Assumptions 1, 3, 4, and 5 be satisfied. If, furthermore, $\{\varepsilon_{i,t}\}_{t=1}^{T}$ is an independent sequence for all $i = 1, ..., N$, then*

$$\frac{\rho' S\left(\tilde{\gamma} - \gamma\right)}{\sqrt{\rho' \hat{\Theta} \hat{\Sigma}_{\Pi\varepsilon} \hat{\Theta}' \rho}} \xrightarrow{d} N(0, 1), \tag{3.19}$$

where $\rho = (\rho_1', \rho_2')'$ is a $(p + N) \times 1$ vector, with $\|\rho\| = 1$, $\rho_1 \in \mathbb{R}^p$ and $\rho_2 \in \mathbb{R}^N$. Moreover,

$$\sup_{\gamma \in \mathcal{F}(s_1, s_2, c_N)} \left| \rho' \hat{\Theta} \hat{\Sigma}_{\Pi\varepsilon} \hat{\Theta}' \rho - \rho' \Theta \Sigma_{\Pi\varepsilon} \Theta' \rho \right| = o_p(1). \tag{3.20}$$

Finally, for every fixed set $H \subseteq \{1, ..., N + p\}$ with cardinality h, we have

$$\left[S_H(\tilde{\gamma}_H - \gamma_H)\right]' \left(\hat{\Theta} \hat{\Sigma}_{\Pi\varepsilon} \hat{\Theta}'\right)_H^{-1} \left[S_H(\tilde{\gamma}_H - \gamma_H)\right] \xrightarrow{d} \chi_h^2. \tag{3.21}$$

Theorem 2 provides sufficient conditions under which our procedure allows for asymptotically gaussian inference. We stress again that hypotheses involving an increasing number of parameters can be tested and that the total number of parameters in the model may be much larger than the sample size. Furthermore, the error terms are allowed to be conditionally heteroskedastic and we provide a consistent estimator of the asymptotic covariance matrix even for the case of hypotheses involving an increasing number of parameters. Indeed, this estimator converges uniformly over $\mathcal{F}(s_1, s_2, c_N)$ which we use in establishing the the honesty of confidence intervals based on (3.19) over this set in Theorem 3. van de Geer et al. (2014) have derived similar results in the setting of the homoskedastic linear cross sectional model for the case of inference on a low-dimensional parameter. Thus, our results can be seen as an extension to dynamic panel data models. Relaxing the homoskedasticity assumption is important as volatility is known to vary over time in dynamic models, see e.g. Engle (1982), and the conditional volatility often depends on the state of the process. Theorem 2 is also related to Belloni et al. (2014) who consider inference in static panel data models for a low-dimensional parameter of interest.

The classical setup where one is only interested in inference on α corresponds to $\rho_2 = 0$ such that $\sqrt{NT}\rho_1' (\tilde{\alpha} - \alpha)$ is asymptotically gaussian with variance equal to the limit of $\rho_1' \Theta_Z \Sigma_{1,N} \Theta_Z' \rho_1$ (assumed to exist for illustration). If furthermore, $\varepsilon_{i,t}$ is homoskedastic with variance σ^2 and independent of $z_{i,t}$ for all $i = 1, ..., N$ and $t = 1, ..., T$, it follows from the definition of $\Sigma_{1,N}$ that this variance equals the limit of $\sigma^2 \rho_1' \Theta_Z \rho_1 = \sigma^2 \rho_1' \Psi_Z^{-1} \rho_1$. The leading special case where one is interested in testing a hypothesis on the j'th entry of α corresponds to $\rho_1 = e_j$. Similar reasoning shows that in the case where one is testing hypotheses involving fixed effects only, corresponding to $\rho_1 = 0$, one has that $\rho_2' \sqrt{T} (\tilde{\eta} - \eta)$ is asymptotically gaussian with variance σ^2. This simple form of the variance follows from the asymptotic independence of the components of $\tilde{\eta}$. Note that the different rates of convergence for $\tilde{\alpha}$ and $\tilde{\eta}$ are in accordance with Theorem 1.

(3.21) is a straightforward consequence of (3.19) and reveals that classical χ^2 inference can be carried out in the usual manner. Thus, asymptotically valid χ^2-inference can be performed in order to test a hypothesis on h parameters simultaneously. Wald tests of general restrictions of the type $H_0 : g(\gamma) = 0$ (where $g : \mathbb{R}^{p+N} \to \mathbb{R}^h$ is differentiable in an open neighborhood around γ and has derivative matrix of rank h) can now also be constructed in the usual manner, see e.g. Davidson (2000) Chapter 12, even when $p + N > NT$ which has hitherto been impossible.

Finally, the independence assumption on $\varepsilon_{i,t}$ across t is needed only if one tests hypotheses involving $\{\eta_i\}_{i=1}^N$ ($h_2 \neq 0$). Weaker assumptions the error terms such as mixing are possible at the expense of more involved expression but will not be pursued here.

4 Honest Confidence Intervals

In this section we show that the confidence bands based on (3.19) are honest (uniformly valid) and contract at the optimal rate. The precise result is contained in the following theorem.

Theorem 3. *Let Assumptions 1, 3, 4, and 5 be satisfied. Then, for all* $\rho \in \mathbb{R}^{p+N}$ *with* $\|\rho\| = 1$,

$$\sup_{t\in\mathbb{R}} \sup_{\gamma\in\mathcal{F}(s_1,s_2,c_N)} \left| \mathbb{P}\left(\frac{\rho' S\left(\tilde{\gamma} - \gamma\right)}{\sqrt{\rho'\hat{\Theta}\hat{\Sigma}_{\Pi\varepsilon}\hat{\Theta}'\rho}} \leq t \right) - \Phi(t) \right| = o(1), \tag{4.1}$$

where $\Phi(\cdot)$ *is the CDF of the standard normal distribution. Furthermore, define* $\tilde{\sigma}_{\alpha,j} := \sqrt{[\hat{\Theta}_Z\hat{\Sigma}_{1,N}\hat{\Theta}_Z]_{jj}}$ *and* $\tilde{\sigma}_{\eta,i} := \sqrt{[\hat{\Sigma}_{3,N}]_{ii}}$ *for* $j = 1, ..., p$ *and* $i = 1, ..., N$*, respectively. Then,*

$$\liminf_{N\to\infty} \inf_{\gamma\in\mathcal{F}(s_1,s_2,c_N)} \mathbb{P}\left(\alpha_j \in \left[\tilde{\alpha}_j - z_{1-\delta/2}\frac{\tilde{\sigma}_{\alpha,j}}{\sqrt{NT}}, \tilde{\alpha}_j + z_{1-\delta/2}\frac{\tilde{\sigma}_{\alpha,j}}{\sqrt{NT}}\right]\right) \geq 1 - \delta, \tag{4.2}$$

$$\liminf_{N\to\infty} \inf_{\gamma\in\mathcal{F}(s_1,s_2,c_N)} \mathbb{P}\left(\eta_i \in \left[\tilde{\eta}_i - z_{1-\delta/2}\frac{\tilde{\sigma}_{\eta,i}}{\sqrt{T}}, \tilde{\eta}_i + z_{1-\delta/2}\frac{\tilde{\sigma}_{\eta,i}}{\sqrt{T}}\right]\right) \geq 1 - \delta, \tag{4.3}$$

for $j = 1, ..., p$ *and* $i = 1, ..., N$*, respectively, where* $z_{1-\delta/2}$ *is the* $1-\delta/2$ *percentile of the standard normal distribution. Finally, letting* $diam([a, b]) = b - a$ *be the length (which coincides with the Lebesgue measure of* $[a, b]$*) of an interval* $[a, b]$ *in the real line, we have*

$$\sup_{\gamma\in\mathcal{F}(s_1,s_2,c_N)} diam\left(\left[\tilde{\alpha}_j - z_{1-\delta/2}\frac{\tilde{\sigma}_{\alpha,j}}{\sqrt{NT}}, \tilde{\alpha}_j + z_{1-\delta/2}\frac{\tilde{\sigma}_{\alpha,j}}{\sqrt{NT}}\right]\right) = O_p\left(\frac{1}{\sqrt{NT}}\right), \tag{4.4}$$

$$\sup_{\gamma\in\mathcal{F}(s_1,s_2,c_N)} diam\left(\left[\tilde{\eta}_i - z_{1-\delta/2}\frac{\tilde{\sigma}_{\eta,i}}{\sqrt{T}}, \tilde{\eta}_i + z_{1-\delta/2}\frac{\tilde{\sigma}_{\eta,i}}{\sqrt{T}}\right]\right) = O_p\left(\frac{1}{\sqrt{T}}\right), \tag{4.5}$$

for $j = 1, ..., p$ *and* $i = 1, ..., N$*, respectively.*

(4.1) reveals that the convergence to the normal distribution in Theorem 2 is actually uniform over $\mathcal{F}(s_1, s_2, c_N)$. Since the desparsified Lasso is not a sparse estimator this uniform convergence does not contradict the work of Leeb and Pötscher (2005). Next, (4.2) is a direct consequence of (4.1) and reveals that the desparsified Lasso produces confidence bands which are *honest* (uniform) over $\mathcal{F}(s_1, s_2, c_N)$. Honest confidence bands are important in practical applications of dynamic panel data models as they guarantee the existence of a known N_0, not depending on $\gamma \in \mathcal{F}(s_1, s_2, c_N)$, such that $\left[\tilde{\alpha}_j - z_{1-\delta/2}\frac{\tilde{\sigma}_{\alpha,j}}{\sqrt{NT}}, \tilde{\alpha}_j + z_{1-\delta/2}\frac{\tilde{\sigma}_{\alpha,j}}{\sqrt{NT}}\right]$ covers α_j with probability not much smaller than $1 - \delta$. Here the important point is that one and the same N_0 guarantees this coverage, irrespective of the true value of $\gamma \in \mathcal{F}(s_1, s_2, c_N)$. On the other hand, pointwise consistent confidence bands only guarantee that

$$\inf_{\gamma\in\mathcal{F}(s_1,s_2,c_N)} \liminf_{N\to\infty} \mathbb{P}\left(\alpha_j \in \left[\tilde{\alpha}_j - z_{1-\delta/2}\frac{\tilde{\sigma}_{\alpha,j}}{\sqrt{NT}}, \tilde{\alpha}_j + z_{1-\delta/2}\frac{\tilde{\sigma}_{\alpha,j}}{\sqrt{NT}}\right]\right) \geq 1 - \delta,$$

implying that the value of N needed in order to guarantee a coverage of close to $1 - \delta$ may depend on the *unknown* true parameter. Thus, for some parameter values one may have to sample more data points to achieve the desired coverage than for others which is unfortunate as one does not know for which parameters this is the case. An honest confidence set S_N for α_j can of course trivially be obtained by setting $S_N = \mathbb{R}$. However, this is clearly not very informative and therefore (4.4) is reassuring as it guarantees that the length of the honest confidence interval contracts at the optimal rate. Therefore, our confidence bands are not only honest, they are also very *informative* as they contract as fast as possible. Furthermore, this contraction is uniform over $\mathcal{F}(s_1, s_2, c_N)$. Since the desparsified Lasso is not a sparse estimator, this fast contraction does not contradict inequality 6 in Theorem 2 of Pötscher (2009) who shows that honest confidence bands based on sparse estimators must be large.

Similarly to the confidence bands pertaining to α, the ones for the fixed effects are also honest and contract at the optimal rate. Note that this rate is again slower than the one for α. It is also worth remarking that the above inference results are valid without any sort of lower bound on the non-zero coefficients as inference is not conducted after model selection.

5 Monte Carlo

In this section we investigate the finite sample properties of our estimator by means of simulations. All calculations are carried out in R using the `glmnet` package and λ_N and λ_{node} are chosen via BIC by the formula given in (9.4.9) in Davidson (2000). We compare the results for our estimator to the least squares oracle which only includes the relevant variables. When sample size allows it, that is when $p+N \leq NT$, we also implement naive least squares including all variables. The number of Monte Carlo replications is 1,000 for all setups and we consider the performance of our estimator along the following dimensions.

1. Estimation error: We compute the root mean square errors (RMSE) of all procedures averaged over the Monte Carlo replications.

2. Coverage rate: We calculate the coverage rate of a gaussian confidence interval constructed as in Theorem 3. This is done for a coefficient belonging to the lags of the left hand side variable, the coefficient of a regressor in $x_{i,t}$ as well as a fixed effect.

3. Length of confidence interval: We calculate the length of the three confidence intervals considered in point 2 above.

4. Size: We evaluate the size of the χ^2-test in Theorem 2 for a hypothesis involving the same three parameters we construct confidence intervals for in point 2 above.

5. Power: We evaluate the power of the χ^2-test in point 4 above.

All tests are carried out at the 5% level of significance and all confidence intervals have a nominal coverage of 95%. Furthermore, as our results regarding estimation error are for the plain Lasso, the root mean square errors are reported for this instead of the desparsified Lasso. As our models are dynamic, we allow for a burn-in period of 1,000 observations when generating the data.

The data generating process is (1.1) and in all experiments $(\alpha_1, \alpha_2, \alpha_3, \alpha_4) = (0.9, 0, 0, -0.3)$ such that the roots of the corresponding lag polynomial lie outside the unit disk. The covariance matrix of $x_{i,t}$ is chosen to have a Toeplitz structure with the (i,j)th entry equal to $\rho^{|i-j|}$ with $\rho = 0.75$. We also experimented with other choices of ρ which did not change the results dramatically. Furthermore, we also tried to let the covariance matrix of $x_{i,t}$ be block-diagonal. Again, this did not alter our results. Finally, we implemented the desparsified conservative Lasso of Caner and Kock (2014). However, this only improved the results slightly and so we do not report these results here.

As our theory allows for heteroskedasticity, we also investigate the effect of this. To be precise, we consider error terms of the form $\varepsilon_{i,t} = u_{i,t}\left(x_{i,t,1}/\sqrt{2} + b_x x_{i,t,2}\right)$ where $u_{i,t}$ is independent of $y_{i,t-1}, ..., y_{i,1-L}$ and $x_{i,t}, ..., x_{i,1}$. b_x is chosen such that the unconditional variance of $\varepsilon_{i,t}$ is the same as the one of $u_{i,t}$ which in turns equals the one from the homoskedastic case. A simple calculation reveals that $b_x = \left(-\sqrt{2}\rho_{12} + \sqrt{2\rho_{12}^2 + 2}\right)/2$, where ρ_{12} is the covariance between $x_{i,t,1}$ and $x_{i,t,2}$. Note that $\varepsilon_{i,t}$ constructed this way satisfies Assumption 1. The reason we ensure that the unconditional variance is the same as in the homoskedastic case is that we do not want any findings in the heteroskedastic case to be driven by a plain change in the unconditional variance. The following experiments were carried out

- Experiment 1: (moderate-dimensional setting): $N = 20$ and $T = 10$. β is 50×1 with five equidistant non-zero entries equaling one. Thus, $p = 54$ and $s_1 = 7$. η is 20×1 with four equidistant non-zero entries equaling 1. Thus, $s_2 = 4$ and $c_N = 0$. In total, $\gamma = (\alpha', \eta')'$ is 74×1. Covariates $x_{i,t}$ and error terms are standard gaussian. We test the true hypothesis

  $$H_0 : (\gamma_1, \gamma_{25}, \gamma_{55}) = (0.9, 1, 1)$$

  by the χ^2_3 test described in Theorem 2 in order to gauge the size of the test. The power is investigated by the hypothesis

  $$H_0 : (\gamma_1, \gamma_{25}, \gamma_{55}) = (0.8, 1, 1).$$

  The following variations of this setting are considered

  (a) The baseline case described so far.
  (b) Same as (a), but assuming ℓ_1-sparsity by replacing the zero components of η with 0.1 implying $c_N = 1.6$.
  (c) Same as (b) but with heteroskedastic errors.

- Experiment 2: (high-dimensional setting). $N = 20$ and $T = 10$. β is 400×1 with ten equidistant non-zero entries equaling one. Thus, $p = 404$ and $s_1 = 12$. η is as in Experiment 1(a). In total, $\gamma = (\alpha', \eta')'$ is 424×1. Covariates $x_{i,t}$ and error terms are standard gaussian. We test the true hypothesis

  $$H_0 : (\gamma_1, \gamma_{85}, \gamma_{405}) = (0.9, 1, 1)$$

  by the χ^2_3 test described in Theorem 2 in order to gauge the size of the test. The power is investigated by the hypothesis

  $$H_0 : (\gamma_1, \gamma_{85}, \gamma_{405}) = (0.8, 1, 1).$$

  The following variations of this setting are considered

  (a) The baseline case described so far.
  (b) Same as (a), but assuming ℓ_1-sparsity by replacing the zero components of η with 0.1 implying $c_N = 1.6$.
  (c) Same as (b) but with heteroskedastic errors.

- Experiment 3: (increase T): As Experiment 2 but with $T = 25$.

- Experiment 4: (increase N): As Experiment 2 but with $N = 40$. η has eight non-zero entries and when ℓ_1-sparsity is assumed all non-zero entries are replaced by 0.1 implying $c_N = 3.2$.

		RMSE		Coverage			Length				
		α	η	γ_1	γ_{25}	γ_{55}	γ_1	γ_{25}	γ_{55}	Size	Power
	LS	8.567	8.044	0.829	0.853	0.864	0.102	0.520	1.183	0.255	0.786
1(a)	DL	2.286	3.948	0.914	0.904	0.938	0.102	0.448	1.426	0.120	0.934
	Ora	0.395	1.326	0.946	0.922	0.917	0.057	0.275	1.166	0.101	1.000
	LS	8.567	8.045	0.829	0.853	0.864	0.102	0.520	1.183	0.255	0.786
1(b)	DL	2.435	4.898	0.917	0.906	0.942	0.107	0.470	1.503	0.108	0.931
	Ora	0.424	6.969	0.841	0.908	0.898	0.060	0.274	1.173	0.172	0.995
	LS	8.522	7.932	0.838	0.866	0.859	0.102	0.515	1.153	0.235	0.805
1(c)	DL	2.438	4.968	0.909	0.917	0.954	0.106	0.464	1.474	0.099	0.920
	Ora	0.471	6.907	0.823	0.927	0.894	0.060	0.269	1.120	0.149	0.990

Table 1: Experiment 1. LS, DL and Ora: least squares including all variables, desparsified Lasso and least squares oracle. RMSE: root mean square error. Coverage: the coverage rate of the asymptotic 95% confidence intervals. Length: the average length of the asymptotic 95% confidence intervals. Size: size of the correct hypothesis $H_0 : (\gamma_1, \gamma_{25}, \gamma_{55}) = (0.9, 1, 1)$. Power: the probability to reject the false $H_0 : (\gamma_1, \gamma_{25}, \gamma_{55}) = (0.8, 1, 1)$.

- Experiment 5: (Heavy tails): As experiment 2 but the covariates $x_{i,t}$ and error terms are t-distributed with 3 degrees of freedom.

Table 1 contains the results of experiment 1. Setting 1(a) reveals that the RMSE of the Lasso are lower than those for least squares including all variables but higher than those of least squares only including the relevant variables. This is the case for α as well as the fixed effects. Next, it is very encouraging that the coverage probabilities for the desparsified Lasso are almost as good as those based on the oracle for γ_1 and γ_{25} and actually superior for γ_{55}. By inspecting the length of the confidence intervals it is seen that this superior coverage is due to wider confidence intervals reflecting a more accurate assessment of uncertainty. Finally, the size and power of the χ^2-test based on the desparsifed Lasso are almost as good as the ones for the oracle and clearly superior to least squares including all variables.

Experiment 1(b) relaxes the sparsity assumption on the fixed effects such that *no* fixed effect is assumed to be zero. In this setting, our procedure actually produces confidence bands with coverage probabilities superior to those of the oracle. In fact, the coverage rates of our procedure are unaffected by relaxing the sparsity assumption – they even increase slightly. Interestingly, the size and power are not affected either, resulting in a less size distorted test than the one based on the least squares oracle. Panel 1(c) adds heteroskedasticity and as expected from our theory the results seem robust towards this.

Next, we turn to experiment 2(a) which is high-dimensional. The results can be found in Table 2. As expected, the estimation error is much higher for the Lasso than for the oracle. However, it is encouraging that the confidence intervals produced by the desparsified Lasso have a coverage which is as good as the one for the the oracle and close to the nominal rate of 95%. The price paid is that the confidence bands are wider than the ones based on the oracle. However, we believe it is preferable to have confidence bands with accurate coverage rates which are wide due to an accurate reflection of uncertainty than to have narrow bands which undercover[4]. Experiment 2(b) relaxes the ℓ_0-sparsity assumption and this results in our confidence bands having more accurate coverage than the oracle based ones. Similarly, the χ^2-test has better size properties, but lower power. These findings confirm the ones from the moderate dimensional setting. Adding heteroskedasticty does not alter the findings.

[4]Note that narrow bands can always be obtained by a confidence bands consisting only of the point estimate. However, in general these have a coverage rate of zero.

		RMSE		Coverage			Length				
		α	η	γ_1	γ_{85}	γ_{405}	γ_1	γ_{85}	γ_{405}	Size	Power
	LS										
2(a)	DL	6.450	3.986	0.921	0.941	0.924	0.122	0.665	2.226	0.100	0.888
	Ora	0.760	1.364	0.939	0.941	0.905	0.043	0.274	1.169	0.108	1.000
	LS										
2(b)	DL	6.519	5.565	0.924	0.937	0.923	0.123	0.672	2.244	0.096	0.885
	Ora	0.800	6.978	0.866	0.934	0.883	0.046	0.274	1.172	0.165	1.000
	LS										
2(c)	DL	6.489	5.567	0.920	0.940	0.942	0.122	0.674	2.266	0.079	0.874
	Ora	0.840	6.795	0.865	0.921	0.884	0.046	0.269	1.128	0.153	0.999

Table 2: LS, DL and Ora: least squares including all variables, desparsified Lasso and least squares oracle. RMSE: root mean square error. Coverage: the coverage rate of the asymptotic 95% confidence intervals. Length: the average length of the asymptotic 95% confidence intervals. Size: size of the correct hypothesis $H_0 : (\gamma_1, \gamma_{85}, \gamma_{405}) = (0.9, 1, 1)$. Power: the probability to reject the false $H_0 : (\gamma_1, \gamma_{85}, \gamma_{405}) = (0.8, 1, 1)$.

		RMSE		Coverage			Length				
		α	η	γ_1	γ_{85}	γ_{405}	γ_1	γ_{85}	γ_{405}	Size	Power
	LS	87.277	10.102	0.557	0.546	0.538	0.050	0.328	0.770	0.763	0.992
3(a)	DL	3.319	2.604	0.912	0.932	0.950	0.051	0.286	0.893	0.090	1.000
	Ora	0.466	0.815	0.941	0.947	0.920	0.027	0.175	0.768	0.082	1.000
	LS	87.277	10.103	0.557	0.546	0.538	0.050	0.328	0.770	0.763	0.992
3(b)	DL	3.393	4.157	0.908	0.925	0.953	0.052	0.295	0.927	0.085	1.000
	Ora	0.475	4.121	0.910	0.942	0.921	0.027	0.175	0.768	0.098	1.000
	LS	87.357	10.123	0.533	0.548	0.535	0.049	0.327	0.769	0.761	0.993
3(c)	DL	3.420	4.160	0.912	0.930	0.959	0.052	0.294	0.920	0.083	1.000
	Ora	0.505	4.096	0.937	0.945	0.940	0.027	0.174	0.749	0.063	1.000

Table 3: Experiment 3. LS, DL and Ora: least squares including all variables, desparsified Lasso and least squares oracle. RMSE: root mean square error. Coverage: the coverage rate of the asymptotic 95% confidence intervals. Length: the average length of the asymptotic 95% confidence intervals. Size: size of the correct hypothesis $H_0 : (\gamma_1, \gamma_{85}, \gamma_{405}) = (0.9, 1, 1)$. Power: the probability to reject the false $H_0 : (\gamma_1, \gamma_{85}, \gamma_{405}) = (0.8, 1, 1)$.

In Table 3, T has been increased to 25 compared to Table 2. This results in lower estimation errors for the Lasso as well as oracle assisted least squares. The coverage rates of the confidence bands are unaltered and still close to the nominal rates. However, the bands are now much narrower, thus making them more informative. The power of the χ^2-test has increased to 1 for the desparsified Lasso and the size is comparable to the one of the oracle procedure. Experiment 3(b) relaxes the ℓ_0-sparsity assumption and shows that our method is robust towards this. The same may be concluded from 3(c) which adds heteroskedasticity.

