High Dimensional Generalised Penalised Least Squares

Ilias Chronopoulos *[1], Katerina Chrysikou † [2] and George Kapetanios ‡[3]

[1]Essex Business School, University of Essex

[2]School of Business, University of Leicester

[3] King's Business School, King's College London

March 2025

Abstract

In this paper we extend the GLS estimator in high dimensional linear models with serially correlated errors and explore its asymptotic properties. We find that the LASSO estimator is consistent under the assumption of α-mixing imposed on the covariate and error processes. However, upon debiasing, is inefficient compared with the proposed feasible DEBIASED GLS LASSO estimator. We show that the feasible DEBIASED GLS LASSO leads to uniformly valid inference in the presence of serial error autocorrelation. Monte Carlo results support the proposed estimator and illustrate significant efficiency gains over the (DEBIASED) LASSO. Finally, an empirical application supports our theoretical results and showcases the superiority of the feasible GLS LASSO in terms of forecasting efficiency against the LASSO and other estimators.

JEL classification: C01, C22, C55

Keywords: Generalised least squares, LASSO, autocorrelation, time series, central limit theorem

*ilias.chronopoulos@essex.ac.uk

†ac988@leicester.ac.uk . Corresponding author

‡george.kapetanios@kcl.ac.uk

1 Introduction

Research in high-dimensional statistics and econometrics has witnessed a surge because the dimensionality of available datasets, models and associated parameter spaces has grown massively, in relation to the sample size. After the seminal work of Tibshirani (1996), the LASSO has become a focus of this continuously growing literature, since it conducts simultaneously model estimation and selection. More recent work establishes the asymptotic behaviour of the LASSO estimator as well as its model selection consistency, known as the oracle property, see, for example, Meinshausen and Yu (2009), Raskutti et al. (2010), Van De Geer and Bühlmann (2009), Van de Geer (2008) and Zhang and Huang (2008).

Although the majority of the theoretical work is centred around the asymptotic behaviour of the LASSO and relevant estimators in generalised linear models, the main interest still lies in simple designs, such as i.i.d. or fixed covariates, while limited work has been done towards the direction of regularised linear models with time series components. Early work of Wang et al. (2007) suggests a linear model with autoregressive error terms and their resulting LASSO estimator satisfies a Knight and Fu (2000)-type asymptotic property, where the number of covariates cannot be larger than the sample size. Hsu et al. (2008) consider the LASSO estimator under a Vector Autoregressive (VAR) process and derive their asymptotic results following the same setting as the former. Various other papers derive asymptotic results ensuring effective model selection for regularised estimators: Nardi and Rinaldo (2011) consider autoregressive structure on the covariates, Basu and Michailidis (2015) consider stochastic regressions and transition matrix estimation in VAR models, and Kock and Callot (2015) consider models with stationary VAR covariates.

Further, indicative papers of research in high-dimensional econometrics which allow for more relaxed assumptions are: Kock (2016a) shows that the adaptive LASSO is oracle efficient in stationary and non-stationary autoregressions, Masini et al. (2022) consider linear time-series models with non-Gaussian errors, and Wong et al. (2020) consider sparse Gaussian VAR models exploring the efficiency of β-mixing assumptions, used to bound the prediction and estimation error of sub-Weibull covariates. Another class of papers focuses entirely on high-dimensional financial econometrics: Babii et al. (2023) and Babii et al. (2022) consider prediction and now-casting with panel data and high-dimensional time series respectively, sampled at different frequencies using the sparse-group LASSO.

Although, the LASSO provides an efficient avenue to estimation and variable selection in high-dimensional datasets, one is unable to perform inference on the estimated parameters. Leeb and Pötscher (2005) have proven that ignoring the model selection step in the LASSO, leads to invalid uniform inference. Recent developments hinge on *post-selection* inference, see, for example, Berk et al. (2013) and Taylor and Tibshirani (2015), among others. The incentive is that *post-selection* methods provide valid inference on the non-zero coefficients obtained after model selection, which is typically carried out at a first step, using the LASSO. Although post-LASSO inference guarantees valid confidence intervals, it is subject to the model selection made prior to that step, which can be misleading when the conditions for a "well-behaved" model are not met, e.g. i.i.d. errors and/or covariates. The latter facilitates the necessity for post-LASSO inference to allow for more relaxed assumptions, as well as uniformity in the limit theory for general penalised models. A way to relax the i.i.d./Gaussianity assumption has been proposed by Tian and Taylor (2017) and Tibshirani et al. (2018), who consider a bootstrap approach for asymptotically valid testing. However, i.i.d. conditions on the covariates and errors need to be assumed in order examine the large sample properties of the method.