Table 4 increases N to 40 compared to Table 2. This results in more fixed effects to be estimated. Thus, it is not surprising that the estimation error for α goes down while the one for η increases. Furthermore, the confidence bands based on the desparsified Lasso actually have better coverage rates than the oracle assisted ones for the fixed effects. Compared to Table 2 the confidence bands for the coefficients belonging to the lagged left hand side variables and $x_{i,t}$

		RMSE		Coverage			Length				
		α	η	γ_1	γ_{85}	γ_{405}	γ_1	γ_{85}	γ_{405}	Size	Power
	LS										
4(a)	DL	4.498	7.999	0.922	0.966	0.937	0.087	0.493	2.224	0.065	0.995
	Ora	0.535	2.613	0.935	0.930	0.898	0.031	0.195	1.165	0.103	1.000
	LS										
4(b)	DL	4.517	11.199	0.919	0.965	0.937	0.087	0.495	2.227	0.064	0.996
	Ora	0.563	13.434	0.835	0.920	0.886	0.033	0.195	1.166	0.192	1.000
	LS										
4(c)	DL	4.495	11.198	0.928	0.967	0.941	0.087	0.494	2.187	0.066	0.996
	Ora	0.593	13.293	0.833	0.925	0.914	0.032	0.193	1.105	0.145	1.000

Table 4: Experiment 4. LS, DL and Ora: least squares including all variables, desparsified Lasso and least squares oracle. RMSE: root mean square error. Coverage: the coverage rate of the asymptotic 95% confidence intervals. Length: the average length of the asymptotic 95% confidence intervals. Size: size of the correct hypothesis $H_0 : (\gamma_1, \gamma_{85}, \gamma_{405}) = (0.9, 1, 1)$. Power: the probability to reject the false $H_0 : (\gamma_1, \gamma_{85}, \gamma_{405}) = (0.8, 1, 1)$.

have become more narrow while the width of the bands of the fixed effects are unchanged as is to be expected from our theory. Finally, tests based on our method have better size properties than the ones based on the oracle while having similar power. As the ℓ_0-sparsity assumption on the fixed effects is relaxed in experiment 4(b) we note that coverage of our procedure is markedly better than the one of the oracle without the confidence bands becoming wider. Size and power remain unchanged and experiment 4(c) reveals robustness to heteroskedasticity.

		RMSE		Coverage			Length				
		α	η	γ_1	γ_{85}	γ_{405}	γ_1	γ_{85}	γ_{405}	Size	Power
	LS										
5(a)	DL	7.272	4.028	0.924	0.943	0.955	0.118	0.725	3.471	0.066	0.856
	Ora	0.788	2.316	0.933	0.934	0.919	0.042	0.272	1.837	0.088	1.000
	LS										
5(b)	DL	7.297	5.603	0.922	0.942	0.956	0.118	0.728	3.480	0.063	0.856
	Ora	0.830	12.004	0.868	0.926	0.909	0.046	0.271	1.844	0.137	0.992
	LS										
5(c)	DL	10.053	5.619	0.933	0.928	0.956	0.137	0.906	4.239	0.078	0.728
	Ora	1.609	21.326	0.845	0.930	0.928	0.069	0.405	2.689	0.113	0.942

Table 5: Experiment 5. LS, DL and Ora: least squares including all variables, desparsified Lasso and least squares oracle. RMSE: root mean square error. Coverage: the coverage rate of the asymptotic 95% confidence intervals. Length: the average length of the asymptotic 95% confidence intervals. Size: size of the correct hypothesis $H_0 : (\gamma_1, \gamma_{85}, \gamma_{405}) = (0.9, 1, 1)$. Power: the probability to reject the false $H_0 : (\gamma_1, \gamma_{85}, \gamma_{405}) = (0.8, 1, 1)$.

Our final experiment adds heavy-tailed error terms and covariates $x_{i,t}$ to the high-dimensional setting presented in Table 2. The results are contained in Table 5. The estimation error of α and η increases for the Lasso as well as the least squares oracle. On the other hand, the coverage rates remain high but the price paid is that the length of the confidence bands pertaining to the fixed effects increases for both procedures. Experiment 5(b) relaxes the ℓ_0-sparsity assumption on the fixed effects. Compared to Table 2, the oracle estimates the fixed effects markedly less precisely while the Lasso is unaffected. The coverage rates are roughly unchanged in the sense that our procedure still performs uniformly better than the oracle. The addition of heteroskedasticity in experiment 5(c) does not alter our conclusions.

6 Conclusion

This paper has considered inference in high-dimensional dynamic panel data models with fixed effects. In particular we have shown how hypotheses involving an increasing number of variables can be tested. These hypotheses can involve parameters from all groups of variables in the model simultaneously. As a stepping stone towards this inference we constructed a uniformly valid estimator of the covariance matrix of the parameter estimates which is robust towards conditional heteroskedasticity.

Next, we showed that confidence bands based on our procedure are asymptotically honest and contract at the optimal rate. This rate of contraction depends on which type of parameter is under consideration. Simulations revealed that our procedure works well in finite samples. Future work may include relaxing the sparsity assumption on the inverse covariance matrix Θ_Z as well as extending our results to non-linear panel data models.

7 Appendix A

7.1 Sufficient Conditions for $y_{i,t}$ to be Subgaussian

The following Lemma provides sufficient conditions for $y_{i,t}$ to inherit the subgaussianity from the covariates and the error terms. It allows for a wide range of models but rules out dynamic panel data models which are explosive or contain unit roots.

Lemma 2. *Let $x_{i,t,k}$ and $\varepsilon_{i,t}$ be uniformly subgaussian for $i = 1, ..., N$, $t = 1, ..., T$ and $k = 1, ..., p_x$ and assume that $\|\beta\|_1 \leq C$ for some $C > 0$ for all N and T. Furthermore,* $\max_{1\leq i\leq N} |\eta_i|$ *is bounded uniformly in N and T. Then, if all roots of $1 - \sum_{j=1}^{L} \alpha_j z^j$ $(\alpha_1, \ldots, \alpha_L$ fixed) are outside the unit disc, $y_{i,t}$ is uniformly subgaussian for $i = 1, ..., N$ and $t = 1, ..., T$.*

Proof of Lemma 2. Let $y_t = \sum_{j=1}^{L} \alpha_j y_{t-j} + u_t$ be an AR(L) process with roots outside the unit disc. Write the companion form as $\xi_t = F\xi_{t-1} + v_t$. Then, by the monotone convergence theorem for Orlicz norms, see van der Vaart and Wellner (1996) exercise 6, page 105, $\big\|\|\xi_t\|\big\|_{\psi_2} \leq \big\|\sum_{j=1}^{\infty} \|F^j\|_{\ell_2} \|v_{t-j}\|\big\|_{\psi_2} = \sum_{j=1}^{\infty} \|F^j\|_{\ell_2} \big\|\|v_{t-j}\|\big\|_{\psi_2} = \sum_{j=1}^{\infty} \|F^j\|_{\ell_2} \|u_{t-j}\|_{\psi_2}$, where $\|\cdot\|_{\ell_2}$ is the ℓ_2 induced norm, and the last equality used that v_t is $L\times 1$ with only one non-zero entry equaling u_t. By Corollary 5.6.14 in Horn and Johnson (1990) there exists a $\delta > 0$ such that $\|F^j\|_{\ell_2} \leq (1-\delta)^j$ for j sufficiently large. Thus, if $\|u_t\|_{\psi_2}$ is uniformly bounded we conclude $\|y_t\|_{\psi_2} \leq \big\|\|\xi_t\|\big\|_{\psi_2} \leq K$ for some $K > 0$. Thus, in our context it suffices to show that $\|x'_{i,t}\beta + \eta_i + \varepsilon_{i,t}\|_{\psi_2}$ is uniformly bounded as $y_{i,t} = \sum_{j=1}^{L} \alpha_j y_{i,t-j} + x'_{i,t}\beta + \eta_i + \varepsilon_{i,t} = \sum_{j=1}^{L} \alpha_j y_{i,t-j} + u_{i,t}$ with $u_{i,t} = x'_{i,t}\beta + \eta_i + \varepsilon_{i,t}$. But $\|x'_{i,t}\beta + \eta_i + \varepsilon_{i,t}\|_{\psi_2} \leq \sum_{j=1}^{p} |\beta_j| \|x_{i,t,k}\|_{\psi_2} + \|\eta_i\|_{\psi_2} + \|\varepsilon_{i,t}\|_{\psi_2}$ which is bounded by the assumptions made. □

7.2 Proof of Theorem 1

In this section we prove Theorem 1. In order to do so we introduce the events

$$\mathcal{A}_N = \left\{ \|Z'\varepsilon\|_\infty \leq \frac{\lambda_N}{2}, \quad \|D'\varepsilon\|_\infty \leq \frac{\lambda_N}{2\sqrt{N}} \right\}, \quad \mathcal{B}_N = \left\{ \kappa^2(\Psi_N, s_1, s_2) \geq \frac{\kappa^2}{2} \right\}.$$

Lemma 3. *On the event $\mathcal{A}_N$, the following inequalities are valid*

$$\|\Pi(\hat{\gamma}-\gamma)\|^2+\lambda_N\|\hat{\alpha}-\alpha\|_1+\frac{\lambda_N}{\sqrt{N}}\|\hat{\eta}-\eta\|_1 \leq 4\lambda_N\|\hat{\alpha}_{J_1}-\alpha_{J_1}\|_1+4\frac{\lambda_N}{\sqrt{N}}\|\hat{\eta}_{J_2}-\eta_{J_2}\|_1+4\frac{\lambda_N}{\sqrt{N}}c_N; \quad (7.1)$$

$$\|\hat{\alpha}_{J_1^c}-\alpha_{J_1^c}\|_1+\frac{1}{\sqrt{N}}\|\hat{\eta}_{J_2^c}-\eta_{J_2^c}\|_1 \leq 3\|\hat{\alpha}_{J_1}-\alpha_{J_1}\|_1+3\frac{1}{\sqrt{N}}\|\hat{\eta}_{J_2}-\eta_{J_2}\|_1+4\frac{1}{\sqrt{N}}c_N. \quad (7.2)$$

Proof. By the minimizing property of the Lasso,

$$\|y-\Pi\hat{\gamma}\|^2+2\lambda_N\|\hat{\alpha}\|_1+2\frac{\lambda_N}{\sqrt{N}}\|\hat{\eta}\|_1 \leq \|y-\Pi\gamma\|^2+2\lambda_N\|\alpha\|_1+2\frac{\lambda_N}{\sqrt{N}}\|\eta\|_1$$

such that inserting $y=\Pi\gamma+\epsilon$ yields

$$\|\Pi(\hat{\gamma}-\gamma)\|^2 \leq 2\varepsilon'\Pi(\hat{\gamma}-\gamma)+2\lambda_N(\|\alpha\|_1-\|\hat{\alpha}\|_1)+2\frac{\lambda_N}{\sqrt{N}}(\|\eta\|_1-\|\hat{\eta}\|_1). \quad (7.3)$$

Note that on $\mathcal{A}_N$

$$2\varepsilon'\Pi(\hat{\gamma}-\gamma) \leq 2\|\varepsilon'Z\|_\infty\|\hat{\alpha}-\alpha\|_1+2\|\varepsilon'D\|_\infty\|\hat{\eta}-\eta\|_1 \leq \lambda_N\|\hat{\alpha}-\alpha\|_1+\frac{\lambda_N}{\sqrt{N}}\|\hat{\eta}-\eta\|_1.$$

Using this and adding $\lambda_N\|\hat{\alpha}-\alpha\|_1+\frac{\lambda_N}{\sqrt{N}}\|\hat{\eta}-\eta\|_1$ to both sides of (7.3) gives

$$\begin{aligned}
&\|\Pi(\hat{\gamma}-\gamma)\|^2+\lambda_N\|\hat{\alpha}-\alpha\|_1+\frac{\lambda_N}{\sqrt{N}}\|\hat{\eta}-\eta\|_1\\
&\leq 2\lambda_N(\|\alpha\|_1-\|\hat{\alpha}\|_1+\|\hat{\alpha}-\alpha\|_1)+2\frac{\lambda_N}{\sqrt{N}}(\|\eta\|_1-\|\hat{\eta}\|_1+\|\hat{\eta}-\eta\|_1)\\
&\leq 2\lambda_N(\|\alpha_{J_1}\|_1-\|\hat{\alpha}_{J_1}\|_1+\|\hat{\alpha}_{J_1}-\alpha_{J_1}\|_1)+2\frac{\lambda_N}{\sqrt{N}}(\|\eta_{J_2}\|_1-\|\hat{\eta}_{J_2}\|_1+\|\hat{\eta}_{J_2}-\eta_{J_2}\|_1+2c_N)\\
&\leq 4\lambda_N\|\hat{\alpha}_{J_1}-\alpha_{J_1}\|_1+4\frac{\lambda_N}{\sqrt{N}}\|\hat{\eta}_{J_2}-\eta_{J_2}\|_1+4\frac{\lambda_N}{\sqrt{N}}c_N,
\end{aligned}$$

where the second inequality used $\|\eta_{J_2^c}\|_1-\|\hat{\eta}_{J_2^c}\|_1+\|\hat{\eta}_{J_2^c}-\eta_{J_2^c}\|_1 \leq 2\|\eta_{J_2^c}\|_1 \leq 2c_N$. This coincides with (7.1). (7.2) follows trivially from this. □

Lemma 4 (Deterministic oracle inequalities). *Let Assumption 2 hold. On the event $\mathcal{A}_N \cap \mathcal{B}_N$ one has for any positive constant λ_N,*

$$\begin{aligned}
\left\|\Pi(\hat{\gamma}-\gamma)\right\|^2 &\leq \frac{120\lambda_N^2(s_1+s_2)}{\kappa^2 NT}+\frac{20\lambda_N}{\sqrt{N}}c_N\\
\|\hat{\alpha}-\alpha\|_1 &\leq \frac{120\lambda_N(s_1+s_2)}{\kappa^2 NT}+\frac{20}{\sqrt{N}}c_N\\
\|\hat{\eta}-\eta\|_1 &\leq \frac{120\lambda_N(s_1+s_2)}{\kappa^2\sqrt{N}T}+20c_N.
\end{aligned}$$

Moreover, the above bounds are valid uniformly over $\mathcal{F}(s_1,s_2,c_N)=\left\{\alpha\in\mathbb{R}^p:\|\alpha\|_0\leq s_1\right\}\times\left\{\eta\in\mathbb{R}^N:\exists\ I\subseteq\{1,...,N\}\ \text{with}\ |I|\leq s_2\ \text{and}\ \|\eta_{I^c}\|_1\leq c_N\right\}$.

Proof. By (7.1) of Lemma 3, which is valid on $\mathcal{A}_N$,

$$\|\Pi(\hat{\gamma}-\gamma)\|^2 \leq 4\lambda_N\|\hat{\alpha}_{J_1}-\alpha_{J_1}\|_1+4\frac{\lambda_N}{\sqrt{N}}\|\hat{\eta}_{J_2}-\eta_{J_2}\|_1+4\frac{\lambda_N}{\sqrt{N}}c_N. \quad (7.4)$$

Consider the auxiliary event

$$\mathcal{C}_N := \left\{ \frac{1}{\sqrt{N}} c_N \leq \frac{1}{4} \|\hat{\alpha}_{J_1} - \alpha_{J_1}\|_1 + \frac{1}{4\sqrt{N}} \|\hat{\eta}_{J_2} - \eta_{J_2}\|_1 \right\}.$$

On the event $\mathcal{A}_N \cap \mathcal{C}_N$, from (7.2) of Lemma 3, we have

$$\|\hat{\alpha}_{J_1^c} - \alpha_{J_1^c}\|_1 + \frac{1}{\sqrt{N}} \|\hat{\eta}_{J_2^c} - \eta_{J_2^c}\|_1 \leq 4\|\hat{\alpha}_{J_1} - \alpha_{J_1}\|_1 + 4\frac{1}{\sqrt{N}} \|\hat{\eta}_{J_2} - \eta_{J_2}\|_1. \tag{7.5}$$

In order to apply the compatibility condition, re-parametrise the vector δ in the definition of the compatibility condition as follows. Let b^1 and b^2 be $p \times 1$ and $N \times 1$ vectors, respectively, with $b = (b^{1'}, b^{2'})'$ defined as

$$\begin{pmatrix} b^1 \\ b^2 \end{pmatrix} := \begin{pmatrix} \mathrm{I}_p & 0 \\ 0 & \sqrt{N}\mathrm{I}_N \end{pmatrix} \begin{pmatrix} \delta^1 \\ \delta^2 \end{pmatrix}.$$

Hence, that $\kappa^2(\Psi_N, r_1, r_2)$ is bounded away from zero for integers $r_1 \in \{1, \ldots, p\}$ and $r_2 \in \{1, \ldots, N\}$ is equivalent to

$$\kappa^2(\Psi_N, r_1, r_2) := \min_{\substack{R_1 \subseteq \{1,\ldots,p\}, |R_1| \leq r_1 \\ R_2 \subseteq \{1,\ldots,N\}, |R_2| \leq r_2 \\ R := R_1 \cup R_2}} \min_{\substack{b \neq 0, \\ \|b^1_{R_1^c}\|_1 + \frac{1}{\sqrt{N}}\|b^2_{R_2^c}\|_1 \\ \leq 4\|b^1_{R_1}\|_1 + \frac{4}{\sqrt{N}}\|b^2_{R_2}\|_1}} \frac{\|\Pi b\|^2}{\frac{NT}{r_1 + r_2} \left\| \begin{pmatrix} b^1_{R_1} \\ b^2_{R_2}/\sqrt{N} \end{pmatrix} \right\|_1^2} > 0. \tag{7.6}$$

By (7.5), our estimator satisfies the constraint of the just introduced version of the compatibility condition and so

$$\begin{aligned} \|\Pi(\hat{\gamma} - \gamma)\|^2 &\geq \frac{\kappa^2(\Psi_N, s_1, s_2)NT}{s_1 + s_2} \left\| \begin{pmatrix} \hat{\alpha}_{J_1} - \alpha_{J_1} \\ (\hat{\eta}_{J_2} - \eta_{J_2})/\sqrt{N} \end{pmatrix} \right\|_1^2 \\ &\geq \frac{\kappa^2(\Psi_N, s_1, s_2)NT}{s_1 + s_2} \left(\|\hat{\alpha}_{J_1} - \alpha_{J_1}\|_1^2 + \frac{1}{N} \|\hat{\eta}_{J_2} - \eta_{J_2}\|_1^2 \right) \\ &\geq \frac{\kappa^2 NT}{2(s_1 + s_2)} \left(\|\hat{\alpha}_{J_1} - \alpha_{J_1}\|_1^2 + \frac{1}{N} \|\hat{\eta}_{J_2} - \eta_{J_2}\|_1^2 \right), \end{aligned}$$

where the last inequality is valid on $\mathcal{B}_N$. Hence, on $\mathcal{A}_N \cap \mathcal{B}_N \cap \mathcal{C}_N$ upon combining with (7.4) one has,

$$\begin{aligned} \frac{\kappa^2 NT}{2(s_1 + s_2)} \left(\|\hat{\alpha}_{J_1} - \alpha_{J_1}\|_1^2 + \frac{1}{N} \|\hat{\eta}_{J_2} - \eta_{J_2}\|_1^2 \right) &\leq 4\lambda_N \|\hat{\alpha}_{J_1} - \alpha_{J_1}\|_1 + \frac{4\lambda_N}{\sqrt{N}} \|\hat{\eta}_{J_2} - \eta_{J_2}\|_1 + \frac{4\lambda_N}{\sqrt{N}} c_N \\ &\leq 5\lambda_N \|\hat{\alpha}_{J_1} - \alpha_{J_1}\|_1 + \frac{5\lambda_N}{\sqrt{N}} \|\hat{\eta}_{J_2} - \eta_{J_2}\|_1, \end{aligned}$$

which, since $\kappa^2 > 0$ by Assumption 2, is equivalent to

$$\|\hat{\alpha}_{J_1} - \alpha_{J_1}\|_1^2 - \frac{10\lambda_N(s_1 + s_2)}{\kappa^2 NT} \|\hat{\alpha}_{J_1} - \alpha_{J_1}\|_1 + \frac{1}{N} \|\hat{\eta}_{J_2} - \eta_{J_2}\|_1^2 - \frac{10\lambda_N(s_1 + s_2)}{\kappa^2 N^{3/2} T} \|\hat{\eta}_{J_2}\|_1 \leq 0.$$

Let $x = \|\hat{\alpha}_{J_1} - \alpha_{J_1}\|_1$, $y = \|\hat{\eta}_{J_2} - \eta_{J_2}\|_1$, $a = \frac{10\lambda_N(s_1+s_2)}{\kappa^2 NT}$, $b = \frac{1}{N}$ and $c = \frac{10\lambda_N(s_1+s_2)}{\kappa^2 N^{3/2} T}$. Thus one has

$$x^2 - ax + by^2 - cy \leq 0. \tag{7.7}$$

First bound $x = \|\hat{\alpha}_{J_1} - \alpha_{J_1}\|_1$. For every y the values of x that satisfy the above quadratic inequality form an interval in $\mathbb{R}_+$. The right end point of this interval is the desired upper bound on x. Clearly, by the solution formula for the roots of a second degree polynomial, this right end point is a decreasing function in $by^2 - cy$. Hence, we first minimize the polynomial

$by^2 - cy$ to find the largest possible value of x which satisfies (7.7). This yields $y = c/2b$ and the corresponding value of $by^2 - cy$ is $-c^2/(4b)$. Hence, our desired upper bound on x is the largest solution of $x^2 - ax - \frac{c^2}{4b} \leq 0$. By the standard solution formula for the roots of a quadratic polynomial this yields

$$\|\hat{\alpha}_{J_1} - \alpha_{J_1}\|_1 = x \leq \frac{a + \sqrt{a^2 + c^2/b}}{2} \leq a + \frac{c}{2\sqrt{b}}. \tag{7.8}$$

Switching the roles of x and y, one gets a similar bound on $y = \|\hat{\eta}_{J_2} - \eta_{J_2}\|_1$, namely

$$\|\hat{\eta}_{J_2} - \eta_{J_2}\|_1 = y \leq \frac{c + \sqrt{c^2 + ba^2}}{2b} \leq \frac{c}{b} + \frac{a}{2\sqrt{b}}. \tag{7.9}$$

Inserting the definitions of a, b and c into (7.8) and (7.9), we get

$$\|\hat{\alpha}_{J_1} - \alpha_{J_1}\|_1 \leq \frac{15\lambda_N(s_1 + s_2)}{\kappa^2 NT} \tag{7.10}$$

$$\|\hat{\eta}_{J_2} - \eta_{J_2}\|_1 \leq \frac{15\lambda_N(s_1 + s_2)}{\kappa^2 N^{1/2} T}. \tag{7.11}$$

Therefore, on $\mathcal{A}_N \cap \mathcal{B}_N \cap \mathcal{C}_N$, it follows from (7.1) that

$$\left\|\Pi(\hat{\gamma} - \gamma)\right\|^2 \leq 4\lambda_N \|\hat{\alpha}_{J_1} - \alpha_{J_1}\|_1 + \frac{4\lambda_N}{\sqrt{N}}\|\hat{\eta}_{J_2} - \eta_{J_2}\|_1 + \frac{4\lambda_N}{\sqrt{N}}c_N \leq \frac{120\lambda_N^2(s_1 + s_2)}{\kappa^2 NT} + \frac{4\lambda_N}{\sqrt{N}}c_N$$

$$\|\hat{\alpha} - \alpha\|_1 \leq 4\|\hat{\alpha}_{J_1} - \alpha_{J_1}\|_1 + \frac{4}{\sqrt{N}}\|\hat{\eta}_{J_2} - \eta_{J_2}\|_1 + \frac{4}{\sqrt{N}}c_N \leq \frac{120\lambda_N(s_1 + s_2)}{\kappa^2 NT} + \frac{4}{\sqrt{N}}c_N$$

$$\|\hat{\eta} - \eta\|_1 \leq 4\sqrt{N}\|\hat{\alpha}_{J_1} - \alpha_{J_1}\|_1 + 4\|\hat{\eta}_{J_2} - \eta_{J_2}\|_1 + 4c_N \leq \frac{120\lambda_N(s_1 + s_2)}{\kappa^2\sqrt{N}T} + 4c_N.$$

On $\mathcal{A}_N \cap \mathcal{C}_N^c$ one has trivial oracle inequalities via (7.1) of Lemma 3. To be precise,

$$\|\Pi(\hat{\gamma} - \gamma)\|^2 < 20\lambda_N \frac{c_N}{\sqrt{N}}, \quad \|\hat{\alpha} - \alpha\|_1 < 20\frac{c_N}{\sqrt{N}}, \quad \|\hat{\eta} - \eta\|_1 < 20c_N.$$

These inequalities are valid on event $\mathcal{A}_N \cap \mathcal{B}_N \cap \mathcal{C}_N^c$ too. Thus the results of the lemma follow upon synchronising constants and using that $(\mathcal{A}_N \cap \mathcal{B}_N \cap \mathcal{C}_N^c) \cup (\mathcal{A}_N \cap \mathcal{B}_N \cap \mathcal{C}_N) = \mathcal{A}_N \cap \mathcal{B}_N$.

To see the uniformity $\mathcal{F}(s_1, s_2, c_N)$, note that only properties s_1, s_2 and c_N characterizing α and η enter the deterministic oracle inequalities. Hence, the deterministic oracle inequalities are uniform over the set $\mathcal{F}(s_1, s_2, c_N)$. □

For the proof of Lemma 5 below, we shall use Orlicz norms as defined in van der Vaart and Wellner (1996): Let ψ be a non-decreasing, convex function with $\psi(0) = 0$. Then, the Orlicz norm of a random variable X is given by

$$\|X\|_\psi = \inf\left\{C > 0 : \mathbb{E}\psi\left(|X|/C\right) \leq 1\right\},$$

where, as usual, $\inf \emptyset = \infty$. We will use Orlicz norms for $\psi(x) = \psi_b(x) = e^{x^b} - 1$ for various values of b. The following Lemma provides a lower bound on the probability of $\mathcal{A}_N$.

Lemma 5. *Let $\lambda_N = \sqrt{4MNT(\log(p \vee N))^3}$ for some $M > 0$. By Assumptions 1 and 3, we have*

$$\mathbb{P}(\mathcal{A}_N) \geq 1 - Ap^{1-BM^{1/3}} - AN^{1-BM^{1/3}},$$

for positive constants A and B.

Proof. Consider the event $\{\|Z'\varepsilon\|_\infty > \lambda_N/2\}$ first. To this end, let $z_{j,l}$ denote the jth entry of the lth column of Z, i.e. the jth entry of $(z_{1,1,l}, z_{1,2,l}, \ldots, z_{1,T,l}, z_{2,1,l}, \ldots, z_{N,T,l})'$. Similarly, we write ε_j for the jth entry of ε. Now note that $j \mapsto (\lceil \frac{j}{T} \rceil, j - \lfloor \frac{j}{T} \rfloor T)$ is a bijection from $\{1, ..., NT\}$ to $\{1, ..., N\} \times \{1, ..., T\}$ where $\lfloor x \rfloor$ denotes the greatest integer strictly less than x and $\lceil x \rceil$ the smallest integer greater than or equal to $x \in \mathbb{R}$. In case the lth column of Z corresponds one of the lags of the left hand side variable, assume for concreteness the kth lag, define $\mathcal{F}_n = \sigma\left(y_{\lceil \frac{j}{T} \rceil, j - \lfloor \frac{j}{T} \rfloor T}, ..., y_{\lceil \frac{j}{T} \rceil, j - \lfloor \frac{j}{T} \rfloor T - L}, \varepsilon_{\lceil \frac{j}{T} \rceil, j - \lfloor \frac{j}{T} \rfloor T}, 1 \leq j \leq n\right)$ and $S_{n,l} = \sum_{j=1}^n z_{j,l}\varepsilon_j = \sum_{j=1}^n y_{\lceil \frac{j}{T} \rceil, j - \lfloor \frac{j}{T} \rfloor T - k} \varepsilon_{\lceil \frac{j}{T} \rceil, j - \lfloor \frac{j}{T} \rfloor T}$. Thus,

$$
\begin{aligned}
\mathbb{E}[S_{n,l}|\mathcal{F}_{n-1}] &= \sum_{j=1}^{n-1} y_{\lceil \frac{j}{T} \rceil, j - \lfloor \frac{j}{T} \rfloor T - k} \varepsilon_{\lceil \frac{j}{T} \rceil, j - \lfloor \frac{j}{T} \rfloor T} + \mathbb{E}\left[y_{\lceil \frac{n}{T} \rceil, j - \lfloor \frac{n}{T} \rfloor T - k} \varepsilon_{\lceil \frac{n}{T} \rceil, n - \lfloor \frac{n}{T} \rfloor T} | \mathcal{F}_{n-1}\right] \\
&= S_{n-1,l} + y_{\lceil \frac{n}{T} \rceil, j - \lfloor \frac{n}{T} \rfloor T - k} \mathbb{E}\left[\varepsilon_{\lceil \frac{n}{T} \rceil, n - \lfloor \frac{n}{T} \rfloor T} | \mathcal{F}_{n-1}\right].
\end{aligned}
$$

Using that $\left(\lceil \frac{n}{T} \rceil, n - \lfloor \frac{n}{T} \rfloor T\right)$ is a unique pair $(i, t) \in \{1, \ldots, N\} \times \{1, \ldots, T\}$ we have that

$$
\mathbb{E}\left[\varepsilon_{\lceil \frac{n}{T} \rceil, n - \lfloor \frac{n}{T} \rfloor T} | \mathcal{F}_{n-1}\right] = \mathbb{E}[\varepsilon_{i,t}|\mathcal{F}_{n-1}] = \mathbb{E}[\varepsilon_{i,t}|\sigma(y_{i,s}, \ldots, y_{i,1-L}, \varepsilon_{i,s}, \ldots, \varepsilon_{i,1}, 1 \leq s \leq t-1)]
$$

[5]where the last equality follows from the assumption of independence across $1 \leq i \leq N$ (Assumption 1). By Assumption 1, this conditional expectation equals zero as the $\varepsilon_{i,s}$ are linear functions of $y_{i,s}, \ldots, y_{i,s-L}$ and $x_{i,s}$. Thus, $S_{n,l}$ is a martingale with mean zero (the increments are martingale differences by the above argument). A similar argument applies when the lth column of Z equals $\{x_{1,1,k}, x_{1,T,k}, x_{2,1,k}, ..., x_{N,T,k}\}'$ for some $1 \leq k \leq p_x$ such that every row of $Z'\varepsilon$ is a zero mean martingale.

Next, note that by Assumption 3, for all $1 \leq j \leq NT$, $1 \leq l \leq p$ and $\epsilon > 0$, one has

$$
\mathbb{P}(|z_{j,l}\varepsilon_j| \geq \epsilon) \leq \mathbb{P}(|z_{j,l}| \geq \sqrt{\epsilon}) + \mathbb{P}(|\varepsilon_j| \geq \sqrt{\epsilon}) \leq Ke^{-C\epsilon}.
$$

It follows from Lemma 2.2.1 in van der Vaart and Wellner (1996) that $\|z_{j,l}\varepsilon_j\|_{\psi_1} \leq (1+K)/C$. Then, by the definition of the Orlicz norm, $\mathbb{E}\left[e^{C/(1+K)|z_{j,l}\varepsilon_j|}\right] \leq 2$. Now use Proposition 2 in Appendix B with $D = C/(1+K)$, $\alpha = 1/3$ and $C_1 = 2$ to conclude

$$
\mathbb{P}\left(\|Z'\varepsilon\|_\infty > \frac{\lambda_N}{2}\right) \leq \sum_{l=1}^p \mathbb{P}\left(\left|\sum_{j=1}^{NT} z_{j,l}\varepsilon_j\right| > \frac{\lambda_N}{2NT}NT\right) = pAe^{-B\log(p \vee N)M^{1/3}} \leq Ap^{1-BM^{1/3}}.
$$

Note also that the upper bound of the preceding probability becomes arbitrarily small for sufficiently large N and M such that we also conclude

$$
\|Z'\varepsilon\|_\infty = O_p(\lambda_N).^6 \tag{7.12}
$$

Next, consider the event $\{\|D'\varepsilon\|_\infty > \lambda_N/(2\sqrt{N})\}$. Using Assumption 1 a small calculation shows that all entries of $D'\varepsilon$ are zero mean martingales with respect to the natural filtration. As above, Assumption 3 and Lemma 2.2.1 in van der Vaart and Wellner (1996) yield $\|\varepsilon_{i,t}\|_{\psi_2} \leq \left(\frac{1+K/2}{C}\right)^{1/2}$ such that by the second to last inequality on page 95 in van der Vaart and Wellner (1996) one has $\|\varepsilon_{i,t}\|_{\psi_1} \leq \|\varepsilon_{i,t}\|_{\psi_2}(\log 2)^{-1/2} \leq \left(\frac{1+K/2}{C}\right)^{1/2}(\log 2)^{-1/2}$ for all i and t. Then using the definition of the Orlicz norm, $\mathbb{E}\left[\exp\left(\left(\frac{C}{1+K/2}\right)^{1/2}(\log 2)^{1/2}|\varepsilon_{i,t}|\right)\right] \leq 2$ and Proposition 2 in Appendix B with $D = \left(\frac{C}{1+K/2}\right)^{1/2}(\log 2)^{1/2}$, $\alpha = 1/3$ and $C_1 = 2$ implies

$$
\mathbb{P}\left(\|D'\varepsilon\|_\infty > \frac{\lambda_N}{2\sqrt{N}}\right) \leq \sum_{i=1}^N \mathbb{P}\left(\left|\sum_{t=1}^T \varepsilon_{i,t}\right| > \frac{\lambda_N}{2\sqrt{N}T}T\right) \leq ANe^{-B(\log(p \vee N)M^{1/3}} \leq AN^{1-BM^{1/3}}.
$$

[5]For $t = 1$, the last expression in the above display is to be read as absence of conditioning on the error terms.
[6]When we use Landau notation, we mean N tending to infinity.

Note also that the upper bound of the preceding probability becomes arbitrarily small for sufficiently large N and M, such that we may also conclude

$$\|D'\varepsilon\|_\infty = O_p\left(\frac{\lambda_N}{\sqrt{N}}\right). \tag{7.13}$$

□

The following lemma shows that $\kappa^2(\Psi_N, s_1, s_2)$ and $\kappa^2(\Psi, s_1, s_2)$ are close if Ψ_N and Ψ are in some sense close.

Lemma 6. *Let A and B be two positive semidefinite $(p+N)\times(p+N)$ matrices and $\delta := \max_{1\le i,j\le p+N} |A_{ij} - B_{ij}|$. For integers $s_1 \in \{1,\dots,p\}$ and $s_2 \in \{1,\dots,N\}$, one has*

$$\kappa^2(B, s_1, s_2) \geq \kappa^2(A, s_1, s_2) - \delta 25(s_1 + s_2).$$

Proof. Let x be a $(p+N)\times 1$ non-zero vector, satisfying $\|x_{J^c}\|_1 \le 4\|x_J\|_1$ for $J = J_1 \cup (J_2 + p)$ where $J_1 \subseteq \{1, ..., p\}$ with $|J_1| \le s_1$, and $J_2 \subseteq \{1, ..., N\}$ with $|J_2| \le s_2$. Now,

$$\begin{aligned}|x'Ax - x'Bx| &= |x'(A-B)x| \le \|x\|_1 \|(A-B)x\|_\infty \le \|x\|_1^2 \delta = \delta\left(\|x_J\|_1 + \|x_{J^c}\|_1\right)^2\\ &\le \delta\left(\|x_J\|_1 + 4\|x_J\|_1\right)^2 \le \delta 25 \|x_J\|_1^2.\end{aligned}$$

Hence,

$$\frac{x'Bx}{\frac{1}{s_1+s_2}\|x_J\|_1^2} \geq \frac{x'Ax}{\frac{1}{s_1+s_2}\|x_J\|_1^2} - \delta 25(s_1+s_2) \geq \kappa^2(A, s_1, s_2) - \delta 25(s_1+s_2),$$

where the last inequality is true because of the definition of $\kappa^2(A, s_1, s_2)$. Minimising the left-hand side over non-zero x satisfying $\|x_{J^c}\|_1 \le 4\|x_J\|_1$ yields the claim. □

Define

$$\tilde{\mathcal{B}}_N = \left\{\max_{1\le i,j\le p+N} \left|\Psi_{N,ij} - \Psi_{ij}\right| \le \frac{\kappa^2(\Psi, s_1, s_2)}{50(s_1+s_2)}\right\}.$$

Setting $A = \Psi$, $B = \Psi_N$ it follows from Lemma 6 that $\tilde{\mathcal{B}}_N \subseteq \mathcal{B}_N$. Thus, we just need to find a lower bound on $\mathbb{P}(\tilde{\mathcal{B}}_N)$ in order to prove Theorem 1.

Lemma 7. *Let Assumptions 1, 2 and 3 hold. Assume that $s_1 + s_2 \le \sqrt{N}$. Then there exist positive constants A, B such that*

$$\mathbb{P}(\mathcal{B}_N^c) \le \mathbb{P}(\tilde{\mathcal{B}}_N^c) \le A(p^2 + pN)e^{-B\left(\frac{N}{(s_1+s_2)^2}\right)^{1/3}}.$$

Proof. Since the lower right $N \times N$ blocks of Ψ_N and Ψ are identical, it suffices to bound the entries of $\frac{1}{NT}Z'Z - \frac{1}{NT}\mathbb{E}[Z'Z]$ and $\frac{1}{T\sqrt{N}}Z'D$. A typical element of $\frac{1}{NT}Z'Z - \frac{1}{NT}\mathbb{E}[Z'Z]$ is of the form $\frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T(z_{i,t,l}z_{i,t,k} - \mathbb{E}[z_{i,t,l}z_{i,t,k}])$ for some $l, k \in \{1,\dots,p\}$. By Assumption 3 we have for every $\epsilon > 0$

$$\mathbb{P}(|z_{i,t,l}z_{i,t,k}| \ge \epsilon) \le \mathbb{P}(|z_{i,t,l}| \ge \sqrt{\epsilon}) + \mathbb{P}(|z_{i,t,k}| \ge \sqrt{\epsilon}) \le Ke^{-C\epsilon}.$$

It follows from Lemma 2.2.1 in van der Vaart and Wellner (1996) that $\|z_{i,t,l}z_{i,t,k}\|_{\psi_1} \le (1+K)/C$. Hence, by subadditivity of the Orlicz norm and Jensen's inequality

$$\left\|\frac{1}{T}\sum_{t=1}^T\left(z_{i,t,l}z_{i,t,k} - \mathbb{E}[z_{i,t,l}z_{i,t,k}]\right)\right\|_{\psi_1} \le 2\max_{1\le t\le T}\|z_{i,t,l}z_{i,t,k}\|_{\psi_1} \le \frac{2(1+K)}{C}.$$

Thus, by the definition of the Orlicz norm, $\mathbb{E}\exp\left(\frac{C}{2(1+K)}\left|\frac{1}{T}\sum_{t=1}^{T}(z_{i,t,l}z_{i,t,k}-\mathbb{E}[z_{i,t,l}z_{i,t,k}])\right|\right)\leq 2$. Using independence across i (Assumption 1) to invoke Proposition 2 in Appendix B with $D=\frac{C}{2(1+K)}$, $\alpha=1/3$ and $C_1=2$ such that for every $x\gtrsim\frac{1}{\sqrt{N}}$

$$\mathbb{P}\left(\left|\sum_{i=1}^{N}\frac{1}{T}\sum_{t=1}^{T}(z_{i,t,l}z_{i,t,k}-\mathbb{E}[z_{i,t,l}z_{i,t,k}])\right|\geq Nx\right)\leq Ae^{-B(x^2N)^{1/3}}, \tag{7.14}$$

for positive constants A and B.

Next, consider $\frac{1}{T\sqrt{N}}Z'D$. A typical element can be written as $\frac{1}{\sqrt{NT}}\sum_{t=1}^{T}z_{i,t,l}$ for some $i\in\{1,\ldots,N\}$ and $l\in\{1,\ldots,p\}$. By Assumption 3, we have $\mathbb{P}(|z_{i,t,l}|\geq\epsilon)\leq\frac{1}{2}Ke^{-C\epsilon^2}$ for all $\epsilon>0$ and it follows from Lemma 2.2.1 in van der Vaart and Wellner (1996) that $\|z_{i,t,l}\|_{\psi_2}\leq\left(\frac{1+K/2}{C}\right)^{1/2}$. Hence,

$$\left\|\frac{1}{\sqrt{NT}}\sum_{t=1}^{T}z_{i,t,l}\right\|_{\psi_2}\leq\frac{1}{\sqrt{N}}\max_{1\leq t\leq T}\|z_{i,t,l}\|_{\psi_2}\leq\frac{1}{\sqrt{N}}\left(\frac{1+K/2}{C}\right)^{1/2}=:\frac{C'}{\sqrt{N}}.$$

Thus, it follows by Markov's inequality, positivity and increasingness of $\psi_2(x)$, as well as $1\wedge\psi_2(x)^{-1}=1\wedge(e^{x^2}-1)^{-1}\leq 2e^{-x^2}$ that for any $x>0$

$$\mathbb{P}\left(\left|\frac{1}{\sqrt{NT}}\sum_{t=1}^{T}z_{i,t,l}\right|>x\right)\leq 1\wedge\frac{1}{e^{(x\sqrt{N}/C')^2}-1}\leq 2e^{-\frac{Nx^2}{C'^2}}\leq Ae^{-Bx^2N}, \tag{7.15}$$

where the last estimate follows by choosing A and B sufficiently large/small for (7.14) and (7.15) both to be valid. Setting $x=\frac{\kappa^2}{50(s_1+s_2)}=\frac{\kappa^2}{50}\frac{1}{s_1+s_2}$, using that $\frac{1}{s_1+s_2}\geq 1/\sqrt{N}$ and κ^2 being bounded away from 0 (Assumption 2), we have

$$\begin{aligned}\mathbb{P}(\mathcal{B}_N^c)\leq\mathbb{P}(\tilde{\mathcal{B}}_N^c)&=\mathbb{P}\left(\max_{1\leq i,j\leq p+N}|\Psi_{N,ij}-\Psi_{ij}|>x\right)\\&\leq A(p^2+pN)\left(e^{-B\left[\left(\frac{\kappa^2}{50(s_1+s_2)}\right)^2N\right]^{1/3}}\vee e^{-B\left(\frac{\kappa^2}{50(s_1+s_2)}\right)^2N}\right)\leq A(p^2+pN)e^{-B\left(\frac{N}{(s_1+s_2)^2}\right)^{1/3}}\end{aligned}$$

where the last estimate has merged $(\kappa^2/50)^{2/3}$ into B and used $\frac{N}{(s_1+s_2)^2}\geq 1$. □

Proof of Theorem 1. Theorem 1 follows by combining Lemmas 4, 5, and 7. □

Corollary 1. *Let conditions of Theorem 1 hold. For large enough $M>0$ and assuming $\frac{(\log(p\vee N))^3(s_1+s_2)^2}{N}=o(1)$, we have the following stochastic orders valid uniformly over $\mathcal{F}(s_1,s_2,c_N)$.*

$$\frac{1}{NT}\left\|\Pi(\hat{\gamma}-\gamma)\right\|^2=O_p\left((s_1+s_2)\left(\frac{\lambda_N}{NT}\right)^2\right)+O_p\left(\frac{\lambda_N}{NT}\frac{c_N}{\sqrt{N}}\right),$$

$$\|\hat{\alpha}-\alpha\|_1=O_p\left((s_1+s_2)\frac{\lambda_N}{NT}\right)+O_p\left(\frac{c_N}{\sqrt{N}}\right),$$

$$\|\hat{\eta}-\eta\|_1=O_p\left((s_1+s_2)\frac{\lambda_N}{\sqrt{NT}}\right)+O_p\left(c_N\right).$$

Proof of Corollary 1. Given positive constants A and B, $Ap^{1-BM^{1/3}}$ and $AN^{1-BM^{1/3}}$ become arbitrarily small for large enough $M>0$. By $\frac{(\log(p\vee N))^3(s_1+s_2)^2}{N}=o(1)$, $A(p^2+pN)e^{-B\left(\frac{N}{(s_1+s_2)^2}\right)^{1/3}}\to 0$ as $N\to\infty$. Thus the lower bound on the probability in Theorem 1 goes to one as $N\to\infty$ for large enough $M>0$ and the conclusion follows from Theorem 1. □

7.3 Proof of Lemma 1

The following lemma gives the rates of the uniform prediction and estimation errors for nodewise regression.

Lemma 8. *Let Assumptions 1, 3 and 4 hold. Let $\lambda_{node} = \sqrt{4M(\log p)^3/N}$ for some $M > 0$. For M sufficiently large, we have*

$$\max_{j\in H_1} \frac{1}{NT}\|Z_{-j}(\hat{\phi}_j - \phi_j)\|^2 = O_p\left(\bar{s}\lambda_{node}^2\right) \tag{7.16}$$

$$\max_{j\in H_1} \|\hat{\phi}_j - \phi_j\|_1 = O_p\left(\bar{s}\lambda_{node}\right) \tag{7.17}$$

$$\max_{j\in H_1} \frac{1}{NT}\|Z_{-j}'\zeta_j\|_\infty = O_p(\lambda_{node}). \tag{7.18}$$

Proof. We say that a $(p-1)\times(p-1)$ matrix A satisfies the *compatibility condition* $CC(r)$ for some integer $r \in \{1,\ldots,p-1\}$ if

$$\kappa^2\left(A,r\right) := \min_{\substack{R\subseteq\{1,\ldots,p-1\}\\ |R|\le r}} \min_{\substack{\delta\in\mathbb{R}^{p-1}\setminus\{0\}\\ \|\delta_{R^c}\|_1\le 3\|\delta_R\|_1}} \frac{\delta' A\delta}{\frac{1}{r}\|\delta_R\|_1^2} > 0$$

and consider the events

$$\mathcal{D}_N = \left\{\max_{j\in H_1} \frac{1}{NT}\|Z_{-j}'\zeta_j\|_\infty \le \frac{\lambda_{node}}{2}\right\}$$

and

$$\mathcal{E}_{N,j} = \left\{\kappa^2\left(\frac{1}{NT}Z_{-j}'Z_{-j}, s_{node,j}\right) \ge \frac{\kappa^2\left(\Psi_{Z,-j,-j}, s_{node,j}\right)}{2}\right\}.$$

By standard arguments or using the same technique as in Lemma 3 it can be shown that for each $j \in H_1$

$$\frac{1}{NT}\|Z_{-j}(\hat{\phi}_j - \phi_j)\|^2 \le \frac{32 s_{node,j}\lambda_{node}^2}{\kappa^2\left(\Psi_{Z,-j,-j}, s_{node,j}\right)}, \qquad \|\hat{\phi}_j - \phi_j\|_1 \le \frac{32 s_{node,j}\lambda_{node}}{\kappa^2\left(\Psi_{Z,-j,-j}, s_{node,j}\right)},$$

on $\mathcal{D}_N \cap \mathcal{E}_{N,j}$. Thus, the inequalities in the above display are valid simultaneously on $\mathcal{D}_N \cap (\cap_{j\in H_1}\mathcal{E}_{N,j})$. Noting that $\kappa^2\left(\Psi_{Z,-j,-j}, r\right) \ge \kappa^2\left(\Psi_Z, r\right)$ for all $j = 1,...,p$ and $r = 1,...,p-1$ we conclude that

$$\max_{j\in H_1} \frac{1}{NT}\|Z_{-j}(\hat{\phi}_j - \phi_j)\|^2 \le \frac{32\bar{s}\lambda_{node}^2}{\kappa^2\left(\Psi_Z,\bar{s}\right)} \qquad \text{and} \qquad \max_{j\in H_1}\|\hat{\phi}_j - \phi_j\|_1 \le \frac{32\bar{s}\lambda_{node}}{\kappa^2\left(\Psi_Z,\bar{s}\right)}.$$

on $\mathcal{D}_N \cap (\cap_{j\in H_1}\mathcal{E}_{N,j})$. Thus, we proceed by establishing a lower bound on this set. Consider $\mathcal{D}_N$ first. A typical element of $Z_{-j}'\zeta_j$ is of the form $\sum_{i=1}^N\sum_{t=1}^T z_{i,t,l}\zeta_{j,i,t}$ for some $l \ne j$. By (3.12), one has $\frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T z_{i,t,l}\zeta_{j,i,t} = \frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T(z_{i,t,l}\zeta_{j,i,t} - \mathbb{E}[z_{i,t,l}\zeta_{j,i,t}])$ for $l \ne j$. By Assumptions 3 and 4(c), it holds for any $\epsilon > 0$ that

$$\mathbb{P}(|z_{i,t,l}\zeta_{j,i,t}| > \epsilon) \le \mathbb{P}(|z_{i,t,l}| > \sqrt{\epsilon}) + \mathbb{P}(|\zeta_{j,i,t}| > \sqrt{\epsilon}) \le Ke^{-C\epsilon}.$$

such that Lemma 2.2.1 in van der Vaart and Wellner (1996) yields that $\|z_{i,t,l}\zeta_{j,i,t}\|_{\psi_1} \le (1+K)/C$. Therefore, by Jensen's inequality and subadditivity of the Orlicz norm

$$\left\|\frac{1}{T}\sum_{t=1}^T\left(z_{i,t,l}\zeta_{j,i,t} - \mathbb{E}[z_{i,t,l}\zeta_{j,i,t}]\right)\right\|_{\psi_1} \le 2\max_{1\le t\le T}\|z_{i,t,l}\zeta_{j,i,t}\|_{\psi_1} \le \frac{2(1+K)}{C}.$$

Using the definition of the Orlicz norm $\mathbb{E}\exp\left(\frac{C}{2(1+K)}\left|\frac{1}{T}\sum_{t=1}^{T}\left(z_{i,t,l}\zeta_{j,i,t}-\mathbb{E}[z_{i,t,l}\zeta_{j,i,t}]\right)\right|\right)\le 2$. Using independence across i (Assumption 1) to invoke Proposition 2 in Appendix B with $D = C/(1+K)$, $\alpha = 1/3$, $C_1 = 2$ and $\epsilon = \lambda_{node}/2 \gtrsim \frac{1}{\sqrt{N}}$, we conclude (using $h_1 \le p$)

$$\mathbb{P}\left(\max_{j\in H_1}\frac{1}{NT}\|Z_{-j}'\zeta_j\|_\infty > \epsilon\right) \le h_1 p\mathbb{P}\left(\left|\sum_{i=1}^{N}\frac{1}{T}\sum_{t=1}^{T}(z_{i,t,l}\zeta_{j,i,t}-\mathbb{E}[z_{i,t,l}\zeta_{j,i,t}])\right| > \epsilon N\right)$$
$$\le Ah_1pe^{-B(\epsilon^2N)^{1/3}} \le Ap^2e^{-BM^{1/3}\log p} = Ap^{2-BM^{1/3}}$$

for positive constants A and B. The upper bound of the preceding probability becomes arbitrarily small for M sufficiently large such that

$$\max_{j\in H_1}\frac{1}{NT}\|Z_{-j}'\zeta_j\|_\infty = O_p(\lambda_{node}),$$

which is (7.18).