Alternative approaches that allow inference on the true parameters without the limitation of a "well-behaved" model, have been developed. These are based on DEBIASED or *desparsified* versions of the LASSO, see, for example, Javanmard and Montanari (2014), Van de Geer et al. (2014), Zhang and Zhang (2014) and on certain assumptions on the covariance matrix, such as fixed design, i.i.d-ness or sub-Gaussianity. Further, extensions to a time series framework have been introduced in the literature: Chernozhukov et al. (2021) consider the covariates and error terms of a high-dimensional model to be temporally and cross-sectionally dependent and apply bootstrap methods towards estimation and inference of the true parameters. Babii et al. (2020) use DEBIASED sparse-group LASSO for inference in a lower dimensional group of parameters. Kock (2016b) studies high-dimensional correlated random effects panel data models, where they allow for correlation between time invariant covariates and unobserved heterogeneity, as under fixed effects. Finally, Adamek et al. (2023) extend the *desparsified* LASSO to a time series setting, using near-epoch dependence assumptions, allowing for non-Gaussian, serially correlated and heteroscedastic processes. Notice though that throughout the literature, no link has been made on the use of Generalised Least Squares (GLS) type of inference to account for non-spherical errors, which is the main focus of this paper.

Inference on Ridge regression instead of the LASSO has been considered by Zhang and Politis (2022), where they propose a wild bootstrap algorithm to construct confidence regions and perform hypothesis testing for a linear combination of parameters. A similar approach has been considered by Zhang and Politis (2023), allowing for inference under non-stationary and heteroscedastic errors.

In this paper, we contribute to the ongoing literature of high-dimensional inference under general conditions. We propose a penalised GLS-type estimator which utilizes estimated autocovariances of the residual in a linear regression, where the residual is allowed to follow a general autoregressive process. LASSO works as a preliminary estimator and is shown to be asymptotically consistent under mild assumptions on the covariates and error processes. We perform uniform inference via the DEBIASED GLS LASSO, imposing mild restrictions on the autocorrelation of the error term. In addition, we relax assumptions on the error and most importantly the covariates processes, commonly used in the existing literature, for example fixed design, i.i.d., sub-Gaussianity, by allowing them to be stationary α-mixing processes.

The remainder of this paper is organised as follows. Sections 2–4 present the model, proposed methodology and theoretical results. Section 5 describes the regularisation parameter tuning, Section 6 contains the Simulation Study, and Section 7 presents an empirical application. Section 8 concludes. Proofs and additional simulation results are relegated to the Supplement.

Notation

For any vector $\boldsymbol{x} \in \mathbb{R}^n$, we denote the ℓ_p-, ℓ_∞- and ℓ_0- norms, as $\|\boldsymbol{x}\|_p = \left(\sum_{i=1}^n |x_i|^p\right)^{1/p}$, $\|\boldsymbol{x}\|_\infty = \max_{i=1,\ldots,n} |x_i|$, $\|\boldsymbol{x}\|_0 = \sum_{i=1}^n \mathbf{1}\{x_i \neq 0\} = \text{supp}(\boldsymbol{x})$, respectively, and $\text{sign}(\boldsymbol{x})$ denotes the sign function applied element-wise on $\boldsymbol{x}$. We use "$\rightarrow_P$" to denote convergence in probability. For two deterministic sequences a_n and b_n we define asymptotic proportionality, "$\asymp$", by writing $a_n \asymp b_n$ if there exist constants $0 < c_1 \leq c_2$ such that $c_1 b_n \leq a_n \leq c_2 b_n$ for all $n \geq 1$. For any set A, $|A|$ denotes its cardinality, while A^c denotes its complement. For $a,\ b \in \mathbb{R}$, we write $a \vee b = \max\{a, b\}$ and $a \wedge b = \min\{a, b\}$. For a real number a, $\lfloor a \rfloor$ denotes the largest integer no greater than a. For a symmetric matrix $\boldsymbol{B}$, we denote its minimum and maximum eigenvalues by $\Lambda_{\min}(\boldsymbol{B})$ and $\Lambda_{\max}(\boldsymbol{B})$ respectively.

2 Theoretical considerations

We consider the following high-dimensional linear regression model,

$$y_t = \boldsymbol{x}_t'\boldsymbol{\beta} + u_t, \tag{1}$$

$$u_t = \sum_{j=1}^{q} \phi_j u_{t-j} + \varepsilon_t, \quad t = q+1, \ldots, T, \tag{2}$$

where $\{\boldsymbol{x}_t\} = \left\{(x_{t,1}, \ldots, x_{t,p})'\right\}$ is a p-dimensional vector-valued stationary time series of predictors, that may include the intercept and $q < \infty$. We provide the following definition and further proceed with assumptions on $\{\boldsymbol{x}_t\}$, $\{\varepsilon_t\}$, and $\{u_t\}$.