In order to provide a lower bound on the probability of $\left(\cap_{j\in H_1}\mathcal{E}_{N,j}\right)$ define the event

$$\tilde{\mathcal{E}}_{N,j} := \left\{\max_{1\le l,k\le p-1}\left|\left[\frac{1}{NT}Z_{-j}'Z_{-j}\right]_{lk} - [\Psi_{Z,-j,-j}]_{lk}\right| \le \frac{\kappa^2(\Psi_{Z,-j,-j}, s_{node,j})}{32s_{node,j}}\right\} \subseteq \mathcal{E}_{N,j}$$

by Proposition 1 in Appendix B with $A = \Psi_{Z,-j,-j}$, $B = \frac{1}{NT}Z_{-j}'Z_{-j}$, $r = s_{node,j}$ and $\delta = \frac{\kappa^2(\Psi_{Z,-j,-j},s_{node,j})}{32s_{node,j}}$. Observe that the relation

$$\max_{1\le l,k\le p-1}\left|\left[\frac{1}{NT}Z_{-j}'Z_{-j}\right]_{lk} - \left[\Psi_{Z,-j,-j}\right]_{lk}\right| \le \max_{1\le l,k\le p}\left|\left[\frac{1}{NT}Z'Z\right]_{lk} - [\Psi_Z]_{lk}\right|$$
$$\le \frac{\kappa^2(\Psi_Z,\bar{s})}{32\bar{s}} \le \frac{\kappa^2(\Psi_{Z,-j,-j}, s_{node,j})}{32s_{node,j}},$$

implies $\mathcal{E}_N := \left\{\max_{1\le l,k\le p}\left|[\frac{1}{NT}Z'Z]_{lk} - [\Psi_Z]_{lk}\right| \le \frac{\kappa^2(\Psi_Z,\bar{s})}{32\bar{s}}\right\} \subseteq \tilde{\mathcal{E}}_{N,j} \subseteq \mathcal{E}_{N,j}$ for all $j \in H_1$ and hence $\mathcal{E}_N \subseteq \cap_{j\in H_1}\mathcal{E}_{N,j}$. It remains to provide a lower bound on $\mathbb{P}(\mathcal{E}_N)$. A typical element of $\frac{1}{NT}Z'Z - \Psi_Z$ is of the form $\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}(z_{i,t,l}z_{i,t,k} - \mathbb{E}[z_{i,t,l}z_{i,t,k}])$ for some $l,k \in \{1,\dots,p\}$. Invoking (7.14) with $x = \frac{\kappa^2(\Psi_Z,\bar{s})}{32\bar{s}} \gtrsim \frac{1}{\sqrt{N}}$ (using $\frac{(\log p)^3\bar{s}^2}{N} = o(1)$ Assumption 4(b))

$$\mathbb{P}\left(\left|\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}(z_{i,t,l}z_{i,t,k} - \mathbb{E}[z_{i,t,l}z_{i,t,k}])\right| \ge x\right) \le Ae^{-B(x^2N)^{1/3}},$$

for positive constants A and B. Therefore,

$$\mathbb{P}(\mathcal{E}_N^c) = \mathbb{P}\left(\max_{1\le l,k\le p}\left|\left[\frac{1}{NT}Z'Z\right]_{lk} - [\Psi_Z]_{lk}\right| \ge x\right) \le p^2Ae^{-B(x^2N)^{1/3}} \to 0$$

as $N \to \infty$ using $\frac{(\log p)^3\bar{s}^2}{N} = o(1)$ (Assumption 4(b)). □

Proof of Lemma 1. Recall (3.6) and use $z_j = Z_{-j}\phi_j + \zeta_j$:

$$\hat{\tau}_j^2 = \frac{1}{NT}\zeta_j'\zeta_j + \frac{1}{NT}\zeta_j'Z_{-j}\phi_j - \frac{1}{NT}(\hat{\phi}_j - \phi_j)'Z_{-j}'\zeta_j - \frac{1}{NT}(\hat{\phi}_j - \phi_j)'Z_{-j}'Z_{-j}\phi_j.$$

Thus,

$$\max_{j\in H_1}|\hat\tau_j^2-\tau_j^2| \le \max_{j\in H_1}\Big|\frac{1}{NT}\zeta_j'\zeta_j-\tau_j^2\Big| + \max_{j\in H_1}\Big|\frac{1}{NT}\zeta_j'Z_{-j}\phi_j\Big| + \max_{j\in H_1}\Big|\frac{1}{NT}(\hat\phi_j-\phi_j)'Z_{-j}'\zeta_j\Big| + \max_{j\in H_1}\Big|\frac{1}{NT}(\hat\phi_j-\phi_j)'Z_{-j}'Z_{-j}\phi_j\Big|. \tag{7.19}$$

Consider the first term on the right of the inequality in (7.19). By Assumption 4(c), we have for all $\epsilon>0$, $\mathbb{P}(|\zeta_{j,i,t}^2|\ge\epsilon)=\mathbb{P}(|\zeta_{j,i,t}|\ge\sqrt{\epsilon})\le\frac{1}{2}Ke^{-C\epsilon}$. It follows from Lemma 2.2.1 in van der Vaart and Wellner (1996) that $\|\zeta_{j,i,t}^2\|_{\psi_1}\le(1+K/2)/C$. Therefore, by Jensen's inequality and subadditivity of the Orlicz norm

$$\left\|\frac{1}{T}\sum_{t=1}^{T}\Big(\zeta_{j,i,t}^2-\mathbb{E}[\zeta_{j,i,t}^2]\Big)\right\|_{\psi_1}\le 2\max_{1\le t\le T}\|\zeta_{j,i,t}^2\|_{\psi_1}\le\frac{2+K}{C}.$$

Using the definition of the Orlicz norm, $\mathbb{E}\exp\big(\frac{C}{2+K}\big|\frac{1}{T}\sum_{t=1}^{T}(\zeta_{j,i,t}^2-\mathbb{E}[\zeta_{j,i,t}^2])\big|\big)\le 2$. Using independence across $i=1,\dots,N$ (Assumption 1) to invoke Proposition 2 in Appendix B with $D=C/(2+K)$, $\alpha=1/3$ and $C_1=2$ for $x\gtrsim\frac{1}{\sqrt{N}}$,

$$\mathbb{P}\Big(\Big|\frac{1}{N}\sum_{i=1}^{N}\frac{1}{T}\sum_{t=1}^{T}(\zeta_{j,i,t}^2-\mathbb{E}[\zeta_{j,i,t}^2])\Big|\ge x\Big)\le Ae^{-B(x^2N)^{1/3}},$$

for positive constants A and B. Setting $x=\sqrt{\frac{M(\log h_1)^3}{N}}$ for some $M>0$, we have

$$\begin{aligned}&\mathbb{P}\left(\max_{j\in H_1}\Big|\frac{1}{N}\sum_{i=1}^{N}\frac{1}{T}\sum_{t=1}^{T}(\zeta_{j,i,t}^2-\mathbb{E}[\zeta_{j,i,t}^2])\Big|\ge\sqrt{\frac{M(\log h_1)^3}{N}}\right)\\&\le\sum_{j\in H_1}\mathbb{P}\left(\Big|\frac{1}{N}\sum_{i=1}^{N}\frac{1}{T}\sum_{t=1}^{T}(\zeta_{j,i,t}^2-\mathbb{E}[\zeta_{j,i,t}^2])\Big|\ge\sqrt{\frac{M(\log h_1)^3}{N}}\right)\le Ah_1^{1-BM^{1/3}}.\end{aligned}$$

Recognising that the upper bound of the preceding probability becomes arbitrarily small for sufficiently large N and M, we have

$$\max_{j\in H_1}\Big|\frac{1}{NT}\zeta_j'\zeta_j-\tau_j^2\Big|=O_p\Big(\sqrt{\frac{(\log h_1)^3}{N}}\Big)=O_p(\lambda_{node}).$$

Now consider the second term on the right of the inequality in (7.19). Recall that

$$C=\begin{pmatrix}1&-\phi_{1,2}&\cdots&-\phi_{1,p}\\-\phi_{2,1}&1&\cdots&-\phi_{2,p}\\\vdots&\vdots&\ddots&\vdots\\-\phi_{p,1}&-\phi_{p,2}&\cdots&1\end{pmatrix}$$

such that C_j is the jth row of C but written as a $p\times1$ vector. Then

$$\begin{aligned}\max_{j\in H_1}\|\phi_j\|_1&\le\max_{j\in H_1}\sqrt{s_{node,j}}\|C_j\|\le\max_{j\in H_1}\sqrt{s_{node,j}}\sqrt{\frac{C_j'\Psi_ZC_j}{\text{mineval}(\Psi_Z)}}\\&=\max_{j\in H_1}\frac{\sqrt{s_{node,j}}}{\sqrt{\text{mineval}(\Psi_Z)}}\sqrt{\Psi_{Z,j,j}-\Psi_{Z,j,-j}\Psi_{Z,-j,-j}^{-1}\Psi_{Z,-j,j}}\\&\le\max_{j\in H_1}\sqrt{s_{node,j}}\frac{\max_{j\in H_1}\sqrt{\Psi_{Z,j,j}}}{\sqrt{\text{mineval}(\Psi_Z)}}\le\max_{j\in H_1}\sqrt{s_{node,j}}\frac{\sqrt{\text{maxeval}(\Psi_Z)}}{\sqrt{\text{mineval}(\Psi_Z)}}=O(\sqrt{\bar{s}}),\end{aligned}\tag{7.20}$$

where the first equality is due to (3.11), and the third inequality is due to that Assumption 4(a) implies that $\Psi^{-1}_{Z,-j,-j}$ is positive definite for all $j \in H_1$. Now,

$$\max_{j\in H_1}\left|\frac{1}{NT}\zeta_j' Z_{-j}\phi_j\right| \leq \max_{j\in H_1}\left(\left\|\frac{1}{NT}\zeta_j' Z_{-j}\right\|_\infty \|\phi_j\|_1\right) = O_p(\lambda_{node})O(\sqrt{\bar{s}}) = O_p(\sqrt{\bar{s}}\lambda_{node}),$$

where the first equality is due to (7.18).

The third term in (7.19) is bounded as

$$\max_{j\in H_1}\left|\frac{1}{NT}(\hat{\phi}_j-\phi_j)'Z_{-j}'\zeta_j\right| \leq \max_{j\in H_1}\left(\left\|\hat{\phi}_j-\phi_j\right\|_1\left\|\frac{1}{NT}Z_{-j}'\zeta_j\right\|_\infty\right) = O_p(\bar{s}\lambda^2_{node}),$$

where the equality is due to (7.17) and (7.18).

To bound the fourth term on the right of the inequality in (7.19), recall (3.8) and manipulate to get $\frac{1}{NT}Z_{-j}'Z_{-j}(\hat{\phi}_j-\phi_j) = \frac{1}{NT}Z_{-j}'\zeta_j - \lambda_{node}w_j$. Thus,

$$\left\|\frac{1}{NT}(\hat{\phi}_j-\phi_j)'Z_{-j}'Z_{-j}\right\|_\infty \leq \left\|\frac{1}{NT}Z_{-j}'\zeta_j\right\|_\infty + \lambda_{node}\|w_j\|_\infty = O_p(\lambda_{node}),$$

where the equality is due to (7.18). Thus,

$$\max_{j\in H_1}\left|\frac{1}{NT}(\hat{\phi}_j-\phi_j)'Z_{-j}'Z_{-j}\phi_j\right| \leq \max_{j\in H_1}\left\|\frac{1}{NT}(\hat{\phi}_j-\phi_j)'Z_{-j}'Z_{-j}\right\|_\infty \max_{j\in H_1}\|\phi_j\|_1 = O_p(\sqrt{\bar{s}}\lambda_{node}),$$

where the last equality is due to (7.20). Summing up all four terms on the right of the inequality in (7.19), we get

$$\max_{j\in H_1}|\hat{\tau}_j^2-\tau_j^2| \leq O_p\left(\lambda_{node}\right) + O_p(\sqrt{\bar{s}}\lambda_{node}) + O_p(\bar{s}\lambda^2_{node}) = O_p(\sqrt{\bar{s}}\lambda_{node}) = o_p(1),$$

where the first equality is due to that $O_p(\sqrt{\bar{s}}\lambda_{node})$ dominates $O_p(\bar{s}\lambda^2_{node})$ by Assumption 4(b) ($\bar{s}\lambda_{node} = o(1)$) implies $\bar{s}\lambda^2_{node} = o(1)$. The second equality is also due to Assumption 4(b). This establishes (3.13).

We now prove (3.14). We first recall

$$\tau_j^2 = \mathbb{E}\left[\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}(z_{i,t,j}-z_{i,t,-j}'\phi_j)^2\right] = \Psi_{Z,j,j} - \Psi_{Z,j,-j}\Psi^{-1}_{Z,-j,-j}\Psi_{Z,-j,j} \tag{7.21}$$
$$= \frac{1}{\Theta_{Z,j,j}},$$

Furthermore,

$$\Theta_{Z,j,j} \equiv \frac{e_j'\Theta_Z e_j}{\|e_j\|^2} \leq \max_{\delta\in\mathbb{R}^p\setminus\{0\}}\frac{\delta'\Theta_Z\delta}{\|\delta\|^2} = \text{maxeval}(\Theta_Z) = \frac{1}{\text{mineval}(\Psi_Z)}.$$

The preceding inequality is uniform in j. Thus, $\min_{j\in H_1}\tau_j^2 \geq \text{mineval}(\Psi_Z)$, which is uniformly bounded away from zero by Assumption 4(a). Therefore,

$$\min_{j\in H_1}\hat{\tau}_j^2 = \min_{j\in H_1}(\hat{\tau}_j^2-\tau_j^2+\tau_j^2) \geq \min_{j\in H_1}\tau_j^2 - \max_{j\in H_1}|\hat{\tau}_j^2-\tau_j^2| \geq \text{mineval}(\Psi_Z) - o_p(1).$$

Hence, we conclude that $\min_{j\in H_1}\hat{\tau}_j^2$ is bounded away from zero for N large enough and $\max_{j\in H_1}\frac{1}{\hat{\tau}_j^2} = O_p(1)$ which establishes (3.14).

Hence,

$$\max_{j\in H_1}\Big|\frac{1}{\hat{\tau}_j^2}-\frac{1}{\tau_j^2}\Big|\leq\frac{\max_{j\in H_1}|\tau_j^2-\hat{\tau}_j^2|}{\min_{j\in H_1}\tau_j^2}\cdot\max_{j\in H_1}\frac{1}{\hat{\tau}_j^2}=\max_{j\in H_1}|\tau_j^2-\hat{\tau}_j^2|O(1)O_p(1)=O_p(\sqrt{\bar{s}}\lambda_{node}),$$

which establishes (3.15).

We can now bound $\max_{j\in H_1}\|\hat{\Theta}_{Z,j}-\Theta_{Z,j}\|_1$. Use the definition of C_j and (3.10) to recognise that $\Theta_{Z,j}=C_j\Theta_{Z,j,j}=C_j/\tau_j^2$.

$$\begin{aligned}
\max_{j\in H_1}\left\|\hat{\Theta}_{Z,j}-\Theta_{Z,j}\right\|_1&=\max_{j\in H_1}\left\|\frac{\hat{C}_j}{\hat{\tau}_j^2}-\frac{C_j}{\tau_j^2}\right\|_1=\max_{j\in H_1}\left|\frac{1}{\hat{\tau}_j^2}-\frac{1}{\tau_j^2}\right|+\max_{j\in H_1}\left\|\frac{\hat{\phi}_j}{\hat{\tau}_j^2}-\frac{\phi_j}{\tau_j^2}\right\|_1\\
&\leq\max_{j\in H_1}\left|\frac{1}{\hat{\tau}_j^2}-\frac{1}{\tau_j^2}\right|+\max_{j\in H_1}\left\|\frac{\hat{\phi}_j}{\hat{\tau}_j^2}-\frac{\phi_j}{\hat{\tau}_j^2}\right\|_1+\max_{j\in H_1}\left\|\frac{\phi_j}{\hat{\tau}_j^2}-\frac{\phi_j}{\tau_j^2}\right\|_1\\
&=\max_{j\in H_1}\Big|\frac{1}{\hat{\tau}_j^2}-\frac{1}{\tau_j^2}\Big|+\max_{j\in H_1}\frac{1}{\hat{\tau}_j^2}\left\|\hat{\phi}_j-\phi_j\right\|_1+\max_{j\in H_1}\|\phi_j\|_1\Big|\frac{1}{\hat{\tau}_j^2}-\frac{1}{\tau_j^2}\Big|=O_p(\bar{s}\lambda_{node}),
\end{aligned}$$

which establishes (3.16). Next, we bound $\max_{j\in H_1}\|\hat{\Theta}_{Z,j}-\Theta_{Z,j}\|$. Since

$$\left|(\hat{C}_j-C_j)'\frac{Z'Z}{NT}(\hat{C}_j-C_j)-(\hat{C}_j-C_j)'\Psi_Z(\hat{C}_j-C_j)\right|\leq\left\|\frac{Z'Z}{NT}-\Psi_Z\right\|_\infty\left\|\hat{C}_j-C_j\right\|_1^2,$$

$$\max_{j\in H_1}\left|(\hat{C}_j-C_j)'\Psi_Z(\hat{C}_j-C_j)\right|\leq\max_{j\in H_1}\left|(\hat{C}_j-C_j)'\frac{Z'Z}{NT}(\hat{C}_j-C_j)\right|+\left\|\frac{Z'Z}{NT}-\Psi_Z\right\|_\infty\max_{j\in H_1}\left\|\hat{C}_j-C_j\right\|_1^2. \tag{7.22}$$

Consider the first term on the right hand side of (7.22).

$$\max_{j\in H_1}\Big|(\hat{C}_j-C_j)'\frac{Z'Z}{NT}(\hat{C}_j-C_j)\Big|=\max_{j\in H_1}\frac{1}{NT}\left\|Z(\hat{C}_j-C_j)\right\|^2=\max_{j\in H_1}\frac{1}{NT}\left\|Z_{-j}(\hat{\phi}_j-\phi_j)\right\|^2=O_p(\bar{s}\lambda_{node}^2),$$

where the last equality is due to (7.16). Next, consider the second term on the right of the inequality (7.22). Invoke (7.14) with $x=\sqrt{\frac{M(\log p)^3}{N}}\gtrsim\frac{1}{\sqrt{N}}$ $(M>0)$,

$$\mathbb{P}\left(\max_{1\leq l,k\leq p}\Big|\Big[\frac{1}{NT}Z'Z\Big]_{lk}-[\Psi_Z]_{lk}\Big|\geq x\right)\leq Ap^2e^{-B(x^2N)^{1/3}}=Ap^{2-BM^{1/3}},$$

for positive constants A and B. The upper bound of the preceding probability becomes arbitrarily small for sufficiently large N and M. Therefore, $\left\|\frac{Z'Z}{NT}-\Psi_Z\right\|_\infty=O_p\left(\sqrt{\frac{(\log p)^3}{N}}\right)$. We have

$$\left\|\frac{Z'Z}{NT}-\Psi_Z\right\|_\infty\max_{j\in H_1}\left\|\hat{C}_j-C_j\right\|_1^2=\left\|\frac{Z'Z}{NT}-\Psi_Z\right\|_\infty\max_{j\in H_1}\left\|\hat{\phi}_j-\phi_j\right\|_1^2=O_p\left(\sqrt{\frac{(\log p)^3}{N}}\bar{s}^2\lambda_{node}^2\right),$$

where the first equality is due to the definitions of $\hat{C}_j$ and C_j, and the second equality is due to (7.17). Adding up the two terms, we have

$$\max_{j\in H_1}\left|(\hat{C}_j-C_j)'\Psi_Z(\hat{C}_j-C_j)\right|\leq O_p(\bar{s}\lambda_{node}^2)+O_p\left(\sqrt{\frac{(\log p)^3}{N}}\bar{s}^2\lambda_{node}^2\right)=O_p(\bar{s}\lambda_{node}^2),$$

where the last equality is due to $\frac{(\log p)^3\bar{s}^2}{N}=o(1)$ by Assumption 4(b). Since $\max_{j\in H_1}|(\hat{C}_j-C_j)'\Psi_Z(\hat{C}_j-C_j)|\geq$ $\text{mineval}(\Psi_Z)\max_{j\in H_1}\|\hat{C}_j-C_j\|^2$ and $\text{mineval}(\Psi_Z)$ is uniformly bounded away from zero we have

$\max_{j\in H_1}\|\hat{\phi}_j-\phi_j\| = \max_{j\in H_1}\|\hat{C}_j - C_j\| = O_p(\sqrt{\bar{s}}\lambda_{node})$. Then,

$$\max_{j\in H_1}\left\|\hat{\Theta}_{Z,j}-\Theta_{Z,j}\right\| = \max_{j\in H_1}\left\|\frac{\hat{C}_j}{\hat{\tau}_j^2}-\frac{C_j}{\tau_j^2}\right\| \leq \max_{j\in H_1}\left|\frac{1}{\hat{\tau}_j^2}-\frac{1}{\tau_j^2}\right| + \max_{j\in H_1}\left\|\frac{\hat{\phi}_j}{\hat{\tau}_j^2}-\frac{\phi_j}{\tau_j^2}\right\|$$
$$\leq \max_{j\in H_1}\left|\frac{1}{\hat{\tau}_j^2}-\frac{1}{\tau_j^2}\right| + \max_{j\in H_1}\frac{1}{\hat{\tau}_j^2}\left\|\hat{\phi}_j-\phi_j\right\| + \max_{j\in H_1}\|\phi_j\|\left|\frac{1}{\hat{\tau}_j^2}-\frac{1}{\tau_j^2}\right| = O_p(\sqrt{\bar{s}}\lambda_{node}),$$

where in the last equality we have used that $\max_{j\in H_1}\|\phi_j\| = O(1)$, which follows from (7.20). We have hence established (3.17). Finally, recall that $\Theta_{Z,j} = C_j\Theta_{Z,j,j} = C_j/\tau_j^2$. Thus,

$$\max_{j\in H_1}\|\Theta_{Z,j}\|_1 = \max_{j\in H_1}|1/\tau_j^2| + \max_{j\in H_1}\|\phi_j\|_1\max_{j\in H_1}1/\tau_j^2 = O(\sqrt{\bar{s}}). \tag{7.23}$$

Therefore,

$$\max_{j\in H_1}\left\|\hat{\Theta}_{Z,j}\right\|_1 \leq \max_{j\in H_1}\left\|\hat{\Theta}_{Z,j}-\Theta_{Z,j}\right\|_1 + \max_{j\in H_1}\|\Theta_{Z,j}\|_1 = O_p(\bar{s}\lambda_{node}) + O(\sqrt{\bar{s}}) = O_p(\sqrt{\bar{s}}),$$

where the last equality is due to $\frac{(\log p)^3\bar{s}^2}{N} = o(1)$ by Assumption 4(b). □

7.4 Proof of Theorem 2

Proof of Theorem 2. The following assumption is implied by Assumption 5. However, as Assumption 5 is much simpler we have chosen to use the latter in the main text even though it is slightly less general than the following assumption. Note again how the assumptions simplifies when wither h_1 or h_2 equals 0.

Assumption 6.

(a) (i) $\frac{h_1^2\bar{s}^2(\log(p\vee T))^5}{N} = o(1)$; *(ii)* $\frac{h_1h_2\bar{s}(\log(p\vee N\vee T))^3}{N} = o(1)$; *(iii)* $\frac{h_1\bar{s}^2(\log p)^3(\log(p\vee N))^3}{N} = o(1)$; *(iv)* $\frac{h_1\bar{s}(\log(N\vee T))^2(\log p)^2}{NT} = o(1)$; *(v)* $\frac{(\log(N\vee T))^2 1\{h_2\neq 0\}}{T} = o(1)$.

(b) Let

$$a := \frac{\left[(s_1+s_2)\vee\frac{\sqrt{T}c_N}{\sqrt{(\log(p\vee N))^3}}\right](\log(p\vee N))^3}{NT}.$$

(i) $h_1^2\bar{s}^2\left(1\vee\sqrt{\frac{(\log(p\vee T))^7}{N}}\right)a = o(1)$; *(ii)* $h_1\bar{s}\log(p\vee N\vee T)a = o(1)$;
(iii) $h_1h_2\bar{s}(\log(p\vee N\vee T))^2a = o(1)$; *(iv)* $\sqrt{h_1h_2\bar{s}N\log(p\vee N\vee T)}a = o(1)$;
(v) $Nh_2^2\left(1\vee\sqrt{\frac{(\log N)^3}{T}}\right)a = o(1)$.

(c)

$$\frac{(h_1\vee h_2)\left[(s_1+s_2)^2(\log(p\vee N))^3\vee Tc_N^2\right]b}{N} = o(1),$$

where $b := \left[\left(\bar{s}\log(p\vee N)\vee(\log p)^3\right)1\{h_1\neq 0\}\right]\vee\left[\log(p\vee N)1\{h_2\neq 0\}\right]$.

(d) mineval$(\Sigma_{\Pi\varepsilon})$ *is uniformly bounded away from zero and maxeval*$(\Sigma_{1,N})$ *is uniformly bounded from above.*

We show that

$$t = \frac{\rho' S\left(\tilde{\gamma}-\gamma\right)}{\sqrt{\rho'\hat{\Theta}\hat{\Sigma}_{\Pi\varepsilon}\hat{\Theta}'\rho}} \xrightarrow{d} N(0,1).$$

To this end, note that by (3.3) one may write $t = t_1 + t_2$, where

$$t_1 = \frac{\rho'\hat{\Theta}S^{-1}\Pi'\varepsilon}{\sqrt{\rho'\hat{\Theta}\hat{\Sigma}_{\Pi\varepsilon}\hat{\Theta}'\rho}} \quad \text{and} \quad t_2 = \frac{-\rho'\Delta}{\sqrt{\rho'\hat{\Theta}\hat{\Sigma}_{\Pi\varepsilon}\hat{\Theta}'\rho}}.$$

Defining

$$t_1' = \frac{\rho'\Theta S^{-1}\Pi'\varepsilon}{\sqrt{\rho'\Theta\Sigma_{\Pi\varepsilon}\Theta'\rho}}$$

it suffices to show that $t_1' \xrightarrow{d} N(0,1)$, $t_1' - t_1 = o_p(1)$, and $t_2 = o_p(1)$. In the sequel we first show that $t_1 - t_1' = o_p(1)$, then $t_1' \xrightarrow{d} N(0,1)$ and finally $t_2 = o_p(1)$. To show that $t_1 - t_1' = o_p(1)$, it suffices to show that the denominators as well as the numerators of t_1 and t_1' are asymptotically equivalent, since

$$\rho'\Theta\Sigma_{\Pi\varepsilon}\Theta'\rho \geq \text{mineval}(\Sigma_{\Pi\varepsilon})\left(\text{mineval}(\Theta)\right)^2 = \frac{\text{mineval}(\Sigma_{\Pi\varepsilon})}{\left(\text{maxeval}(\Psi)\right)^2} \tag{7.24}$$

which is uniformly bounded away from zero by Assumptions 4(a) and 6(d).

7.4.1 Denominators of t_1 and t_1'

We first show that the denominators of t_1 and t_1' are asymptotically equivalent, i.e.,

$$|\rho'\hat{\Theta}\hat{\Sigma}_{\Pi\varepsilon}\hat{\Theta}'\rho - \rho'\Theta\Sigma_{\Pi\varepsilon}\Theta'\rho| = o_p(1). \tag{7.25}$$

Write

$$\left|(\rho_1', \rho_2')\begin{pmatrix} \hat{\Theta}_Z\hat{\Sigma}_{1,N}\hat{\Theta}_Z' & \hat{\Theta}_Z\hat{\Sigma}_{2,N} \\ \hat{\Sigma}_{2,N}'\hat{\Theta}_Z' & \hat{\Sigma}_{3,N} \end{pmatrix}\begin{pmatrix} \rho_1 \\ \rho_2 \end{pmatrix} - (\rho_1', \rho_2')\begin{pmatrix} \Theta_Z\Sigma_{1,N}\Theta_Z' & \Theta_Z\Sigma_{2,N} \\ \Sigma_{2,N}'\Theta_Z' & \Sigma_{3,N} \end{pmatrix}\begin{pmatrix} \rho_1 \\ \rho_2 \end{pmatrix}\right|$$

$$\leq |\rho_1'\hat{\Theta}_Z\hat{\Sigma}_{1,N}\hat{\Theta}_Z'\rho_1 - \rho_1'\Theta_Z\Sigma_{1,N}\Theta_Z'\rho_1| \tag{7.26}$$

$$+ 2|\rho_1'\hat{\Theta}_Z\hat{\Sigma}_{2,N}\rho_2 - \rho_1'\Theta_Z\Sigma_{2,N}\rho_2| \tag{7.27}$$

$$+ |\rho_2'\hat{\Sigma}_{3,N}\rho_2 - \rho_2'\Sigma_{3,N}\rho_2|. \tag{7.28}$$

To establish (7.25), we show that (7.26), (7.27) and (7.28) are $o_p(1)$, respectively.

(7.26) is $o_p(1)$:

Define $\tilde{\Sigma}_{1,N} := \frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T \varepsilon_{i,t}^2 z_{i,t}z_{i,t}'$. To show that (7.26) is $o_p(1)$, it suffices to show that

$$|\rho_1'\hat{\Theta}_Z\hat{\Sigma}_{1,N}\hat{\Theta}_Z'\rho_1 - \rho_1'\hat{\Theta}_Z\tilde{\Sigma}_{1,N}\hat{\Theta}_Z'\rho_1| = o_p(1) \tag{7.29}$$

$$|\rho_1'\hat{\Theta}_Z\tilde{\Sigma}_{1,N}\hat{\Theta}_Z'\rho_1 - \rho_1'\hat{\Theta}_Z\Sigma_{1,N}\hat{\Theta}_Z'\rho_1| = o_p(1) \tag{7.30}$$

$$|\rho_1'\hat{\Theta}_Z\Sigma_{1,N}\hat{\Theta}_Z'\rho_1 - \rho_1'\Theta_Z\Sigma_{1,N}\Theta_Z'\rho_1| = o_p(1). \tag{7.31}$$

We prove (7.29) first. Note that

$$|\rho_1'\hat{\Theta}_Z\hat{\Sigma}_{1,N}\hat{\Theta}_Z'\rho_1 - \rho_1'\hat{\Theta}_Z\tilde{\Sigma}_{1,N}\hat{\Theta}_Z'\rho_1| \leq \|\hat{\Sigma}_{1,N} - \tilde{\Sigma}_{1,N}\|_\infty \|\hat{\Theta}_Z'\rho_1\|_1^2.$$

First,

$$\|\hat{\Theta}_Z'\rho_1\|_1 = \left\|\sum_{j\in H_1}\hat{\Theta}_{Z,j}\rho_{1j}\right\|_1 \leq \sum_{j\in H_1}|\rho_{1j}|\left\|\hat{\Theta}_{Z,j}\right\|_1 = O_p(\sqrt{h_1\bar{s}}), \tag{7.32}$$

where the last equality is due to (3.18). We now bound $\left\|\hat{\Sigma}_{1,N}-\tilde{\Sigma}_{1,N}\right\|_\infty$. Since $\hat{\varepsilon}_{i,t}=y_{i,t}-z_{i,t}'\hat{\alpha}-\hat{\eta}_i=\varepsilon_{i,t}-z_{i,t}'(\hat{\alpha}-\alpha)-(\hat{\eta}_i-\eta_i)=:\varepsilon_{i,t}-\pi_{i,t}(\hat{\gamma}-\gamma)$, substituting for $\hat{\varepsilon}_{i,t}$, we have

$$\begin{aligned}\left\|\hat{\Sigma}_{1,N}-\tilde{\Sigma}_{1,N}\right\|_\infty &= \left\|\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\hat{\varepsilon}_{i,t}^2 z_{i,t}z_{i,t}'-\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\varepsilon_{i,t}^2 z_{i,t}z_{i,t}'\right\|_\infty \\ &\leq 2\left\|\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}z_{i,t}z_{i,t}'\varepsilon_{i,t}\pi_{i,t}'(\hat{\gamma}-\gamma)\right\|_\infty+\left\|\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}z_{i,t}z_{i,t}'[\pi_{i,t}'(\hat{\gamma}-\gamma)]^2\right\|_\infty. \end{aligned} \tag{7.33}$$

Consider the first term of (7.33). A typical element of $\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}z_{i,t}z_{i,t}'\varepsilon_{i,t}\pi_{i,t}'(\hat{\gamma}-\gamma)$ is

$$\begin{aligned}\frac{1}{NT}\sum_{j=1}^{NT}z_{j,l}z_{j,k}\varepsilon_j\pi_j'(\hat{\gamma}-\gamma) &\leq \frac{1}{NT}\left(\sum_{j=1}^{NT}z_{j,l}^2z_{j,k}^2\varepsilon_j^2\right)^{1/2}\left(\sum_{j=1}^{NT}[\pi_j'(\hat{\gamma}-\gamma)]^2\right)^{1/2} \\ &= \left(\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}z_{i,t,l}^2z_{i,t,k}^2\varepsilon_{i,t}^2\right)^{1/2}\left(\frac{1}{NT}\left\|\Pi(\hat{\gamma}-\gamma)\right\|^2\right)^{1/2} \end{aligned} \tag{7.34}$$

for some $l,k\in\{1,\dots,p\}$, where the inequality is due to Cauchy-Schwarz inequality. Use independence across i (Assumption 1) and subgaussianity (Assumption 3) to invoke Proposition 3 in Appendix B, such that

$$\max_{1\leq l\leq p}\max_{1\leq k\leq p}\left|\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}(z_{i,t,l}^2z_{i,t,k}^2\varepsilon_{i,t}^2-\mathbb{E}[z_{i,t,l}^2z_{i,t,k}^2\varepsilon_{i,t}^2])\right|=O_p\left(\sqrt{\frac{(\log(p^2T))^7}{N}}\right)$$

and

$$\max_{1\leq l\leq p}\max_{1\leq k\leq p}\max_{1\leq i\leq N}\max_{1\leq t\leq T}\mathbb{E}[z_{i,t,l}^2z_{i,t,k}^2\varepsilon_{i,t}^2]\leq A=O(1)$$

for some positive constant A. Then, by the triangle inequality,

$$\max_{1\leq l\leq p}\max_{1\leq k\leq p}\left|\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}z_{i,t,l}^2z_{i,t,k}^2\varepsilon_{i,t}^2\right|=O_p\left(\sqrt{\frac{(\log(p\vee T))^7}{N}}\right)+O(1). \tag{7.35}$$

Combining (7.34) and (7.35), we have

$$\left\|\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}z_{i,t}z_{i,t}'\varepsilon_{i,t}\pi_{i,t}'(\hat{\gamma}-\gamma)\right\|_\infty=O_p\left(\frac{(\log(p\vee T))^{7/4}}{N^{1/4}}\vee 1\right)\left(\frac{1}{NT}\left\|\Pi(\hat{\gamma}-\gamma)\right\|^2\right)^{1/2}. \tag{7.36}$$

We now consider the second term of (7.33). A typical element of $\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}z_{i,t}z_{i,t}'[\pi_{i,t}'(\hat{\gamma}-\gamma)]^2$ is $\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}z_{i,t,l}z_{i,t,k}[\pi_{i,t}'(\hat{\gamma}-\gamma)]^2\leq\max_{1\leq i\leq N}\max_{1\leq t\leq T}|z_{i,t,l}z_{i,t,k}|\frac{1}{NT}\|\Pi(\hat{\gamma}-\gamma)\|^2$ for some $l,k\in\{1,\dots,p\}$. Recall that we have proved in the proof of Lemma 7 that $\|z_{i,t,l}z_{i,t,k}\|_{\psi_1}\leq(1+K)/C$. Using the definition of the Orlicz norm, we have $\mathbb{E}e^{\frac{C}{1+K}|z_{i,t,l}z_{i,t,k}|}\leq 2$. Using Markov's inequality, we have for any $\epsilon>0$

$$\mathbb{P}\left(\max_{1\leq l\leq p}\max_{1\leq k\leq p}\max_{1\leq i\leq N}\max_{1\leq t\leq T}|z_{i,t,l}z_{i,t,k}|\geq\epsilon\right)\leq\sum_{l=1}^{p}\sum_{k=1}^{p}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{\mathbb{E}e^{\frac{C}{1+K}|z_{i,t,l}z_{i,t,k}|}}{e^{\frac{C}{1+K}\epsilon}}\leq 2NTp^2e^{-\frac{C}{1+K}\epsilon}.$$

Set $\epsilon=M\log(p^2NT)$ for some $M>0$ and note that the upper bound of the preceding probability becomes arbitrarily small for N and M sufficiently large. Thus,

$$\max_{1\leq l\leq p}\max_{1\leq k\leq p}\max_{1\leq i\leq N}\max_{1\leq t\leq T}|z_{i,t,l}z_{i,t,k}|=O_p(\log(p^2NT))$$

and we get

$$\left\|\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} z_{i,t}z_{i,t}'[\pi_{i,t}'(\hat{\gamma}-\gamma)]^2\right\|_\infty = O_p(\log(p\vee N\vee T))\frac{1}{NT}\|\Pi(\hat{\gamma}-\gamma)\|^2. \tag{7.37}$$

Combining (7.36) and (7.37), conclude

$$\begin{aligned}&\left\|\hat{\Sigma}_{1,N}-\tilde{\Sigma}_{1,N}\right\|_\infty \\ &= O_p\left(\frac{(\log(p\vee T))^{7/4}}{N^{1/4}}\vee 1\right)\left(\frac{1}{NT}\left\|\Pi(\hat{\gamma}-\gamma)\right\|^2\right)^{1/2} + O_p(\log(p\vee N\vee T))\frac{1}{NT}\|\Pi(\hat{\gamma}-\gamma)\|^2.\end{aligned}$$

Therefore, combining the preceding rates with (7.32) one gets

$$\begin{aligned}&|\rho_1'\hat{\Theta}_Z\hat{\Sigma}_{1,N}\hat{\Theta}_Z'\rho_1-\rho_1'\hat{\Theta}_Z\tilde{\Sigma}_{1,N}\hat{\Theta}_Z'\rho_1| \\ &= O_p(h_1\bar{s})O_p\left(\frac{(\log(p\vee T))^{7/4}}{N^{1/4}}\vee 1\right)\left[\frac{1}{NT}\left\|\Pi(\hat{\gamma}-\gamma)\right\|^2\right]^{1/2} + O_p(h_1\bar{s})O_p(\log(p\vee N\vee T))\frac{1}{NT}\|\Pi(\hat{\gamma}-\gamma)\|^2 \\ &= o_p(1),\end{aligned}$$

where the last equality is also due to Assumption 6(b)(i)-(ii), which establishes (7.29).

Next, turn to (7.30). Note that

$$|\rho_1'\hat{\Theta}_Z\tilde{\Sigma}_{1,N}\hat{\Theta}_Z'\rho_1-\rho_1'\hat{\Theta}_Z\Sigma_{1,N}\hat{\Theta}_Z'\rho_1| \le \left\|\tilde{\Sigma}_{1,N}-\Sigma_{1,N}\right\|_\infty\left\|\hat{\Theta}_Z'\rho_1\right\|_1^2.$$

Given (7.32), we only need to consider $\|\tilde{\Sigma}_{1,N}-\Sigma_{1,N}\|_\infty$. Using independence across i (Assumption 1) and subgaussianity (Assumption 3) to invoke Proposition 3 in Appendix B such that

$$\left\|\tilde{\Sigma}_{1,N}-\Sigma_{1,N}\right\|_\infty = \max_{1\le l\le p}\max_{1\le k\le p}\left|\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}(z_{i,t,l}z_{i,t,k}\varepsilon_{it}^2-\mathbb{E}[z_{i,t,l}z_{i,t,k}\varepsilon_{i,t}^2])\right| = O_p\left(\sqrt{\frac{(\log(p^2T))^5}{N}}\right). \tag{7.38}$$

Thus,

$$|\rho_1'\hat{\Theta}_Z\tilde{\Sigma}_{1,N}\hat{\Theta}_Z'\rho_1-\rho_1'\hat{\Theta}_Z\Sigma_{1,N}\hat{\Theta}_Z'\rho_1| = O_p\left(\sqrt{\frac{(\log(p\vee T))^5}{N}}h_1\bar{s}\right) = o_p(1),$$

where the last equality is due to Assumption 6(a)(i), establishing (7.30).

To prove (7.31) invoke Lemma 9 in Appendix B:

$$\begin{aligned}&|\rho_1'\hat{\Theta}_Z\Sigma_{1,N}\hat{\Theta}_Z'\rho_1-\rho_1'\Theta_Z\Sigma_{1,N}\Theta_Z'\rho_1| \le \|\Sigma_{1,N}\|_\infty\|(\hat{\Theta}_Z'-\Theta_Z')\rho_1\|_1^2+2\|\Sigma_{1,N}\Theta_Z'\rho_1\|\|(\hat{\Theta}_Z'-\Theta_Z')\rho_1\| \\ &\le \|\Sigma_{1,N}\|_\infty\|(\hat{\Theta}_Z'-\Theta_Z')\rho_1\|_1^2+2\mathrm{maxeval}(\Sigma_{1,N})\|\Theta_Z'\rho_1\|\|(\hat{\Theta}_Z'-\Theta_Z')\rho_1\|.\end{aligned}$$

First, note that $\|\Sigma_{1,N}\|_\infty$ is uniformly bounded as every entry is an average of uniformly bounded population moments (see Proposition 3 in appendix B).

$$\begin{aligned}\|(\hat{\Theta}_Z'-\Theta_Z')\rho_1\|_1 &\le \sum_{j\in H_1}\left\|\hat{\Theta}_{Z,j}-\Theta_{Z,j}\right\|_1|\rho_{1j}| \le \max_{j\in H_1}\left\|\hat{\Theta}_{Z,j}-\Theta_{Z,j}\right\|_1\sqrt{h_1} \\ &= O_p(\bar{s}\lambda_{node}\sqrt{h_1}) = o_p(1),\end{aligned} \tag{7.39}$$

where the first equality is due to (3.16), and the last equality is due to Assumption 6(a)(i). Next, $\|\Theta'_Z\rho_1\| \leq \text{maxeval}(\Theta_Z)\|\rho_1\| \leq \text{maxeval}(\Theta_Z) = 1/\text{mineval}(\Psi_Z)$, which is uniformly bounded from above by Assumption 4(a). Furthermore,

$$\begin{aligned}\|(\hat{\Theta}'_Z - \Theta'_Z)\rho_1\| &= \Big\|\sum_{j\in H_1}(\hat{\Theta}_{Z,j} - \Theta_{Z,j})\rho_{1j}\Big\| \leq \sum_{j\in H_1}\Big\|\hat{\Theta}_{Z,j} - \Theta_{Z,j}\Big\| |\rho_{1j}| \\ &\leq \max_{j\in H_1}\big\|\hat{\Theta}_{Z,j} - \Theta_{Z,j}\big\|\sqrt{h_1} = O_p(\sqrt{\bar{s}}\lambda_{node}\sqrt{h_1}) = o_p(1),\end{aligned}$$

where the second last to equality is due to (3.17), and the last equality is due to (7.39). Thus, we have established (7.31) concluding the proof of (7.26) is $o_p(1)$.