Definition 1. *Consider a sequence of r.v.s', $\{z_t\}$ such that $z_t - Ez_t$ is a stationary, ergodic α-mixing sequence with mixing coefficients $\alpha_k \leq c\rho^k$, $k \geq 1$, for some $0 < \rho < 1$ and $c > 0$. In addition, assume that the r.v.'s, $\{z_t\}$ have thin- or heavy-tailed distributions.*

1. *We write $\{z_t\} \in \mathcal{T}(r), r > 0$ to denote a thin-tailed distribution for z_t, for some $d, r > 0$, and $E\left[\exp\left(d\,|z_t|^r\right)\right] < \infty$*
2. *We write $\{z_t\} \in \mathcal{H}(\theta), \theta > 4$ to denote a heavy-tailed distribution for z_t and $E\,|z_t|^\theta < \infty$.*

The definitions in 1 and 2 imply different properties for $\{z_t\}$, such that for $c_0, c_1, \xi > 0$ we write

$$P\left(|z_t| \geq \xi\right) \leq \begin{cases} c_0 \exp\left(-c_1 \xi^r\right), & \text{if } z_t \in \mathcal{T}(r), r > 0, \\ c_0\, \xi^{-\theta}, & \text{if } z_t \in \mathcal{H}(\theta), \theta > 4. \end{cases} \tag{3}$$

Exponential inequalities for sums of unbounded r.v.'s, z_t with thin- and heavy-tailed distributions can be found in Lemma 1 of Dendramis et al. (2021).

Assumption 1. *Let $\{\varepsilon_t\}$ be a sequence of independent and identically (i.i.d) random variables, with $E(\varepsilon_t) = 0$, that satisfies the conditions in Theorem 14.9 of Davidson (1994), with probability density function (p.d.f), f_ε, that exists and is continuous almost everywhere. f_ε can either have thin or heavy tails, which would imply different properties for $P(|\varepsilon_t| \geq \xi)$, which are summarised in (3).*

REMARK 1. In Assumption 1, we outline the properties and tail behaviour of $\{\varepsilon_t\}$. The requirement of $\{\varepsilon_t\}$ to satisfy the conditions in Theorem 14.9 of Davidson (1994) allows us to define $\{u_t\}$ as a strong mixing AR(q) process, with properties and tail behaviour outlined in Definition 1. This is the reason why we also need to make an explicit distributional assumption for $\{\varepsilon_t\}$, where its p.d.f, should be continuous almost everywhere. If this is not assumed, then $\{u_t\}$ may not be strong mixing, see e.g. Andrews (1984).

We proceed and place the following Assumption on $\{u_t\}$.

ASSUMPTION 2. *$\{u_t\}$ is a second order stationary process that satisfies (2), where we assume that the roots of the polynomial $\phi(z) = (1 - \sum_{i=1}^{q} \phi_i z^i)$, all lie outside the unit root circle. Then u_t can have either thin- or heavy-tailed distributions, such that $u_t \in \mathcal{T}(r), r > 0$, if Definition 1.1 is satisfied, or $u_t \in \mathcal{H}(\theta), \theta > 4$, if Definition is 1.2 satisfied.*

Note that the tail behaviour of $\{u_t\}$ in Assumption 2 is governed by the tail behaviour of $\{\varepsilon_t\}$.

We propose to estimate the unknown parameters associated with (1) feasibly within the scope of a high-dimensional model. To describe the method in detail, let

$$\widetilde{u}_t = y_t - \boldsymbol{x}_t' \widetilde{\boldsymbol{\beta}}, \quad t = 1, \dots, T,$$

where $\widetilde{\boldsymbol{\beta}}$ is a preliminary LASSO estimate of $\boldsymbol{\beta}$. Further, let $\widehat{\boldsymbol{\phi}} = (\widehat{\phi}_1, \dots, \widehat{\phi}_q)'$ be the q^{th} order OLS estimator of the autoregressive parameters for $\{\widetilde{u}_t\}$ obtained as the the solution to the minimisation of

$$T^{-1} \sum_{t=q+1}^{T} \left(\widetilde{u}_t - \phi_1 \widetilde{u}_{t-1} - \dots - \phi_q \widetilde{u}_{t-q} \right)^2,$$

over $\boldsymbol{\phi} = (\phi_1, \dots, \phi_q)'$. Then, an asymptotically valid feasible and penalised GLS estimate of $\boldsymbol{\beta}$ can be obtained minimising the following loss function, with respect to the parameters of interest, $\boldsymbol{\beta}$:

$$\mathcal{L}\left(\boldsymbol{\beta}; \widehat{\boldsymbol{\phi}}\right) = \arg\min_{\boldsymbol{\beta} \in \mathbb{R}^p} \frac{1}{2T} \left\| \widetilde{\boldsymbol{y}} - \widetilde{\boldsymbol{X}} \boldsymbol{\beta} \right\|_2^2 + \lambda \left\| \boldsymbol{\beta} \right\|_1, \tag{4}$$

where $\widetilde{\boldsymbol{X}} = \widehat{\boldsymbol{L}} \boldsymbol{X}$, $\widetilde{\boldsymbol{y}} = \widehat{\boldsymbol{L}} \boldsymbol{y}$, $\boldsymbol{y} = (y_1, \dots, y_T)'$ and $\widehat{\boldsymbol{L}}$ is a $(T-q) \times T$ matrix defined as:

$$\widehat{\boldsymbol{L}} = \begin{pmatrix} -\widehat{\phi}_q & -\widehat{\phi}_{q-1} & \cdots & \cdots & -\widehat{\phi}_1 & 1 & \cdots & 0 & 0 \\ 0 & -\widehat{\phi}_q & -\widehat{\phi}_{q-1} & \cdots & \cdots & -\widehat{\phi}_1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & -\widehat{\phi}_q & -\widehat{\phi}_{q-1} & \cdots & -\widehat{\phi}_1 & 1 \end{pmatrix}, \quad \widehat{\boldsymbol{\phi}} = \begin{pmatrix} \widehat{\phi}_1 \\ \vdots \\ \widehat{\phi}_q \end{pmatrix}. \tag{5}$$

One can use the scalar representation of $\widetilde{\boldsymbol{y}}$ and vector representation of $\widetilde{\boldsymbol{X}}$, to obtain

$$\widetilde{y}_t = y_t - \sum_{j=1}^{q} \widehat{\phi}_j y_{t-j}, \quad \widetilde{\boldsymbol{x}}_t = \boldsymbol{x}_t - \sum_{j=1}^{q} \widehat{\phi}_j \boldsymbol{x}_{t-j}, \quad t = q+1, \dots, T.$$

The loss function in (4) corresponds to the ℓ_1-penalised loss function using the estimates of $\widehat{\boldsymbol{\phi}}$, where the additional penalty is added to the least squares objective. Note that the asymptotic consistency of $\widehat{\boldsymbol{\beta}}$ can be established even prior to the estimation of $\boldsymbol{\phi}$. We prove this result in Lemma 1. Further, following the same argument that we used to define $\widehat{\boldsymbol{L}}$ in (5), we define $\boldsymbol{L}$, a $(T-q)\times T$ matrix, with $\widehat{\boldsymbol{\phi}}$ replaced by $\boldsymbol{\phi}$, which is used to derive the infeasible GLS LASSO estimates, when the degree of autocorrelation in u_t is known, using $\boldsymbol{X}^* = \boldsymbol{L}\boldsymbol{X}$ and $\boldsymbol{y}^* = \boldsymbol{L}\boldsymbol{y}$. These quantities will be useful for our analysis in Section 4.

REMARK 2. Notice that Yule–Walker or Burg-type estimates (or maximum entropy estimates), see, for example, Burg (1968), Brockwell et al. (2005), can be used instead of the OLS estimates to obtain $\widehat{\boldsymbol{\phi}}$, used in the construction of $\widehat{\boldsymbol{L}}$, without changing the asymptotic properties of neither the preliminary estimate, $\widetilde{\boldsymbol{\beta}}$ nor the GLS-type estimate, $\widehat{\boldsymbol{\beta}}$.

ASSUMPTION 3. *The components of $\boldsymbol{x}_t = (x_{t,1},\ldots,x_{t,p})'$, can have either thin- or heavy-tailed distributions, such that for $i = 1,\ldots,p$,*

$$\{x_{t,i}\} \in \begin{cases} \mathcal{T}(r), r>0, \text{ satisfies Definition 1.1.} \\ \mathcal{H}(\theta), \theta>4, \text{ satisfies Definition 1.2} \end{cases}. \tag{6}$$

Assumption 3 imposes moment conditions and defines the tail behaviour of $\{x_{t,i}\}$. Notice that with heavy-tailed distributions for $\{\boldsymbol{x}_t\}$ and $\{u_t\}$, the probability inequalities in (3) become polynomial rather than exponential. By allowing heavy tails in $\{\boldsymbol{x}_t\}$ and $\{u_t\}$, we permit departures from exponential probability tails, implying that p should be relatively small compared to T, see, for example, Assumption 6. In the Supplement we explore the small sample properties of the proposed estimators in cases where the sampling distributions of $\{\boldsymbol{x}_t\}$ and $\{u_t\}$ have heavier tails compared to the normal distribution, for example, $\boldsymbol{x}_t, u_t \sim t_d,\ d := \{4, 8, 16\}$, using simulations.