(7.27) is $o_p(1)$:

Define $\tilde{\Sigma}_{2,N} := \frac{1}{\sqrt{NT}}\sum_{i=1}^N\sum_{t=1}^T \varepsilon_{i,t}^2 z_{i,t}d'_{i,t}$. It suffices to show

$$|\rho_1'\hat{\Theta}_Z\hat{\Sigma}_{2,N}\rho_2 - \rho_1'\hat{\Theta}_Z\tilde{\Sigma}_{2,N}\rho_2| = o_p(1) \tag{7.40}$$

$$|\rho_1'\hat{\Theta}_Z\tilde{\Sigma}_{2,N}\rho_2 - \rho_1'\hat{\Theta}_Z\Sigma_{2,N}\rho_2| = o_p(1) \tag{7.41}$$

$$|\rho_1'\hat{\Theta}_Z\Sigma_{2,N}\rho_2 - \rho_1'\Theta_Z\Sigma_{2,N}\rho_2| = o_p(1). \tag{7.42}$$

Consider (7.40) first. Note that

$$\begin{aligned}|\rho_1'\hat{\Theta}_Z\hat{\Sigma}_{2,N}\rho_2 - \rho_1'\hat{\Theta}_Z\tilde{\Sigma}_{2,N}\rho_2| &\leq \Big\|\rho_1'\hat{\Theta}_Z\left(\hat{\Sigma}_{2,N} - \tilde{\Sigma}_{2,N}\right)\Big\|_\infty \|\rho_2\|_1 \\ &\leq \big\|\rho_1'\hat{\Theta}_Z\big\|_1\big\|\hat{\Sigma}_{2,N} - \tilde{\Sigma}_{2,N}\big\|_\infty\sqrt{h_2} = O_p\left(\sqrt{h_1h_2\bar{s}}\right)\big\|\hat{\Sigma}_{2,N} - \tilde{\Sigma}_{2,N}\big\|_\infty,\end{aligned}$$

where the last equality is due to (7.32). In addition,

$$\begin{aligned}\Big\|\hat{\Sigma}_{2,N} - \tilde{\Sigma}_{2,N}\Big\|_\infty &= \Bigg\|\frac{1}{\sqrt{NT}}\sum_{i=1}^N\sum_{t=1}^T\hat{\varepsilon}_{i,t}^2 z_{i,t}d'_{i,t} - \frac{1}{\sqrt{NT}}\sum_{i=1}^N\sum_{t=1}^T\varepsilon_{i,t}^2 z_{i,t}d'_{i,t}\Bigg\|_\infty \\ &\leq 2\Bigg\|\frac{1}{\sqrt{NT}}\sum_{i=1}^N\sum_{t=1}^T z_{i,t}d'_{i,t}\varepsilon_{i,t}\pi'_{i,t}(\hat{\gamma}-\gamma)\Bigg\|_\infty + \Bigg\|\frac{1}{\sqrt{NT}}\sum_{i=1}^N\sum_{t=1}^T z_{i,t}d'_{i,t}[\pi'_{i,t}(\hat{\gamma}-\gamma)]^2\Bigg\|_\infty \end{aligned}\tag{7.43}$$

Consider the first term of (7.43). A typical element of $\frac{1}{\sqrt{NT}}\sum_{i=1}^N\sum_{t=1}^T z_{i,t}d'_{i,t}\varepsilon_{i,t}\pi'_{i,t}(\hat{\gamma}-\gamma)$ is

$$\begin{aligned}\frac{1}{\sqrt{NT}}\sum_{j=1}^{NT} z_{j,l}d_{j,k}\varepsilon_j\pi'_j(\hat{\gamma}-\gamma) &\leq \frac{1}{\sqrt{NT}}\left(\sum_{j=1}^{NT} z_{j,l}^2d_{j,k}^2\varepsilon_j^2\right)^{1/2}\left(\sum_{j=1}^{NT}[\pi'_j(\hat{\gamma}-\gamma)]^2\right)^{1/2} \\ &= \left(\frac{1}{T}\sum_{i=1}^N\sum_{t=1}^T z_{i,t,l}^2d_{i,t,k}^2\varepsilon_{i,t}^2\right)^{1/2}\frac{1}{\sqrt{NT}}\|\Pi(\hat{\gamma}-\gamma)\| = \left(\frac{1}{T}\sum_{t=1}^T z_{k,t,l}^2\varepsilon_{k,t}^2\right)^{1/2}\frac{1}{\sqrt{NT}}\|\Pi(\hat{\gamma}-\gamma)\|\end{aligned}$$

for some $l \in \{1,\ldots,p\}$ and $k \in \{1,\ldots,N\}$ where the inequality is due to Cauchy-Schwarz inequality. By subgaussianity, Assumption 3, we can use the same technique as in (8.3) in Proposition 3 in Appendix B to prove $\mathbb{E}e^{D\left|\frac{1}{T}\sum_{t=1}^T z_{i,t,l}^2\varepsilon_{i,t}^2\right|^{1/2}} \leq BT$ for positive constants D, B. Using Markov's inequality, we have for $\epsilon > 0$

$$\mathbb{P}\left(\max_{1\leq l\leq p}\max_{1\leq k\leq N}\left|\frac{1}{T}\sum_{t=1}^T z_{k,t,l}^2\varepsilon_{k,t}^2\right| \geq \epsilon\right) \leq \sum_{l=1}^p\sum_{k=1}^N\frac{\mathbb{E}e^{D\left|\frac{1}{T}\sum_{t=1}^T z_{k,t,l}^2\varepsilon_{k,t}^2\right|^{1/2}}}{e^{D\epsilon^{1/2}}} \leq BpNTe^{-D\epsilon^{1/2}}.$$

Set $\epsilon = M(\log(pNT))^2$ for some $M > 0$ and note that the upper bound of the preceding probability becomes arbitrarily small for N and M sufficiently large. Thus, $\max_{1\le l\le p}\max_{1\le k\le N}\left|\frac{1}{T}\sum_{t=1}^T z_{k,t,l}^2\varepsilon_{k,t}^2\right| = O_p((\log(pNT))^2)$. Therefore,

$$\begin{aligned}
&\left\|\frac{1}{\sqrt{NT}}\sum_{i=1}^N\sum_{t=1}^T z_{i,t}d_{i,t}'\varepsilon_{i,t}\pi_{i,t}'(\hat{\gamma}-\gamma)\right\|_\infty \le \left(\max_{1\le l\le p}\max_{1\le k\le N}\frac{1}{T}\sum_{t=1}^T z_{k,t,l}^2\varepsilon_{k,t}^2\right)^{1/2}\frac{1}{\sqrt{NT}}\left\|\Pi(\hat{\gamma}-\gamma)\right\| \\
&\le O_p(\log(pNT))\frac{1}{\sqrt{NT}}\left\|\Pi(\hat{\gamma}-\gamma)\right\|. \qquad (7.44)
\end{aligned}$$

Now consider the second term of (7.43). A typical element of $\frac{1}{\sqrt{NT}}\sum_{i=1}^N\sum_{t=1}^T z_{i,t}d_{i,t}'[\pi_{i,t}'(\hat{\gamma}-\gamma)]^2$ is

$$\begin{aligned}
&\frac{1}{\sqrt{NT}}\sum_{i=1}^N\sum_{t=1}^T z_{i,t,l}d_{i,t,k}[\pi_{i,t}'(\hat{\gamma}-\gamma)]^2 \le \max_{1\le i\le N}\max_{1\le t\le T}\sqrt{N}|z_{i,t,l}d_{i,t,k}|\frac{1}{NT}\|\Pi(\hat{\gamma}-\gamma)\|^2 \\
&\le \max_{1\le t\le T}\sqrt{N}|z_{k,t,l}|\frac{1}{NT}\|\Pi(\hat{\gamma}-\gamma)\|^2
\end{aligned}$$

for some $l\in\{1,\dots,p\}$, $k\in\{1,\dots,N\}$. Using Markov's inequality, we have for any $\epsilon > 0$

$$\mathbb{P}\left(\max_{1\le l\le p}\max_{1\le k\le N}\max_{1\le t\le T}|z_{k,t,l}|\ge\epsilon\right)\le\sum_{l=1}^p\sum_{k=1}^N\sum_{t=1}^T\mathbb{P}\left(|z_{k,t,l}|\ge\epsilon\right)\le pNT\frac{K}{2}e^{-C\epsilon^2}.$$

Set $\epsilon = \sqrt{M\log(pNT)}$ for some $M > 0$ to see that the upper bound of the preceding probability becomes arbitrarily small for N and M sufficiently large. Thus, $\max_{1\le l\le p}\max_{1\le k\le N}\max_{1\le t\le T}|z_{k,t,l}| = O_p(\sqrt{\log(pNT)})$. In total,

$$\begin{aligned}
&\left\|\frac{1}{\sqrt{NT}}\sum_{i=1}^N\sum_{t=1}^T z_{i,t}d_{i,t}'[\pi_{i,t}'(\hat{\gamma}-\gamma)]^2\right\|_\infty \le \max_{1\le l\le p}\max_{1\le k\le N}\max_{1\le t\le T}\sqrt{N}|z_{k,t,l}|\frac{1}{NT}\|\Pi(\hat{\gamma}-\gamma)\|^2 \\
&= O_p(\sqrt{N\log(pNT)})\frac{1}{NT}\|\Pi(\hat{\gamma}-\gamma)\|^2. \qquad (7.45)
\end{aligned}$$

Therefore, combining (7.44) and (7.45)

$$\begin{aligned}
&|\rho_1'\hat{\Theta}_Z\hat{\Sigma}_{2,N}\rho_2 - \rho_1'\hat{\Theta}_Z\tilde{\Sigma}_{2,N}\rho_2| \le \left\|\hat{\Sigma}_{2,N}-\tilde{\Sigma}_{2,N}\right\|_\infty O_p(\sqrt{h_1h_2\bar{s}}) \\
&= O_p\left(\sqrt{h_1h_2\bar{s}}\log(pNT)\right)\frac{1}{\sqrt{NT}}\left\|\Pi(\hat{\gamma}-\gamma)\right\| + O_p\left(\sqrt{h_1h_2\bar{s}N\log(pNT)}\right)\frac{1}{NT}\|\Pi(\hat{\gamma}-\gamma)\|^2 \\
&= o_p(1),
\end{aligned}$$

where the last equality is due to Assumption 6(b)(iii)-(iv), which establishes (7.40).

Next, turn to (7.41). Note that

$$|\rho_1'\hat{\Theta}_Z\tilde{\Sigma}_{2,N}\rho_2 - \rho_1'\hat{\Theta}_Z\Sigma_{2,N}\rho_2| \le \left\|\tilde{\Sigma}_{2,N}-\Sigma_{2,N}\right\|_\infty\left\|\hat{\Theta}_Z'\rho_1\right\|_1\sqrt{h_2}.$$

Given (7.32), it suffices to consider

$$\begin{aligned}
\left\|\tilde{\Sigma}_{2,N}-\Sigma_{2,N}\right\|_\infty &= \max_{1\le l\le p}\max_{1\le k\le N}\left|\frac{1}{\sqrt{NT}}\sum_{i=1}^N\sum_{t=1}^T(z_{i,t,l}d_{i,t,k}\varepsilon_{i,t}^2 - \mathbb{E}[z_{i,t,l}d_{i,t,k}\varepsilon_{i,t}^2])\right| \\
&= \max_{1\le l\le p}\max_{1\le k\le N}\left|\frac{1}{\sqrt{NT}}\sum_{t=1}^T(z_{k,t,l}\varepsilon_{k,t}^2 - \mathbb{E}[z_{k,t,l}\varepsilon_{k,t}^2])\right|.
\end{aligned}$$

By subgaussianity, Assumption 3, we can use the same technique as in (8.3) in Proposition 3 in Appendix B to prove $\mathbb{E}e^{D|\frac{1}{T}\sum_{t=1}^{T}(z_{k,t,l}\varepsilon_{k,t}^2-\mathbb{E}[z_{k,t,l}\varepsilon_{k,t}^2])|^{2/3}} \leq BT$ for some positive constant B. Using Markov's inequality, we have for any $\epsilon > 0$

$$\mathbb{P}\left(\max_{1\leq l\leq p}\max_{1\leq k\leq N}\left|\frac{1}{T}\sum_{t=1}^{T}(z_{k,t,l}\varepsilon_{k,t}^2-\mathbb{E}[z_{k,t,l}\varepsilon_{k,t}^2])\right|\geq\epsilon\right)\leq\sum_{l=1}^{p}\sum_{k=1}^{N}\mathbb{P}\left(\left|\frac{1}{T}\sum_{t=1}^{T}(z_{k,t,l}\varepsilon_{k,t}^2-\mathbb{E}[z_{k,t,l}\varepsilon_{k,t}^2])\right|\geq\epsilon\right)$$
$$\leq\sum_{l=1}^{p}\sum_{k=1}^{N}\frac{\mathbb{E}e^{D\left|\frac{1}{T}\sum_{t=1}^{T}(z_{k,t,l}\varepsilon_{k,t}^2-\mathbb{E}[z_{k,t,l}\varepsilon_{k,t}^2])\right|^{2/3}}}{e^{D\epsilon^{2/3}}}\leq BpNTe^{-D\epsilon^{2/3}}.$$

Set $\epsilon=\sqrt{M(\log(pNT))^3}$ for some $M>0$ and note that the upper bound of the preceding probability becomes arbitrarily small for N and M sufficiently large. Thus,

$$\max_{1\leq l\leq p}\max_{1\leq k\leq N}\left|\frac{1}{T}\sum_{t=1}^{T}(z_{k,t,l}\varepsilon_{k,t}^2-\mathbb{E}[z_{k,t,l}\varepsilon_{k,t}^2])\right|=O_p\left(\sqrt{(\log(pNT))^3}\right)$$

and so

$$\left\|\tilde{\Sigma}_{2,N}-\Sigma_{2,N}\right\|_\infty=\frac{1}{\sqrt{N}}\max_{1\leq l\leq p}\max_{1\leq k\leq N}\left|\frac{1}{T}\sum_{t=1}^{T}(z_{k,t,l}\varepsilon_{k,t}^2-\mathbb{E}[z_{k,t,l}\varepsilon_{k,t}^2])\right|=O_p\left(\sqrt{\frac{(\log(pNT))^3}{N}}\right). \tag{7.46}$$

In total,

$$|\rho_1'\hat{\Theta}_Z\tilde{\Sigma}_{2,N}\rho_2-\rho_1'\hat{\Theta}_Z\Sigma_{2,N}\rho_2|=O_p\left(\sqrt{\frac{(\log(p\vee N\vee T))^3h_1h_2\bar{s}}{N}}\right)=o_p(1),$$

where the last equality is due to Assumption 6(a)(ii), establishing (7.41).

We now establish (7.42).

$$|\rho_1'\hat{\Theta}_Z\Sigma_{2,N}\rho_2-\rho_1'\Theta_Z\Sigma_{2,N}\rho_2|\leq\|\Sigma_{2,N}\|_\infty\|(\hat{\Theta}_Z'-\Theta_Z')\rho_1\|_1\sqrt{h_2}$$
$$=\|\Sigma_{2,N}\|_\infty O_p(\bar{s}\lambda_{node}\sqrt{h_1h_2})=O\left(1/\sqrt{N}\right)O_p(\bar{s}\lambda_{node}\sqrt{h_1h_2})=o_p(1),$$

where the first equality is due to (7.39), the second equality is due to the definition of $\Sigma_{2,N}$ and (8.1), and the last equality is due to Assumption 6(a)(ii). Thus, we have established (7.42), concluding the proof of that (7.27) is $o_p(1)$.

(7.28) is $o_p(1)$:

We now prove that (7.28) is $o_p(1)$. First,

$$|\rho_2'\hat{\Sigma}_{3,N}\rho_2-\rho_2'\Sigma_{3,N}\rho_2|\leq\left\|\hat{\Sigma}_{3,N}-\Sigma_{3,N}\right\|_\infty h_2\leq h_2\left(\left\|\hat{\Sigma}_{3,N}-\tilde{\Sigma}_{3,N}\right\|_\infty+\left\|\tilde{\Sigma}_{3,N}-\Sigma_{3,N}\right\|_\infty\right),$$

where $\tilde{\Sigma}_{3,N}:=\frac{1}{T}\sum_{i=1}^{N}\sum_{t=1}^{T}\varepsilon_{i,t}^2d_{i,t}d_{i,t}'$. We consider $\left\|\hat{\Sigma}_{3,N}-\tilde{\Sigma}_{3,N}\right\|_\infty$ first.

$$\left\|\hat{\Sigma}_{3,N}-\tilde{\Sigma}_{3,N}\right\|_\infty=\left\|\frac{1}{T}\sum_{i=1}^{N}\sum_{t=1}^{T}\hat{\varepsilon}_{i,t}^2d_{i,t}d_{i,t}'-\frac{1}{T}\sum_{i=1}^{N}\sum_{t=1}^{T}\varepsilon_{i,t}^2d_{i,t}d_{i,t}'\right\|_\infty$$
$$\leq2\left\|\frac{1}{T}\sum_{i=1}^{N}\sum_{t=1}^{T}d_{i,t}d_{i,t}'\varepsilon_{i,t}\pi_{i,t}'(\hat{\gamma}-\gamma)\right\|_\infty+\left\|\frac{1}{T}\sum_{i=1}^{N}\sum_{t=1}^{T}d_{i,t}d_{i,t}'[\pi_{i,t}'(\hat{\gamma}-\gamma)]^2\right\|_\infty. \tag{7.47}$$

Consider the first term of (7.47). A typical element of $\frac{1}{T}\sum_{i=1}^N\sum_{t=1}^T d_{i,t}d_{i,t}'\varepsilon_{i,t}\pi_{i,t}'(\hat{\gamma}-\gamma)$ is

$$\begin{aligned}\frac{1}{T}\sum_{j=1}^{NT} d_{j,l}d_{j,k}\varepsilon_j\pi_j'(\hat{\gamma}-\gamma) &\leq \frac{1}{T}\left(\sum_{j=1}^{NT} d_{j,l}^2 d_{j,k}^2\varepsilon_j^2\right)^{1/2}\left(\sum_{j=1}^{NT}[\pi_j'(\hat{\gamma}-\gamma)]^2\right)^{1/2} \\ &= \left(\frac{1}{T}\sum_{i=1}^N\sum_{t=1}^T d_{i,t,l}^2 d_{i,t,k}^2\varepsilon_{i,t}^2\right)^{1/2}\frac{1}{\sqrt{T}}\left\|\Pi(\hat{\gamma}-\gamma)\right\| = \left(\frac{1}{T}\sum_{t=1}^T\varepsilon_{k,t}^2\right)^{1/2}\frac{1}{\sqrt{T}}\left\|\Pi(\hat{\gamma}-\gamma)\right\|\end{aligned}$$

for some $l,k\in\{1,\dots,N\}$, where the inequality is due to Cauchy-Schwarz inequality. By Assumption 3 we have $\mathbb{P}(|\varepsilon_{i,t}^2|\geq\epsilon)\leq\mathbb{P}(|\varepsilon_{i,t}|\geq\epsilon^{1/2})\leq\frac{1}{2}Ke^{-C\epsilon}$ for every $\epsilon>0$. It follows from Lemma 2.2.1 in van der Vaart and Wellner (1996) that $\|\varepsilon_{i,t}^2\|_{\psi_1}\leq(1+K/2)/C$ for all i and t. Hence, by subadditivity of the Orlicz norm and Jensen's inequality, $\|\varepsilon_{i,t}^2-\mathbb{E}[\varepsilon_{i,t}^2]\|_{\psi_1}\leq 2\|\varepsilon_{i,t}^2\|_{\psi_1}\leq(2+K)/C$. Using the definition of the Orlicz norm, we have $\mathbb{E}\exp(\frac{C}{2+K}|\varepsilon_{i,t}^2-\mathbb{E}[\varepsilon_{i,t}^2]|)\leq 2$. Use independence of $\varepsilon_{i,t}$ across t to invoke Proposition 2 in Appendix B for $D=\frac{C}{2+K}$ and $\alpha=1/3$ to conclude

$$\mathbb{P}\left(\left|\sum_{t=1}^T(\varepsilon_{i,t}^2-\mathbb{E}[\varepsilon_{i,t}^2])\right|\geq T\epsilon\right)\leq Ae^{-B(\epsilon^2T)^{1/3}},$$

for positive constants A and B. Setting $\epsilon=\sqrt{\frac{M(\log N)^3}{T}}$ for some $M>0$ $\left(\epsilon\gtrsim\frac{1}{\sqrt{T}}\right)$, one has

$$\mathbb{P}\left(\max_{1\leq k\leq N}\left|\sum_{t=1}^T(\varepsilon_{k,t}^2-\mathbb{E}[\varepsilon_{k,t}^2])\right|\geq T\epsilon\right)\leq\sum_{k=1}^N\mathbb{P}\left(\left|\sum_{t=1}^T(\varepsilon_{k,t}^2-\mathbb{E}[\varepsilon_{k,t}^2])\right|\geq T\epsilon\right)\leq AN^{1-BM^{1/3}}.$$

The upper bound of the preceding probability becomes arbitrarily small for N and M sufficiently large. Hence,

$$\max_{1\leq k\leq N}\left|\frac{1}{T}\sum_{t=1}^T(\varepsilon_{k,t}^2-\mathbb{E}[\varepsilon_{k,t}^2])\right|=O_p\left(\sqrt{\frac{(\log N)^3}{T}}\right). \tag{7.48}$$

Furthermore, since $\max_{1\leq k\leq N}\max_{1\leq t\leq T}\mathbb{E}[\varepsilon_{k,t}^2]\leq\max_{1\leq k\leq N}\max_{1\leq t\leq T}\|\varepsilon_{k,t}^2\|_{\psi_1}\leq(1+K/2)/C=O(1)$

$$\max_{1\leq k\leq N}\left|\frac{1}{T}\sum_{t=1}^T\varepsilon_{k,t}^2\right|\leq\max_{1\leq k\leq N}\left|\frac{1}{T}\sum_{t=1}^T(\varepsilon_{k,t}^2-\mathbb{E}[\varepsilon_{k,t}^2])\right|+\max_{1\leq k\leq N}\max_{1\leq t\leq T}\mathbb{E}[\varepsilon_{k,t}^2]=O_p\left(\sqrt{\frac{(\log N)^3}{T}}\right)+O(1). \tag{7.49}$$

Therefore,

$$\left\|\frac{1}{T}\sum_{i=1}^N\sum_{t=1}^T d_{i,t}d_{i,t}'\varepsilon_{i,t}\pi_{i,t}'(\hat{\gamma}-\gamma)\right\|_\infty=O_p\left(\frac{(\log N)^{3/4}}{T^{1/4}}\vee 1\right)\frac{1}{\sqrt{T}}\|\Pi(\hat{\gamma}-\gamma)\|. \tag{7.50}$$

Now consider the second term of (7.47). A typical element of $\frac{1}{T}\sum_{i=1}^N\sum_{t=1}^T d_{i,t}d_{i,t}'[\pi_{i,t}'(\hat{\gamma}-\gamma)]^2$ is

$$\frac{1}{T}\sum_{i=1}^N\sum_{t=1}^T d_{i,t,l}d_{i,t,k}[\pi_{i,t}'(\hat{\gamma}-\gamma)]^2\leq\max_{1\leq i\leq N}\max_{1\leq t\leq T}|d_{i,t,l}d_{i,t,k}|\frac{1}{T}\|\Pi(\hat{\gamma}-\gamma)\|^2=\frac{1}{T}\|\Pi(\hat{\gamma}-\gamma)\|^2, \tag{7.51}$$

uniformly over $l,k\in\{1,\dots,N\}$. Combining (7.50) and (7.51), we have

$$\left\|\hat{\Sigma}_{3,N}-\tilde{\Sigma}_{3,N}\right\|_\infty=O_p\left(\frac{(\log N)^{3/4}}{T^{1/4}}\vee 1\right)\frac{1}{\sqrt{T}}\|\Pi(\hat{\gamma}-\gamma)\|+\frac{1}{T}\|\Pi(\hat{\gamma}-\gamma)\|^2. \tag{7.52}$$

Next, consider $\left\|\tilde{\Sigma}_{3,N} - \Sigma_{3,N}\right\|_\infty$.

$$\left\|\tilde{\Sigma}_{3,N} - \Sigma_{3,N}\right\|_\infty = \max_{1\le l\le N}\max_{1\le k\le N}\left|\frac{1}{T}\sum_{i=1}^{N}\sum_{t=1}^{T}(\varepsilon_{i,t}^2 d_{i,t,l}d_{i,t,k} - \mathbb{E}[\varepsilon_{i,t}^2 d_{i,t,l}d_{i,t,k}])\right|$$
$$= \max_{1\le k\le N}\left|\frac{1}{T}\sum_{t=1}^{T}(\varepsilon_{k,t}^2 - \mathbb{E}[\varepsilon_{k,t}^2])\right| = O_p\left(\sqrt{\frac{(\log N)^3}{T}}\right), \tag{7.53}$$

where the last equality is due to (7.48). Summing up (7.52) and (7.53) yields

$$|\rho_2'\hat{\Sigma}_{3,N}\rho_2 - \rho_2'\Sigma_{3,N}\rho_2|$$
$$= h_2 O_p\left(\frac{(\log N)^{3/4}}{T^{1/4}} \vee 1\right)\frac{1}{\sqrt{T}}\|\Pi(\hat{\gamma}-\gamma)\| + h_2\frac{1}{T}\|\Pi(\hat{\gamma}-\gamma)\|^2 + O_p\left(h_2\sqrt{\frac{(\log N)^3}{T}}\right)$$
$$= o_p(1),$$

where the last equality is due to Assumptions 6(b)(v), which, in turns, implies that (7.28) is $o_p(1)$.

Thus, we have proved (7.25). (3.20) then follows trivially since the conclusions of Theorem 1 and Corollary 1 are uniform over the set $\mathcal{F}(s_1, s_2, c_N)$ and the true parameter vector only entered the above arguments when these results were used.

7.4.2 Numerators of t_1 and t_1'

We now show that the numerators of t_1 and t_1' are asymptotically equivalent, i.e.,

$$|\rho'\hat{\Theta}S^{-1}\Pi'\varepsilon - \rho'\Theta S^{-1}\Pi'\varepsilon| = o_p(1). \tag{7.54}$$

Note that

$$|\rho'\hat{\Theta}S^{-1}\Pi'\varepsilon - \rho'\Theta S^{-1}\Pi'\varepsilon| \le \|\rho'(\hat{\Theta}-\Theta)\|_1\|S^{-1}\Pi'\varepsilon\|_\infty = \|\rho_1'(\hat{\Theta}_Z - \Theta_Z)\|_1\|S^{-1}\Pi'\varepsilon\|_\infty$$
$$= O_p(\bar{s}\lambda_{node}\sqrt{h_1})\left(\frac{1}{\sqrt{NT}}\left\|Z'\varepsilon\right\|_\infty \vee \frac{1}{\sqrt{T}}\left\|D'\varepsilon\right\|_\infty\right) = O_p(\bar{s}\lambda_{node}\sqrt{h_1})O_p\left(\sqrt{(\log(p\vee N))^3}\right) = o_p(1),$$

where the second equality is due to (7.39), and the third equality is due to (7.12) and (7.13), and the last equality is due to Assumption 6(a)(iii).