ASSUMPTION 4. *For all t,s (i) $E(u_t|\boldsymbol{x}_s) = 0$; (ii) $\sup_x E(u_t^2|\boldsymbol{x}_s) = O(1)$; (iii) The smallest eigenvalue of $\boldsymbol{\Sigma}_{u|x}$ is bounded away from zero, almost surely, by some constant $\delta_{\min} > 0$, $\Lambda_{\min}(\boldsymbol{\Sigma}_{u|x}) > \delta_{\min}$, where $\boldsymbol{\Sigma}_{u|x} = \left(\gamma_{|j-k|}\right)_{j,k=1,\ldots,T}$ and $\gamma_k = E(u_k u_0|\boldsymbol{x}_s)$, for some $k \geq 0$.*

In Assumption 4 we place a strict exogeneity assumption and rule out the addition of lagged endogenous variables in the RHS of (1), that is we assume that $\{u_t\}$ is uncorrelated with $\{\boldsymbol{x}_s\}$, and that $E(\varepsilon_t|\boldsymbol{x}_s) = 0$, and $\sup_x E(\varepsilon_t^2|\boldsymbol{x}_s) = O(1)$, for all t,s. We now further consider the overall tail behaviour of the empirical process, $\{\zeta_{t,i}\}$, which is defined as $\zeta_{t,i} = \{x_{t,i}\varepsilon_t\}$:

ASSUMPTION 5. *For every $i = 1,\ldots,p$, there are two possible tail behaviours for $\{\zeta_{t,i}\}$; either i) $\{\zeta_{t,i}\} \in \mathcal{T}(r)$, if both $\{x_{t,i}\} \in \mathcal{T}(r)$ and $\varepsilon_t \in \mathcal{T}(r)$, for some $r > 0$, or ii) $\{\zeta_{t,i}\} \in \mathcal{H}(\theta)$, if at least one (or both) of $\{x_{t,i}\}, \{\varepsilon_t\} \in \mathcal{H}(\theta)$, for some $\theta > 4$, and $\mathcal{T}(r)$ and $\mathcal{H}(\theta)$ are given in Definition 1.*

REMARK 3. Alternatives to α-mixing processes can also be considered, such as τ-mixing processes, which allow for heavier than Gaussian tails. Babii et al. (2024) establish a version of the Fuk–Nagaev inequality for τ-mixing processes, to derive the debiased central limit theorem for low-dimensional groups of regression coefficients.

Our objective is to use a GLS-type of estimator to address the issue of serial-autocorrelation in the error term, u_t, that will enable sharper inference following the paradigm of the DEBIASED LASSO. Such methods have been examined in the past in low dimensional cases, where $T >> p$, see, for example, Amemiya (1973), and Kapetanios and Psaradakis (2016). We extend this framework to allow for $p > T$, ensuring that $\widetilde{u}_t$ is asymptotically consistent using a penalised estimate of $\boldsymbol{\beta}$, while we limit our research to finite autoregressions.

REMARK 4. The theoretical results presented in this section can be established even when $\{u_t\}$ admits an $AR(\infty)$ representation, which could arise in practice by inverting an MA(1) process. This would entail certain additional assumptions to be made about $\{u_t\}$, and the theoretical proofs would come at an increased mathematical cost. Given the very good performance of an AR(1) model in instances as above, see, for example, Kapetanios and Psaradakis (2016), and Perron and González-Coya (2022), in low-dimensional settings, we choose not to pursue this and leave it for future work. However, it is of interest to examine the properties of the proposed estimator in the case where the order, q, used for the estimation is misspecified. In the Supplement, we explore the small sample properties of the GLS and DEBIASED GLS estimators when the error is generated from an MA(1) process to examine the impact of the misspecification of q in estimation and inference.

The use of penalised models highlights the necessity of imposing sparsity conditions on the parameter set, $\boldsymbol{\beta}$, which in turn allow for a degree of misspecification in the model. The latter can be approximated by a sparse linear model following certain boundedness conditions on the smallest eigenvalues of the sample variance-covariance matrix, $\sum_{t=1}^{T} \boldsymbol{x}_t \boldsymbol{x}_t'/T$.

ASSUMPTION 6. *For some $\theta > 4$, we define the regularisation parameter:*

$$\lambda = O(T^{-1/2} \log^{1/2} p) \vee O(p^{1/\theta} T^{-(\frac{\theta}{2}+1)}) \tag{7}$$

The regularization parameter in Assumption 6 is determined by the distribution of the data, namely on whether the decay of the probability tails of the data are either exponential, or polynomial.

We now define the following index set for $i = 1, \ldots, p$:

$$S_0 = \{i : \beta_i \neq 0\}, \tag{8}$$

with cardinality $s_0 = |S_0|$. The set S_0, is the active set, and s_0 the sparsity index of β. We make use of S_0 in the following assumption, which forms the compatibility condition, as seen in Bickel et al. (2009), Raskutti et al. (2010) and Chapter 6 of Bühlmann and Van De Geer (2011). We make the following assumption under S_0.

ASSUMPTION 7. *For $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)'$, denote $\boldsymbol{\beta}_{s_0} := \beta_i \mathbf{1}\{i \in S_0\}$, $\boldsymbol{\beta}_{s_0^c} := \beta_i \mathbf{1}\{i \notin S_0\}$, such that $\boldsymbol{\beta}_{s_0}$ has zeroes outside the set S_0, such that $\|\boldsymbol{\beta}_{s_0^c}\|_1 \leq 3\|\boldsymbol{\beta}_{s_0}\|_1$, $\boldsymbol{\Sigma} = E[T^{-1}\sum_{t=1}^{T}(\boldsymbol{x}_t -$*

$\sum_{j=1}^{q}\phi_j\boldsymbol{x}_{t-j})(\boldsymbol{x}_t-\sum_{j=1}^{q}\phi_j\boldsymbol{x}_{t-j})']$. *We define the following compatibility constant*

$$\zeta_*^2(s_0,\boldsymbol{\phi})=\min_{\substack{\left\|\boldsymbol{\beta}_{s_0^c}\right\|_1\leq 3\left\|\boldsymbol{\beta}_{s_0}\right\|_1\\ \boldsymbol{\beta}\in\mathbb{R}^p\setminus\{0\}}}\frac{|S_0|\boldsymbol{\beta}'\boldsymbol{\Sigma}\boldsymbol{\beta}}{\left\|\boldsymbol{\beta}_{s_0}\right\|_1^2}. \tag{9}$$

Consider $\zeta_*^2(s_0,\boldsymbol{\phi})>0$, *then* $\left\|\boldsymbol{\beta}_{s_0}\right\|_1^2\leq[\boldsymbol{\beta}'\boldsymbol{\Sigma}\boldsymbol{\beta}]\,\zeta_*^{-2}(s_0,\boldsymbol{\phi})$, *holds. The constant 3 might appear arbitrary and can be replaced with a number larger than 1, at the cost of changing the lower bound for* λ.

Assumption 7 implies the "restricted" positive definiteness of the variance-covariance matrix, which is valid only for the vectors satisfying $\|\boldsymbol{\beta}_{s_0^c}\|_1\leq 3\|\boldsymbol{\beta}_{s_0}\|_1$. Assumption 7 excludes cases of a perfectly multicollinear design, while Toeplitz covariance structures can be allowed in our setting. We further note that Assumption 7 is a relaxation of explicitly imposing a bound on the precision matrix, such that the $\Lambda_{\min}(E[\sum_{t=1}^{T}\boldsymbol{x}_t\boldsymbol{x}_t'/T])>\delta>0$.

Notice that in Assumption 7 we present a modified restricted eigenvalue condition, based on the population variance-covariance matrix, $\boldsymbol{\Sigma}$. In Lemma 10.7 of the Supplement, we show that the population covariance $\boldsymbol{\Sigma}=E\left[T^{-1}(\sum_{t=1}^{T}(\boldsymbol{x}_t-\sum_{j=1}^{q}\phi_j\boldsymbol{x}_{t-j}))(\sum_{t=1}^{T}(\boldsymbol{x}_t-\sum_{j=1}^{q}\phi_j\boldsymbol{x}_{t-j}))'\right]$ can be approximated well by the sample covariance estimate, $T^{-1}\sum_{t=1}^{T}(\sum_{t=1}^{T}(\boldsymbol{x}_t-\sum_{j=1}^{q}\phi_j\boldsymbol{x}_{t-j}))(\sum_{t=1}^{T}(\boldsymbol{x}_t-\sum_{j=1}^{q}\phi_j\boldsymbol{x}_{t-j}))'$, such that

$$\mathrm{P}\left(\max_{0<i,k\leq p}\left|\widehat{\boldsymbol{\Sigma}}_{i,k}-\boldsymbol{\Sigma}_{i,k}\right|>\nu\right)\to 0,\quad\text{for some}\quad\nu>0. \tag{10}$$

We provide a proof of (10) in the Supplement. Note that a statement similar to (10) can be shown using $\ddot{\boldsymbol{\Sigma}}=E\left(T^{-1}\sum_{t=1}^{T}\boldsymbol{x}_t\boldsymbol{x}_t'\right)$ and its estimator $\widehat{\ddot{\boldsymbol{\Sigma}}}=T^{-1}\sum_{t=1}^{T}\boldsymbol{x}_t\boldsymbol{x}_t'$.

Lemma 1. *For some parameter* $\ddot{\lambda}$, *let*

$$\mathcal{L}(\boldsymbol{\beta})=\|\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta}\|_2^2+\ddot{\lambda}\,\|\boldsymbol{\beta}\|_1 \tag{11}$$

be the LASSO *regression prior to obtaining* $\widetilde{u}_t$. *Consider* $\ddot{\boldsymbol{\Sigma}}$ *and the following restricted eigenvalue condition,*

$$s_0\left\|\boldsymbol{\beta}_{s_0}\right\|_2^2\leq\boldsymbol{\beta}'\ddot{\boldsymbol{\Sigma}}\boldsymbol{\beta}\phi_0^{-2}, \tag{12}$$

for some $\phi_0^{-2}>0$ *and* $s_0=|S_0|$.