7.4.3 $t_1' \xrightarrow{d} N(0,1)$

We now prove that t_1' is asymptotically distributed as a standard normal by verifying (i)-(iii) of Theorem 5 in Appendix B. Note that

$$t_1' := \frac{\rho'\Theta S^{-1}\Pi'\varepsilon}{\sqrt{\rho'\Theta\Sigma_{\Pi\varepsilon}\Theta'\rho}} = \frac{\rho'\Theta S^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}\begin{pmatrix} z_{i,t}\varepsilon_{i,t} \\ d_{i,t}\varepsilon_{i,t}\end{pmatrix}}{\sqrt{\rho'\Theta\Sigma_{\Pi\varepsilon}\Theta'\rho}} = \frac{\rho'\Theta S^{-1}\sum_{j=1}^{k_N}\begin{pmatrix} z_j\varepsilon_j \\ d_j\varepsilon_j\end{pmatrix}}{\sqrt{\rho'\Theta\Sigma_{\Pi\varepsilon}\Theta'\rho}},$$

where $k_N := NT_N = NT$. In the proof of Lemma 5, we have shown that t_1' is a martingale difference array with variance

$$\operatorname{var}(t_1') = \mathbb{E}[t_1'^2] = \frac{\rho'\Theta S^{-1}\mathbb{E}[\Pi'\varepsilon\varepsilon'\Pi]S^{-1}\Theta'\rho}{\rho'\Theta\Sigma_{\Pi\varepsilon}\Theta'\rho} = 1$$

where we have used the definition of $\Sigma_{\Pi\varepsilon}$. We have already shown in (7.24) that the denominator of t_1' is uniformly bounded away from zero. Thus, verifying that t_1' satisfies (i) and (ii) of

Theorem 5 in Appendix B is equivalent to verifying that the numerator of t_1' satisfies (i) and (ii) of Theorem 5. First, note that

$$\left\|\rho_1'\Theta_Z\right\|_1=\left\|\sum_{j\in H_1}\rho_{1j}\Theta_{Z,j}'\right\|_1\leq\sum_{j\in H_1}|\rho_{1j}|\left\|\Theta_{Z,j}'\right\|_1=O(\sqrt{h_1\bar{s}}),\tag{7.55}$$

where the last equality is due to (7.23). Next,

$$\left|\rho'\Theta S^{-1}\begin{pmatrix} z_{i,t}\varepsilon_{i,t}\\ d_{i,t}\varepsilon_{i,t}\end{pmatrix}\right|\leq\left|\rho_1'\Theta_Z\frac{z_{i,t}\varepsilon_{i,t}}{\sqrt{NT}}\right|+\left|\frac{\rho_{2,i}\varepsilon_{i,t}}{\sqrt{T}}\right|\leq\left\|\rho_1'\Theta_Z\right\|_1\max_{l\in\Xi}\left|\frac{z_{i,t,l}\varepsilon_{i,t}}{\sqrt{NT}}\right|+\frac{\|\rho_2\|_\infty|\varepsilon_{i,t}|}{\sqrt{T}}$$
$$\lesssim\sqrt{h_1\bar{s}}\max_{l\in\Xi}\left|\frac{z_{i,t,l}\varepsilon_{i,t}}{\sqrt{NT}}\right|+\frac{\|\rho_2\|_\infty|\varepsilon_{i,t}|}{\sqrt{T}},$$

for $\Xi:=\bigcup_{j\in H_1}(S_{node,j}\cup\{j\})$, where the last inequality due to (7.55). We have already shown in the proof of Lemma 5 that $z_{i,t,l}\varepsilon_{i,t}$ has uniformly bounded ψ_1-Orlicz norm. The same is the case for $\varepsilon_{i,t}$. Hence,

$$\left\|\sqrt{h_1\bar{s}}\max_{l\in\Xi}\left|\frac{z_{i,t,l}\varepsilon_{i,t}}{\sqrt{NT}}\right|+\frac{\|\rho_2\|_\infty|\varepsilon_{i,t}|}{\sqrt{T}}\right\|_{\psi_1}\leq\sqrt{\frac{h_1\bar{s}}{NT}}\left\|\max_{l\in\Xi}z_{i,t,l}\varepsilon_{i,t}\right\|_{\psi_1}+\frac{\|\rho_2\|_\infty}{\sqrt{T}}\left\|\varepsilon_{i,t}\right\|_{\psi_1}$$
$$\lesssim\sqrt{\frac{h_1\bar{s}}{NT}}\log(1+|\Xi|)\max_{l\in\Xi}\left\|z_{i,t,l}\varepsilon_{i,t}\right\|_{\psi_1}+\frac{\|\rho_2\|_\infty}{\sqrt{T}}\left\|\varepsilon_{i,t}\right\|_{\psi_1}$$
$$\lesssim\sqrt{\frac{h_1\bar{s}}{NT}}\log(1+[h_1(\bar{s}+1)\wedge p])+\frac{\|\rho_2\|_\infty}{\sqrt{T}},$$

for all i and T, where the second inequality is due to Lemma 2.2.2 in van der Vaart and Wellner (1996).[7] Using Lemma 2.2.2 in van der Vaart and Wellner (1996) one more time,

$$\left\|\max_{1\leq i\leq N}\max_{1\leq t\leq T}\left|\rho'\Theta S^{-1}\begin{pmatrix} z_{i,t}\varepsilon_{i,t}\\ d_{i,t}\varepsilon_{i,t}\end{pmatrix}\right|\right\|_{\psi_1}\lesssim\log(1+NT)\left[\sqrt{\frac{h_1\bar{s}}{NT}}\log(1+[h_1(\bar{s}+1)\wedge p])+\frac{\|\rho_2\|_\infty}{\sqrt{T}}\right]=o(1),$$

where the last equality is due to Assumption 6(a)(iv)-(v). Since $\|U\|_{L_r}\leq r!\|U\|_{\psi_1}$ for any random variable U (van der Vaart and Wellner (1996), p95), we conclude that (i) and (ii) of Theorem 5 are satisfied.

We now verify (iii) of Theorem 5. That is,

$$\frac{\sum_{j=1}^{k_N}\left[\rho'\Theta S^{-1}\begin{pmatrix} z_j\varepsilon_j\\ d_j\varepsilon_j\end{pmatrix}\right]^2}{\rho'\Theta\Sigma_{\Pi\varepsilon}\Theta'\rho}=\frac{\rho'\Theta\begin{pmatrix}\tilde{\Sigma}_{1,N} & \tilde{\Sigma}_{2,N}\\ \tilde{\Sigma}_{2,N}' & \tilde{\Sigma}_{3,N}\end{pmatrix}\Theta'\rho}{\rho'\Theta\Sigma_{\Pi\varepsilon}\Theta'\rho}\xrightarrow{p}1.$$

Since we have already shown in (7.24) that the denominator of t_1' is uniformly bounded away from zero, it suffices to show

$$\left|\rho'\Theta\begin{pmatrix}\tilde{\Sigma}_{1,N} & \tilde{\Sigma}_{2,N}\\ \tilde{\Sigma}_{2,N}' & \tilde{\Sigma}_{3,N}\end{pmatrix}\Theta'\rho-\rho'\Theta\Sigma_{\Pi\varepsilon}\Theta'\rho\right|=o_p(1).\tag{7.56}$$

The left-hand side of (7.56) can be bounded by

$$\left|\rho'\Theta\begin{pmatrix}\tilde{\Sigma}_{1,N} & \tilde{\Sigma}_{2,N}\\ \tilde{\Sigma}_{2,N}' & \tilde{\Sigma}_{3,N}\end{pmatrix}\Theta'\rho-\rho'\Theta\Sigma_{\Pi\varepsilon}\Theta'\rho\right|$$
$$\leq|\rho_1'\Theta_Z\tilde{\Sigma}_{1,N}\Theta_Z'\rho_1-\rho_1'\Theta_Z\Sigma_{1,N}\Theta_Z'\rho_1|\tag{7.57}$$
$$+2|\rho_1'\Theta_Z\tilde{\Sigma}_{2,N}\rho_2-\rho_1'\Theta_Z\Sigma_{2,N}\rho_2|\tag{7.58}$$
$$+|\rho_2'\tilde{\Sigma}_{3,N}\rho_2-\rho_2'\Sigma_{3,N}\rho_2|.\tag{7.59}$$

[7] $|\Xi|$ has at most $h_1(\bar{s}+1)$ elements given its definition and is also bounded from above by the dimension of $z_{i,t}$, p.

Thus, we establish that (7.57), (7.58) and (7.59) are $o_p(1)$. Consider (7.57) first.

$$|\rho_1'\Theta_Z\tilde{\Sigma}_{1,N}\Theta_Z'\rho_1 - \rho_1'\Theta_Z\Sigma_{1,N}\Theta_Z'\rho_1| \leq \|\tilde{\Sigma}_{1,N} - \Sigma_{1,N}\|_\infty \|\Theta_Z'\rho_1\|_1^2 = O_p\left(\sqrt{\frac{(\log(p^2T))^5}{N}}\right) O(h_1\bar{s}) = o_p(1)$$

where the first equality is due to (7.55) and (7.38), and the last equality is due to Assumption 6(a)(i). Now consider (7.58).

$$|\rho_1'\Theta_Z\tilde{\Sigma}_{2,N}\rho_2 - \rho_1'\Theta_Z\Sigma_{2,N}\rho_2| \leq \|\tilde{\Sigma}_{2,N} - \Sigma_{2,N}\|_\infty \|\Theta_Z'\rho_1\|_1 \|\rho_2\|_1 = O_p\left(\sqrt{\frac{(\log(pNT))^3 h_1 h_2 \bar{s}}{N}}\right) = o_p(1),$$

where the first equality is due to (7.46), and the last equality is due to Assumption 6(a)(ii). Finally, consider (7.59).

$$|\rho_2'\tilde{\Sigma}_{3,N}\rho_2 - \rho_2'\Sigma_{3,N}\rho_2| \leq \|\tilde{\Sigma}_{3,N} - \Sigma_{3,N}\|_\infty \|\rho_2\|_1^2 = O_p\left(\sqrt{\frac{(\log N)^3}{T}}\right) O(h_2) = o_p(1),$$

where the first equality is due to (7.53), and the last equality is due to Assumption 6(b)(v). Therefore, we have established (7.56) and t_1' is asymptotically standard gaussian.

7.4.4 $t_2 = o_p(1)$

Last, we prove that $t_2 = o_p(1)$. Since the denominator of t_2 is bounded away from zero by a positive constant with probability approaching one by (7.24) and (7.25), it suffices to show $\rho'\Delta = o_p(1)$.

$$\begin{aligned}
|\rho'\Delta| &= \Big|\sum_{j\in H} \rho_j \Delta_j\Big| \leq \sqrt{h} \max_{j\in H} |\Delta_j| \leq \sqrt{h}\|S(\hat{\gamma} - \gamma)\|_1 \max_{j\in H} \big\|\hat{\Theta}_j'\Psi_N - \mathrm{I}_{p+N,j}'\big\|_\infty \\
&= \sqrt{h}\|S(\hat{\gamma} - \gamma)\|_1 \left(\max_{j\in H_1}\left\|\begin{pmatrix} \frac{1}{NT}Z'Z\hat{\Theta}_{Z,j} - e_j \\ \frac{1}{T\sqrt{N}}D'Z\hat{\Theta}_{Z,j}\end{pmatrix}\right\|_\infty \vee \max_{i\in H_2}\left\|\begin{pmatrix}\frac{1}{T\sqrt{N}}Z'De_i \\ 0\end{pmatrix}\right\|_\infty\right) \\
&= \sqrt{h}\|S(\hat{\gamma} - \gamma)\|_1 \left(\max_{j\in H_1}\left(\left\|\frac{1}{NT}Z'Z\hat{\Theta}_{Z,j} - e_j\right\|_\infty \vee \left\|\frac{1}{T\sqrt{N}}D'Z\hat{\Theta}_{Z,j}\right\|_\infty\right) \vee \max_{i\in H_2}\left\|\frac{1}{T\sqrt{N}}Z'D\right\|_\infty\right) \\
&\leq \sqrt{h}\|S(\hat{\gamma} - \gamma)\|_1 \left(\max_{j\in H_1}\left(\left\|\frac{1}{NT}Z'Z\hat{\Theta}_{Z,j} - e_j\right\|_\infty \vee \big\|\hat{\Theta}_{Z,j}\big\|_1\left\|\frac{1}{T\sqrt{N}}D'Z\right\|_\infty\right) \vee \max_{i\in H_2}\left\|\frac{1}{T\sqrt{N}}Z'D\right\|_\infty\right)
\end{aligned}$$

where $\hat{\Theta}_j$ is the jth row of $\hat{\Theta}$ but written as a $(p+N)\times 1$ vector, and $\mathrm{I}_{p+N,j}$ is the jth row of I_{p+N} but written as a $(p+N)\times 1$ vector. Note that

$$\max_{j\in H_1}\left\|\frac{1}{NT}Z'Z\hat{\Theta}_{Z,j} - e_j\right\|_\infty \leq \max_{j\in H_1}\frac{\lambda_{node}}{\hat{\tau}_j^2} = O_p(\lambda_{node}),$$

where the inequality is due to the extended KKT conditions (3.9), and the equality is due to (3.14). Recall that by (7.15) we have that for every $\epsilon > 0$

$$\mathbb{P}\left(\max_{1\leq i\leq N}\max_{1\leq l\leq p}\left|\frac{1}{\sqrt{NT}}\sum_{t=1}^T z_{i,t,l}\right| \geq \epsilon\right) \leq \sum_{i=1}^N\sum_{l=1}^p \mathbb{P}\left(\left|\frac{1}{\sqrt{NT}}\sum_{t=1}^T z_{i,t,l}\right| \geq \epsilon\right) \leq ApNe^{-B\epsilon^2 N},$$

for positive constants A, B. Setting $\epsilon = \sqrt{\frac{M\log(pN)}{N}}$ $(M > 0)$ makes the upper bound of the preceding inequality arbitrarily small for sufficiently large N and M, such that

$$\big\|\hat{\Theta}_{Z,j}\big\|_1\left\|\frac{1}{T\sqrt{N}}D'Z\right\|_\infty = O_p\left(\sqrt{\frac{\bar{s}\log(pN)}{N}}\right).$$

Thus, $|\rho'\Delta| = o_p(1)$ by Assumption 6(c). For later reference,

$$\sup_{\gamma\in\mathcal{F}(s_1,s_2,c_N)} |\rho'\Delta| = o_p(1) \tag{7.60}$$

by the same reasoning leading to the uniform validity of (3.20). □

7.5 Proof of Theorem 3

Proof of Theorem 3. For every $\epsilon > 0$, define

$$A_{1,N} := \left\{ \sup_{\gamma \in \mathcal{F}(s_1,s_2,c_N)} |\rho'\Delta| < \epsilon \right\} \qquad A_{2,N} := \left\{ \sup_{\gamma \in \mathcal{F}(s_1,s_2,c_N)} \left| \frac{\sqrt{\rho'\hat{\Theta}\hat{\Sigma}_{\Pi\varepsilon}\hat{\Theta}'\rho}}{\sqrt{\rho'\Theta\Sigma_{\Pi\varepsilon}\Theta'\rho}} - 1 \right| < \epsilon \right\}$$
$$A_{3,N} := \left\{ |\rho'\hat{\Theta}S^{-1}\Pi'\varepsilon - \rho'\Theta S^{-1}\Pi'\varepsilon| < \epsilon \right\}.$$

By (7.60), (3.20), (7.24) and (7.54), the probabilities of the preceding three events all tend to one. Thus, for every $t \in \mathbb{R}$,

$$\left| \mathbb{P}\left(\frac{\rho' S(\tilde{\gamma} - \gamma)}{\sqrt{\rho'\hat{\Theta}\hat{\Sigma}_{\Pi\varepsilon}\hat{\Theta}'\rho}} \leq t \right) - \Phi(t) \right|$$
$$\leq \left| \mathbb{P}\left(\frac{\rho'\hat{\Theta}S^{-1}\Pi'\varepsilon}{\sqrt{\rho'\hat{\Theta}\hat{\Sigma}_{\Pi\varepsilon}\hat{\Theta}'\rho}} - \frac{\rho'\Delta}{\sqrt{\rho'\hat{\Theta}\hat{\Sigma}_{\Pi\varepsilon}\hat{\Theta}'\rho}} \leq t, A_{1,N}, A_{2,N}, A_{3,N} \right) - \Phi(t) \right| + \mathbb{P}\left(\cup_{i=1}^3 A_{i,N}^c \right).$$

We consider $\mathbb{P}\left(\frac{\rho'\hat{\Theta}S^{-1}\Pi'\varepsilon}{\sqrt{\rho'\hat{\Theta}\hat{\Sigma}_{\Pi\varepsilon}\hat{\Theta}'\rho}} - \frac{\rho'\Delta}{\sqrt{\rho'\hat{\Theta}\hat{\Sigma}_{\Pi\varepsilon}\hat{\Theta}'\rho}} \leq t, A_{1,N}, A_{2,N}, A_{3,N} \right)$ first.

$$\mathbb{P}\left(\frac{\rho'\hat{\Theta}S^{-1}\Pi'\varepsilon}{\sqrt{\rho'\hat{\Theta}\hat{\Sigma}_{\Pi\varepsilon}\hat{\Theta}'\rho}} - \frac{\rho'\Delta}{\sqrt{\rho'\hat{\Theta}\hat{\Sigma}_{\Pi\varepsilon}\hat{\Theta}'\rho}} \leq t, A_{1,N}, A_{2,N}, A_{3,N} \right)$$
$$\leq \mathbb{P}\left(\frac{\rho'\Theta S^{-1}\Pi'\varepsilon}{\sqrt{\rho'\Theta\Sigma_{\Pi\varepsilon}\Theta'\rho}} \leq t(1+\epsilon) + \frac{\epsilon + \epsilon}{\sqrt{\rho'\Theta\Sigma_{\Pi\varepsilon}\Theta'\rho}} \right) \leq \mathbb{P}\left(\frac{\rho'\Theta S^{-1}\Pi'\varepsilon}{\sqrt{\rho'\Theta\Sigma_{\Pi\varepsilon}\Theta'\rho}} \leq t(1+\epsilon) + 2D\epsilon \right)$$

for some positive constant D, where the first and second inequalities are due to the fact that $\rho'\Theta\Sigma_{\Pi\varepsilon}\Theta'\rho$ is uniformly bounded away from zero, see (7.24). Since the last inequality in the above does not depend on γ,

$$\sup_{\gamma \in \mathcal{F}(s_1,s_2,c_N)} \mathbb{P}\left(\frac{\rho'\hat{\Theta}S^{-1}\Pi'\varepsilon}{\sqrt{\rho'\hat{\Theta}\hat{\Sigma}_{\Pi\varepsilon}\hat{\Theta}'\rho}} - \frac{\rho'\Delta}{\sqrt{\rho'\hat{\Theta}\hat{\Sigma}_{\Pi\varepsilon}\hat{\Theta}'\rho}} \leq t, A_{1,N}, A_{2,N}, A_{3,N} \right)$$
$$\leq \mathbb{P}\left(\frac{\rho'\Theta S^{-1}\Pi'\varepsilon}{\sqrt{\rho'\Theta\Sigma_{\Pi\varepsilon}\Theta'\rho}} \leq t(1+\epsilon) + 2D\epsilon \right).$$

By the asymptotic normality of t_1', for N sufficiently large,

$$\sup_{\gamma \in \mathcal{F}(s_1,s_2,c_N)} \mathbb{P}\left(\frac{\rho'\hat{\Theta}S^{-1}\Pi'\varepsilon}{\sqrt{\rho'\hat{\Theta}\hat{\Sigma}_{\Pi\varepsilon}\hat{\Theta}'\rho}} - \frac{\rho'\Delta}{\sqrt{\rho'\hat{\Theta}\hat{\Sigma}_{\Pi\varepsilon}\hat{\Theta}'\rho}} \leq t, A_{1,N}, A_{2,N}, A_{3,N} \right) \leq \Phi(t(1+\epsilon) + 2D\epsilon) + \epsilon.$$

As the above arguments are valid for every $\epsilon > 0$, we can use the continuity of $q \mapsto \Phi(q)$ to conclude that for every $\delta > 0$, one can choose ϵ sufficiently small such that

$$\sup_{\gamma \in \mathcal{F}(s_1,s_2,c_N)} \mathbb{P}\left(\frac{\rho'\hat{\Theta}S^{-1}\Pi'\varepsilon}{\sqrt{\rho'\hat{\Theta}\hat{\Sigma}_{\Pi\varepsilon}\hat{\Theta}'\rho}} - \frac{\rho'\Delta}{\sqrt{\rho'\hat{\Theta}\hat{\Sigma}_{\Pi\varepsilon}\hat{\Theta}'\rho}} \leq t, A_{1,N}, A_{2,N}, A_{3,N} \right) \leq \Phi(t) + \delta + \epsilon. \quad (7.61)$$

We next find a lower bound for $\mathbb{P}\left(\frac{\rho'\hat{\Theta}S^{-1}\Pi'\varepsilon}{\sqrt{\rho'\hat{\Theta}\hat{\Sigma}_{\Pi\varepsilon}\hat{\Theta}'\rho}} - \frac{\rho'\Delta}{\sqrt{\rho'\hat{\Theta}\hat{\Sigma}_{\Pi\varepsilon}\hat{\Theta}'\rho}} \le t, A_{1,N}, A_{2,N}, A_{3,N}\right)$.

$$\begin{aligned}
&\mathbb{P}\left(\frac{\rho'\hat{\Theta}S^{-1}\Pi'\varepsilon}{\sqrt{\rho'\hat{\Theta}\hat{\Sigma}_{\Pi\varepsilon}\hat{\Theta}'\rho}} - \frac{\rho'\Delta}{\sqrt{\rho'\hat{\Theta}\hat{\Sigma}_{\Pi\varepsilon}\hat{\Theta}'\rho}} \le t, A_{1,N}, A_{2,N}, A_{3,N}\right)\\
&\ge \mathbb{P}\left(\frac{\rho'\Theta S^{-1}\Pi'\varepsilon}{\sqrt{\rho'\Theta\Sigma_{\Pi\varepsilon}\Theta'\rho}} \le t(1-\epsilon) - \frac{\epsilon+\epsilon}{\sqrt{\rho'\Theta\Sigma_{\Pi\varepsilon}\Theta'\rho}}, A_{1,N}, A_{2,N}, A_{3,N}\right)\\
&\ge \mathbb{P}\left(\frac{\rho'\Theta S^{-1}\Pi'\varepsilon}{\sqrt{\rho'\Theta\Sigma_{\Pi\varepsilon}\Theta'\rho}} \le t(1-\epsilon) - 2D\epsilon, A_{1,N}, A_{2,N}, A_{3,N}\right)\\
&\ge \mathbb{P}\left(\frac{\rho'\Theta S^{-1}\Pi'\varepsilon}{\sqrt{\rho'\Theta\Sigma_{\Pi\varepsilon}\Theta'\rho}} \le t(1-\epsilon) - 2D\epsilon\right) + \mathbb{P}(\cap_{i=1}^{3} A_{i,N}) - 1
\end{aligned}$$

for some positive constant D, where the first and second inequalities are due to the fact that $\rho'\Theta\Sigma_{\Pi\varepsilon}\Theta'\rho$ is uniformly bounded away from zero, see (7.24). Since the last inequality in the above display does not depend on γ, and $\mathbb{P}(\cap_{i=1}^{3} A_{i,N})$ can be made arbitrarily close to one for sufficiently large N,

$$\begin{aligned}
&\inf_{\gamma\in\mathcal{F}(s_1,s_2,c_N)} \mathbb{P}\left(\frac{\rho'\hat{\Theta}S^{-1}\Pi'\varepsilon}{\sqrt{\rho'\hat{\Theta}\hat{\Sigma}_{\Pi\varepsilon}\hat{\Theta}'\rho}} - \frac{\rho'\Delta}{\sqrt{\rho'\hat{\Theta}\hat{\Sigma}_{\Pi\varepsilon}\hat{\Theta}'\rho}} \le t, A_{1,N}, A_{2,N}, A_{3,N}\right)\\
&\ge \mathbb{P}\left(\frac{\rho'\Theta S^{-1}\Pi'\varepsilon}{\sqrt{\rho'\Theta\Sigma_{\Pi\varepsilon}\Theta'\rho}} \le t(1-\epsilon) - 2D\epsilon\right) - \epsilon.
\end{aligned}$$

By the asymptotic normality of t_1', for N sufficiently large,

$$\inf_{\gamma\in\mathcal{F}(s_1,s_2,c_N)} \mathbb{P}\left(\frac{\rho'\hat{\Theta}S^{-1}\Pi'\varepsilon}{\sqrt{\rho'\hat{\Theta}\hat{\Sigma}_{\Pi\varepsilon}\hat{\Theta}'\rho}} - \frac{\rho'\Delta}{\sqrt{\rho'\hat{\Theta}\hat{\Sigma}_{\Pi\varepsilon}\hat{\Theta}'\rho}} \le t, A_{1,N}, A_{2,N}, A_{3,N}\right) \ge \Phi\left(t(1-\epsilon) - 2D\epsilon\right) - 2\epsilon.$$

As the above arguments are valid for every $\epsilon > 0$, we can use the continuity of $q \mapsto \Phi(q)$ to conclude that for every $\delta > 0$, one can choose ϵ sufficiently small such that

$$\inf_{\gamma\in\mathcal{F}(s_1,s_2,c_N)} \mathbb{P}\left(\frac{\rho'\hat{\Theta}S^{-1}\Pi'\varepsilon}{\sqrt{\rho'\hat{\Theta}\hat{\Sigma}_{\Pi\varepsilon}\hat{\Theta}'\rho}} - \frac{\rho'\Delta}{\sqrt{\rho'\hat{\Theta}\hat{\Sigma}_{\Pi\varepsilon}\hat{\Theta}'\rho}} \le t, A_{1,N}, A_{2,N}, A_{3,N}\right) \ge \Phi\left(t\right) - \delta - 2\epsilon. \quad (7.62)$$

Thus, by (7.61), (7.62) and the fact that $\sup_{\gamma\in\mathcal{F}(s_1,s_2,c_N)} \mathbb{P}(\cup_{i=1}^{3} A_{i,N}^c) = \mathbb{P}(\cup_{i=1}^{3} A_{i,N}^c) = o(1)$, we have proved (4.1) (the uniformity over $t \in \mathbb{R}$ follows from the fact that $\Phi(t)$ is continuous). To see (4.2), note that

$$\begin{aligned}
&\mathbb{P}\left(\alpha_j \notin \left[\tilde{\alpha}_j - z_{1-\delta/2}\frac{\tilde{\sigma}_{\alpha,j}}{\sqrt{NT}}, \tilde{\alpha}_j + z_{1-\delta/2}\frac{\tilde{\sigma}_{\alpha,j}}{\sqrt{NT}}\right]\right) = \mathbb{P}\left(\left|\frac{\sqrt{NT}(\tilde{\alpha}_j - \alpha_j)}{\tilde{\sigma}_{z,j}}\right| > z_{1-\delta/2}\right)\\
&\le 1 - \mathbb{P}\left(\frac{\sqrt{NT}(\tilde{\alpha}_j - \alpha_j)}{\tilde{\sigma}_{z,j}} \le z_{1-\delta/2}\right) + \mathbb{P}\left(\frac{\sqrt{NT}(\tilde{\alpha}_j - \alpha_j)}{\tilde{\sigma}_{z,j}} \le -z_{1-\delta/2}\right).
\end{aligned}$$

Thus, taking the supremum over $\gamma \in \mathcal{F}(s_1, s_2, c_N)$ and letting N tend to infinity yields (4.2) via (4.1). The proof is the same for (4.3). Next, we turn to (4.4).

$$\begin{aligned}
&\sqrt{NT} \sup_{\gamma\in\mathcal{F}(s_1,s_2,c_N)} \operatorname{diam}\left(\left[\tilde{\alpha}_j - z_{1-\delta/2}\frac{\tilde{\sigma}_{\alpha,j}}{\sqrt{NT}}, \tilde{\alpha}_j + z_{1-\delta/2}\frac{\tilde{\sigma}_{\alpha,j}}{\sqrt{NT}}\right]\right)\\
&= 2z_{1-\delta/2}\left(\sqrt{[\Theta_Z\Sigma_{1,N}\Theta_Z]_{jj}} + o_p(1)\right) \le 2z_{1-\delta/2}\left(\frac{\sqrt{\text{maxeval}(\Sigma_{1,N})}}{\text{mineval}(\Psi_Z)} + o_p(1)\right) = O_p(1),
\end{aligned}$$

where the first equality is due to (3.20), and the last equality is due to Assumptions 4(a) and 6(d). Similarly, we can prove (4.5):

$$\sqrt{T} \sup_{\gamma \in \mathcal{F}(s_1, s_2, c_N)} \operatorname{diam}\left(\left[\tilde{\eta}_i - z_{1-\delta/2}\frac{\tilde{\sigma}_{\eta,i}}{\sqrt{T}}, \tilde{\eta}_i + z_{1-\delta/2}\frac{\tilde{\sigma}_{\eta,i}}{\sqrt{T}}\right]\right) = 2z_{1-\delta/2}\left(\sqrt{[\Sigma_{3,N}]_{ii}} + o_p(1)\right)$$

$$= 2z_{1-\delta/2}\left(\left[\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}[\varepsilon_{i,t}^2]\right]^{1/2} + o_p(1)\right) = O_p(1),$$

where the third equality follows from the arguments above (7.49). □

8 Appendix B

Proposition 1. *Let A and B be two positive semidefinite $(p-1)\times(p-1)$ matrices and $\delta := \max_{1\leq l,k\leq p-1}|A_{lk} - B_{lk}|$. For some integer $r \in \{1, \ldots, p-1\}$, one has*

$$\kappa^2(B, r) \geq \kappa^2(A, r) - \delta 16 r.$$

Proof. The proof is exactly the same as that of Lemma 6. □

Theorem 4 (Fan et al. (2012)). *Let $\alpha \in (0,1)$. Assume that $(X_i, \mathcal{F}_i)_{i=1}^n$ is a sequence of supermartingale differences satisfying $\sup_i \mathbb{E}[e^{|X_i|^{\frac{2\alpha}{1-\alpha}}}] \leq C_1$ for some constant $C_1 \in (0, \infty)$. Define $S_k := \sum_{i=1}^k X_i$. Then, for all $\epsilon > 0$,*

$$\mathbb{P}\left(\max_{1\leq k\leq n} S_k \geq n\epsilon\right) \leq C(\alpha, n, \epsilon)e^{-\left(\epsilon/4\right)^{2\alpha} n^{\alpha}},$$

where

$$C(\alpha, n, \epsilon) := 2 + 35C_1\left[\frac{1}{16^{1-\alpha}(n\epsilon^2)^{\alpha}} + \frac{1}{n\epsilon^2}\left(\frac{3(1-\alpha)}{2\alpha}\right)^{\frac{1-\alpha}{\alpha}}\right].$$

The preceding theorem is not exactly the same as Theorem 2.1 in Fan et al. (2012), but taken from the proof of Theorem 2.1 in Fan et al. (2012). This theorem generalises Theorem 3.2 in Lesigne and Volny (2001).

Proposition 2. *Let $\alpha \in (0,1)$. Assume that $(X_i, \mathcal{F}_i)_{i=1}^n$ is a sequence of martingale differences satisfying satisfying $\sup_i \mathbb{E}[e^{D|X_i|^{\frac{2\alpha}{1-\alpha}}}] \leq C_1$ for some positive constant D. (C_1 could change with the sample size n.) Then, for all $\epsilon \gtrsim \frac{1}{\sqrt{n}}$,*

$$\mathbb{P}\left(\left|\sum_{i=1}^n X_i\right| \geq n\epsilon\right) \leq AC_1 e^{-K(\epsilon^2 n)^{\alpha}},$$

for positive constants A and K.

Proof. This proposition is a simple adaptation of preceding theorem. Note that for some positive constant D,

$$\mathbb{P}\left(\sum_{i=1}^n X_i \geq n\epsilon\right) = \mathbb{P}\left(\sum_{i=1}^n D^{\frac{1-\alpha}{2\alpha}} X_i \geq nD^{\frac{1-\alpha}{2\alpha}}\epsilon\right) = \mathbb{P}\left(\sum_{i=1}^n Y_i \geq n\delta\right),$$

where $Y_i := D^{\frac{1-\alpha}{2\alpha}} X_i$ and $\delta := D^{\frac{1-\alpha}{2\alpha}}\epsilon$. Now $(Y_i)_{i=1}^n$ is a sequence of martingale differences satisfying $\sup_i \mathbb{E}[e^{|Y_i|^{\frac{2\alpha}{1-\alpha}}}] \leq C_1$. Invoking the preceding theorem, we have

$$\mathbb{P}\left(\sum_{i=1}^n Y_i \geq n\delta\right) \leq C(\alpha, n, \delta)e^{-\left(\delta/4\right)^{2\alpha} n^{\alpha}}.$$

$(-Y_i)_{i=1}^n$ is also a sequence of martingale differences satisfying the same exponential moment condition. Thus,

$$\mathbb{P}\left(\Big|\sum_{i=1}^n X_i\Big| \geq n\epsilon\right) = \mathbb{P}\left(\Big|\sum_{i=1}^n Y_i\Big| \geq n\delta\right) \leq 2C(\alpha,n,\delta)e^{-(\delta/4)^{2\alpha}n^\alpha}$$
$$= 2C(\alpha,n,D^{\frac{1-\alpha}{2\alpha}}\epsilon)e^{-(D^{\frac{1-\alpha}{2\alpha}}\epsilon/4)^{2\alpha}n^\alpha} \leq AC_1e^{-K\epsilon^{2\alpha}n^\alpha},$$

for positive constants A, K, where the last inequality used that if $\epsilon \gtrsim \frac{1}{\sqrt{n}}$ then $2C(\alpha,n,D^{\frac{1-\alpha}{2\alpha}}\epsilon) \leq AC_1$ for some positive constant A. □

Proposition 3. *Suppose we have random variables $Z_{l,i,t,j}$ uniformly subgaussian for $l = 1,\dots,L$ ($L \geq 2$ fixed), $i = 1,\dots,N$, $t = 1,\dots,T$ and $j = 1,\dots,p$. Both p and T increase with N (functions of N). Z_{l_1,i_1,t_1,j_1} and Z_{l_2,i_2,t_2,j_2} are independent as long as $i_1 \neq i_2$ regardless of the values of other subscripts. Then,*

$$\max_{1\leq j\leq p}\max_{1\leq t\leq T}\max_{1\leq i\leq N}\mathbb{E}\Big[\prod_{l=1}^L Z_{l,i,t,j}\Big] \leq A = O(1), \tag{8.1}$$

for some positive constant A and

$$\max_{1\leq j\leq p}\Big|\frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T\Big(\prod_{l=1}^L Z_{l,i,t,j} - \mathbb{E}\Big[\prod_{l=1}^L Z_{l,i,t,j}\Big]\Big)\Big| = O_p\Big(\sqrt{\frac{(\log(pT))^{L+1}}{N}}\Big). \tag{8.2}$$

Proof. For every $\epsilon \geq 0$, $\mathbb{P}\left(|\prod_{l=1}^L Z_{l,i,t,j}| \geq \epsilon\right) \leq \sum_{l=1}^L \mathbb{P}\left(|Z_{l,i,t,j}| \geq \epsilon^{1/L}\right) \leq L\frac{K}{2}e^{-C\epsilon^{2/L}}$ for positive constants K, C. Next, using Hölder's inequaliy, we have

$$\max_{1\leq j\leq p}\max_{1\leq t\leq T}\max_{1\leq i\leq N}\mathbb{E}\Big[\prod_{l=1}^L Z_{l,i,t,j}\Big] \leq \max_{1\leq j\leq p}\max_{1\leq t\leq T}\max_{1\leq i\leq N}\prod_{l=1}^L\Big(\mathbb{E}\Big[|Z_{l,i,t,j}|^L\Big]\Big)^{\frac{1}{L}}.$$

Uniform subgaussianity implies that $\Big(\mathbb{E}\Big[|Z_{l,i,t,j}|^L\Big]\Big)^{\frac{1}{L}}$ is uniformly bounded. That is, $\Big(\mathbb{E}\Big[|Z_{l,i,t,j}|^L\Big]\Big)^{\frac{1}{L}} \leq L!\|Z_{l,i,t,j}\|_{\psi_1} \leq L!(\log 2)^{-1/2}\|Z_{l,i,t,j}\|_{\psi_2} \leq L!(\log 2)^{-1/2}\left(\frac{1+K/2}{C}\right)^{1/2}$, where the first two inequalities are taken from p95 of van der Vaart and Wellner (1996), and the third inequality is due to Lemma 2.2.1 in van der Vaart and Wellner (1996). (8.1) then follows.

For every $\epsilon \geq 0$,

$$\mathbb{P}\Big(\Big|\frac{1}{T}\sum_{t=1}^T\Big(\prod_{l=1}^L Z_{l,i,t,j} - \mathbb{E}\Big[\prod_{l=1}^L Z_{l,i,t,j}\Big]\Big)\Big| \geq \epsilon\Big) \leq \mathbb{P}\Big(\max_{1\leq t\leq T}\Big|\prod_{l=1}^L Z_{l,i,t,j}\Big| \geq \epsilon - A\Big)$$
$$\leq \sum_{t=1}^T \mathbb{P}\Big(\Big|\prod_{l=1}^L Z_{l,i,t,j}\Big| \geq \epsilon - A\wedge\epsilon\Big) \leq \frac{L}{2}TKe^{-C(\epsilon-A\wedge\epsilon)^{2/L}} \leq \frac{L}{2}TKe^{-C[\epsilon^{2/L}-(A\wedge\epsilon)^{2/L}]} \leq TK'e^{-C\epsilon^{2/L}},$$

for $K' = \frac{L}{2}Ke^{CA^{2/L}}$ and where the second to last inequality is due to subadditivity of $x \mapsto x^{2/L}$ such that $x^{2/L} \leq y^{2/L}+(x-y)^{2/L}$ for $x \geq y \geq 0$, $L \geq 2$. Let $X_{i,j}$ denote $\frac{1}{T}\sum_{t=1}^T\left(\prod_{l=1}^L Z_{l,i,t,j} - \mathbb{E}[\prod_{l=1}^L Z_{l,i,t,j}]\right)$. Consider some positive constant $D < C$.

$$\mathbb{E}\Big[e^{D|X_{i,j}|^{2/L}}\Big] = \int_{x\in\mathbb{R}}\int_0^{|x|^{2/L}} De^{Ds}dsP(dx) + 1 = \int_0^\infty De^{Ds}\mathbb{P}(|X_{i,j}| > s^{L/2})ds + 1$$
$$\leq \int_0^\infty TK'De^{(D-C)s}ds + 1 = \frac{TK'D}{C-D} + 1 \leq BT, \tag{8.3}$$

for some positive constant B, where the second equality is by Fubini's theorem. Then we can use independence across i to invoke Proposition 2 in Appendix B with $\alpha = \frac{1}{L+1}$ and $C_1 = BT$, for $\epsilon \gtrsim \frac{1}{\sqrt{N}}$,

$$\mathbb{P}\Big(\Big|\sum_{i=1}^{N}\frac{1}{T}\sum_{t=1}^{T}\Big(\prod_{l=1}^{L} Z_{l,i,t,j} - \mathbb{E}\Big[\prod_{l=1}^{L} Z_{l,i,t,j}\Big]\Big)\Big| \geq N\epsilon\Big) \leq A'Te^{-K\left(\epsilon^2 N\right)^{\frac{1}{L+1}}}$$

for positive constants A' and K. Setting $\epsilon = \sqrt{\frac{M(\log(pT))^{L+1}}{N}}$ $\left(\gtrsim \frac{1}{\sqrt{N}}\right)$ for some $M > 0$, we have

$$\mathbb{P}\Big(\max_{1\leq l\leq p}\Big|\sum_{i=1}^{N}\frac{1}{T}\sum_{t=1}^{T}\Big(\prod_{l=1}^{L} Z_{l,i,t,j} - \mathbb{E}\Big[\prod_{l=1}^{L} Z_{l,i,t,j}\Big]\Big)\Big| \geq N\epsilon\Big) \leq pA'Te^{-K\left(\epsilon^2 N\right)^{\frac{1}{L+1}}} = A'(pT)^{1-KM^{\frac{1}{L+1}}}.$$

The upper bound of the preceding probability becomes arbitrarily small for N and M sufficiently large. Hence (8.2) follows. □

Lemma 9. *Let A be a symmetric $p \times p$ matrix, and $\hat{v}$ and $v \in \mathbb{R}^p$. Then*

$$|\hat{v}'A\hat{v} - v'Av| \leq \|A\|_\infty \|\hat{v} - v\|_1^2 + 2\|Av\|\|\hat{v} - v\|.$$

Proof. See Lemma 6.1 in the working-paper version of van de Geer et al. (2014). □

Theorem 5 (McLeish (1974)). *Let $\{X_{n,i}, i = 1, ..., k_n\}$ be a martingale difference array with respect to the triangular array of σ-algebras $\{\mathcal{F}_{n,i}, i = 0, ..., k_n\}$ (i.e., $X_{n,i}$ is $\mathcal{F}_{n,i}$-measurable and $\mathbb{E}[X_{n,i}|\mathcal{F}_{n,i-1}] = 0$ almost surely for all n and i) satisfying $\mathcal{F}_{n,i-1} \subseteq \mathcal{F}_{n,i}$ for all $n \geq 1$. Assume,*

(i) $\max_{i\leq k_n} |X_{n,i}|$ *is uniformly bounded in L_2 norm,*

(ii) $\max_{i\leq k_n} |X_{n,i}| \xrightarrow{p} 0$*, and*

(iii) $\sum_{i=1}^{k_n} X_{n,i}^2 \xrightarrow{p} 1$.

Then, $S_n = \sum_{i=1}^{k_n} X_{n,i} \xrightarrow{d} N(0,1)$ as $n \to \infty$.

References

Arellano, M. (2003). Panel data econometrics. *OUP Catalogue*.

Baltagi, B. (2008). *Econometric analysis of panel data*, Volume 1. John Wiley & Sons.

Belloni, A., D. Chen, V. Chernozhukov, and C. Hansen (2012). Sparse models and methods for optimal instruments with an application to eminent domain. *Econometrica 80*(6), 2369–2429.

Belloni, A., V. Chernozhukov, and C. Hansen (2014). Inference on treatment effects after selection among high-dimensional controls. *The Review of Economic Studies 81*(2), 608–650.

Belloni, A., V. Chernozhukov, C. Hansen, and D. Kozbur (2014). Inference in high dimensional panel models with an application to gun control. *arXiv preprint arXiv:1411.6507*.

Bickel, P. J., Y. Ritov, and A. B. Tysbakov (2009). Simultaneous analysis of lasso and dantzig selector. *The Annals of Statistics 37*(4), 1705–1732.

Bonhomme, S. and E. Manresa (2012). Grouped patterns of heterogeneity in panel data. Technical report, Citeseer.

Bunea, F., A. Tsybakov, M. Wegkamp, et al. (2007). Sparsity oracle inequalities for the lasso. *Electronic Journal of Statistics 1*, 169–194.

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 A. B. Kock (2014). Asymptotically honest confidence regions for high dimensional parameters by the desparsified conservative lasso. *arXiv preprint arXiv:1410.4208*.

Caner, M. and H. H. Zhang (2014). Adaptive elastic net for generalized methods of moments. *Journal of Business & Economic Statistics 32*(1), 30–47.

Davidson, J. (2000). *Econometric Theory.* Blackwell Publishers.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. *Econometrica: Journal of the Econometric Society*, 987–1007.

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, X., I. Grama, and Q. Liu (2012). Large deviation exponential inequalities for supermartingales. *Electronic Communications in Probability 17*(59), 1–8.

Horn, R. A. and C. R. Johnson (1990). *Matrix analysis.* Cambridge university press.

Huang, J., J. L. Horowitz, and S. Ma (2008). Asymptotic properties of bridge estimators in sparse high-dimensional regression models. *The Annals of Statistics*, 587–613.

Islam, N. (1995). Growth empirics: a panel data approach. *The Quarterly Journal of Economics*, 1127–1170.

Javanmard, A. and A. Montanari (2013). Confidence intervals and hypothesis testing for high-dimensional regression. *arXiv preprint arXiv:1306.3171*.

Knight, K. and W. Fu (2000). Asymptotics for Lasso-type estimators. *The Annals of Statistics 28*(5), 1356–1378.

Kock, A. B. (2013). Oracle inequalities for high-dimensional panel data models. *Submitted*.

Leeb, H. and B. M. Pötscher (2005). Model selection and inference: Facts and fiction. *Econometric Theory 21*, 21–59.

Lesigne, E. and D. Volny (2001). Large deviations for martingales. *Stochastic Processes and Their Applications 96*(1), 143–159.

Li, K.-C. (1989). Honest confidence regions for nonparametric regression. *The Annals of Statistics*, 1001–1008.

Lockhart, R., J. Taylor, R. J. Tibshirani, and R. Tibshirani (2014). A significance test for the lasso. *The Annals of Statistics 42*(2), 413–468.

McLeish, D. L. (1974). Dependent central limit theorems and invariance principles. *The Annals of Probability 2*(4), 620–628.

Meinshausen, N. (2013). Assumption-free confidence intervals for groups of variables in sparse high-dimensional regression. *arXiv preprint arXiv:1309.3489*.

Meinshausen, N. and B. Yu (2009). Lasso-type recovery of sparse representations for high-dimensional data. *The Annals of Statistics*, 246–270.

Negahban, S. N., P. Ravikumar, M. J. Wainwright, and B. Yu (2012). A unified framework for high-dimensional analysis of m-estimators with decomposable regularizers. *Statistical Science 27*(4), 538–557.

Nickl, R. and S. van de Geer (2013). Confidence sets in sparse regression. *The Annals of Statistics 41*(6), 2852–2876.

Pötscher, B. M. (2009). Confidence sets based on sparse estimators are necessarily large. *Sankhyā: The Indian Journal of Statistics, Series A (2008-)*, 1–18.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society Series B 58*, 267–288.

van de Geer, S., P. Buhlmann, Y. Ritov, and R. Dezeure (2014). On asymptotically optimal confidence regions and tests for high-dimensional models. *The Annals of Statistics 42*(3), 1166–1202.

van de Geer, S. A. (2008). High-dimensional generalized linear models and the lasso. *The Annals of Statistics*, 614–645.

van de Geer, S. A., P. Bühlmann, et al. (2009). On the conditions used to prove oracle results for the lasso. *Electronic Journal of Statistics 3*, 1360–1392.

van der Vaart, A. and J. A. Wellner (1996). *Weak Convergence and Empirical Processes.* Springer.

Wooldridge, J. M. (2010). *Econometric analysis of cross section and panel data.* MIT press.

Yuan, M. (2010). High dimensional inverse covariance matrix estimation via linear programming. *Journal of Machine Learning Research*.

Zhang, C.-H. and J. Huang (2008). The sparsity and bias of the lasso selection in high-dimensional linear regression. *The Annals of Statistics*, 1567–1594.

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.

Zou, H. (2006). The adaptive LASSO and its oracle properties. *Journal of the American Statistical Association 101*(476), 1418–1429.

Research Papers 2013

2014-41: Laurent Callot and Johannes Tang Kristensen: Vector Autoregressions with Parsimoniously Time Varying Parameters and an Application to Monetary Policy

2014-42: Laurent A. F. Callot, Anders B. Kock and Marcelo C. Medeiros: Estimation and Forecasting of Large Realized Covariance Matrices and Portfolio Choice

2014-43: Paolo Santucci de Magistris and Federico Carlini: On the identification of fractionally cointegrated VAR models with the F(d) condition

2014-44: Laurent Callot, Niels Haldrup and Malene Kallestrup Lamb: Deterministic and stochastic trends in the Lee-Carter mortality model

2014-45: Nektarios Aslanidis, Charlotte Christiansen, Neophytos Lambertides and Christos S. Savva: Idiosyncratic Volatility Puzzle: Infl¤uence of Macro-Finance Factors

2014-46: Alessandro Giovannelli and Tommaso Proietti: On the Selection of Common Factors for Macroeconomic Forecasting

2014-47: Martin M. Andreasen and Andrew Meldrum: Dynamic term structure models: The best way to enforce the zero lower bound

2014-48: Tim Bollerslev, Sophia Zhengzi Li and Viktor Todorov: Roughing up Beta: Continuous vs. Discontinuous Betas, and the Cross-Section of Expected Stock Returns

2914-49: Tim Bollerslev, Viktor Todorov and Lai Xu: Tail Risk Premia and Return Predictability

2014-50: Kerstin Gärtner and Mark Podolskij: On non-standard limits of Brownian semi-stationary

2014-51: Mark Podolskij : Ambit fields: survey and new challenges

2014-52: Tobias Fissler and Mark Podolskij: Testing the maximal rank of the volatility process for continuous diffusions observed with noise

2014-53: Cristina M. Scherrer: Cross listing: price discovery dynamics and exchange rate effects

2014-54: Claudio Heinrich and Mark Podolskij: On spectral distribution of high dimensional covariation matrices

2014-55: Gustavo Fruet Dias and Fotis Papailias: Forecasting Long Memory Series Subject to Structural Change: A Two-Stage Approach

2014-56: Torben G. Andersen, Nicola Fusari and Viktor Todorov: The Risk Premia Embedded in Index Options

2014-57: Eduardo Rossi and Paolo Santucci de Magistris: Indirect inference with time series observed with error

2014-58: Anders Bredahl Kock and Haihan Tang: Inference in High-dimensional Dynamic Panel Data Models