Under Assumptions 1, 3, 4, specifically for $i=1,\dots,p$ *and if* $\{x_{t,i}u_t\}\in\mathcal{T}(r)$, $r>0$ *we show that (13) and (14) hold with probability at least* $1-cp^{-\epsilon}$, *for some* $\ddot{\lambda}\asymp T^{-1/2}\log^{1/2}p$. *Alternatively, if* $\{x_{t,i}u_t\}\in\mathcal{H}(\theta)$, $\theta>4$, *then (13) and (14) hold with probability at least* $1-cp\left(\frac{\lambda_0\sqrt{T}}{2}\right)^{-\theta}T^{-\frac{\theta}{2}-1}$ *and for some* $\lambda_0=p^{\frac{1}{\theta}}T^{-(\frac{\theta}{2}+1)}<\ddot{\lambda}/2$, $\ddot{\lambda}\asymp p^{1/\theta}T^{-(\frac{\theta}{2}+1)}$, *positive constants,* c,ϵ *and* $(s_0\lambda_0)=o(1)$ *(in both cases of* $\ddot{\lambda}$*), we have the following non-asymptotic bound:*

$$T^{-1}\left\|\boldsymbol{X}\left(\widetilde{\boldsymbol{\beta}}-\boldsymbol{\beta}\right)\right\|_2^2\leq 4\ddot{\lambda}^2 s_0\phi_0^{-2} \tag{13}$$

$$\left\|\widetilde{\boldsymbol{\beta}}-\boldsymbol{\beta}\right\|_1 \leq 4\ddot{\lambda}s_0\phi_0^{-2}, \tag{14}$$

where $\widetilde{\boldsymbol{\beta}}$ is the LASSO *estimate obtained from the solution of* (11)*, and* (12) *holds. A detailed proof of Lemma 1 can be found in the Supplement.*

Further, in Lemma 1, we show that the oracle inequalities hold prior to estimating $\widehat{\boldsymbol{\phi}}$. Notice that in (12) we use the population covariance matrix instead of the sample, $T^{-1}\sum_{t=1}^{T}\boldsymbol{x}_t\boldsymbol{x}_t'$, for which an argument similar to (10) holds and is formalised in the Supplement, see Lemma 10.7. Notice that (12) is the restricted eigenvalue condition on the population covariance, $E(T^{-1}\sum_{t=1}^{T}\boldsymbol{x}_t\boldsymbol{x}_t')$, while ϕ_0^{-2} is the compatibility constant. Lemma 1 is of paramount importance and is required in order to establish the following corollary and Theorem 1.

Corollary 1. *Let Assumptions 1, 3,4, and 7 hold. Specifically for $i=1,\dots,p$ and if $\{x_{t,i}u_t\}\in\mathcal{T}(r)$, $r>0$, then, $\widehat{\boldsymbol{\phi}}=(\phi_1,\dots,\phi_q)'$ is an asymptotically consistent estimate of the autoregressive parameters, for some $\ddot{\lambda}\asymp\sqrt{\log p/T}$, and $s_0=o(\sqrt{T}/\log p)$,*

$$\left\|\widehat{\boldsymbol{\phi}}-\boldsymbol{\phi}\right\|_1 = O_P\left(s_0^2T^{-1}\log p \vee s_0T^{-1}\log p\right) = O_P\left(T^{-1/2}\right),$$

Alternatively, if $\{x_{t,i}u_t\}\in\mathcal{H}(\theta)$, $\theta>4$, $\ddot{\lambda}=O(p^{1/\theta}T^{-(\frac{\theta}{2}+1)})$, $\theta>4$ and $s_0=o(T^{(\theta-1)/2}p^{-1/\theta})$

$$\left\|\widehat{\boldsymbol{\phi}}-\boldsymbol{\phi}\right\|_1 = O_P\left(s_0^2p^{\frac{1}{\theta^2}}T^{-\left(\frac{\theta}{2}+1\right)^2} \vee s_0p^{\frac{1}{\theta^2}}T^{-\left(\frac{\theta}{2}+1\right)^2}\right) = O_P\left(T^{-1/2}\right).$$

Remark 5. In Corollary 1, the order, q, though finite, is unknown. To address this we adopt a sequential testing technique towards the selection of q, similar to Kapetanios and Psaradakis (2016). To add flexibility to our model we allow this technique to be dependent on the generation of $\{u_t\}$. To minimize dependence on these conditions, we consider the following bound on q, $1\leq q<q^*$, where $q^*=\lfloor T^{1/2}\rfloor$, which implies that the case of $u_t\sim$ i.i.d. is not included in our testing procedure. Alternatively, information criteria can also be used to determine the order q.

2.1 A feasible penalised GLS

We start with the main theorem that provides non-asymptotic guarantees for the estimation and prediction errors of the LASSO under a modified compatibility condition, i.e. Assumption 7, which restricts the smallest eigenvalue of the variance-covariance matrix.

Theorem 1. *We introduce non-asymptotic prediction and estimation bounds corresponding to the solution of* (4)*, $\widehat{\boldsymbol{\beta}}$,*

$$T^{-1}\left\|\widetilde{\boldsymbol{X}}\left(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}\right)\right\|_2^2 \leq 4\zeta_*^{-2}(s_0,\boldsymbol{\phi})s_0\lambda^2+\frac{\lambda^2}{T}, \tag{15}$$

$$\left\|\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}\right\|_1 \leq 4\zeta_*^{-2}(s_0,\boldsymbol{\phi})s_0\lambda+\frac{\lambda}{\sqrt{T}}, \tag{16}$$

which hold with probability at least $1-cp^{1-\frac{c_4}{(12C_0)^2}}$ for some $\lambda\asymp\sqrt{\log p/T}$, if Asumption 5 i) is satisfied for $c,C_0>0$ and c_4 a large enough constant, for $i=1,\dots,p$.

Conversely, (15)–(16) *hold with probability* $1-pc[(T^{1/2}\lambda_0/12C_0)^{-\theta}T^{-\frac{\theta}{2}-1}]$*, for some positive* $\lambda_0 = p^{\frac{1}{\theta}}T^{-(\frac{\theta}{2}+1)} < \lambda/2$*, where* $\lambda \asymp p^{\frac{1}{\theta}}T^{-(\frac{\theta}{2}+1)}$*, if Assumption* 5 *ii) is satisfied, with finite constants* $c_1, C_0 > 0$*, and* $(s_0\lambda_0) = o(1)$*.*

Notice that the non-asymptotic bounds presented in Theorem 1 consider both cases where $\{\boldsymbol{x}_t\}$ and (or) $\{u_t\}$ have thin- or heavy-tails. In each respective case of tail behaviour, the asymptotic order of the tuning parameter changes in order for the statements in (15)–(16) to hold with probability approaching 1. The following Corollary is a consequence of Theorem 1, providing asymptotic rates for the bounded processes in (15) and (16).

COROLLARY 2. *Let Assumptions 1–7 hold. Then for the feasible* GLS LASSO*, we obtain the following asymptotic rates:*

$$\frac{1}{T}\left\|\widetilde{\boldsymbol{X}}\left(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}\right)\right\|_2^2 = O_P\left(s_0\lambda^2 \vee \lambda^2 T^{-1}\right), \quad \left\|\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}\right\|_1 = O_P(s_0\lambda \vee \lambda T^{-1/2}), \tag{17}$$

hold for p, T *sufficiently large, and two cases of* λ *and* s_0*:*

$$\{\lambda, s_0\} = \begin{cases} \left\{O\left(\sqrt{\log p/T}\right),\ o\left(\sqrt{T}/\log p\right)\right\}, & \text{if } \{\zeta_{t,i}\} \in \mathcal{T}(r),\ r>0, \text{ hold}, \\ \left\{O\left(p^{\frac{1}{\theta}}T^{-(\frac{\theta}{2}+1)}\right),\ o\left(T^{\frac{\theta+1}{2}}p^{-\frac{1}{\theta}}\right)\right\}, & \text{if } \{\zeta_{t,i}\} \in \mathcal{H}(\theta),\ \theta>4, \text{ hold}, \end{cases} \tag{18}$$

where $\{\zeta_{t,i}\}$ *and its properties are defined in Assumption* 5*.*

3 Point-wise valid inference based on the (feasible) GLS LASSO

A natural avenue to inference, having obtained the estimates of $\boldsymbol{\beta}$ from (4) is to re-estimate the parameters of the estimated active set, $\widehat{S}_0 = \{i : \ \widehat{\beta}_i \neq 0\}$, via OLS. Formally, let $\ddot{\boldsymbol{\beta}}_{S_0}$ be the vector whose i^{th} element equals the least square re-estimate for all $i \in \widehat{S}_0$ and zero otherwise, while $\widehat{\boldsymbol{\beta}}_{S_0}$ denotes the oracle assisted least squares estimates only including the relevant variables, those indexed by S_0. The following theorem shows that this indeed leads to point-wise valid confidence bands for the non-zero entries of $\boldsymbol{\beta}$, i.e. those indexed by S_0, since $\widehat{S}_0 = S_0$ asymptotically.

THEOREM 2. *Let the Assumptions of Theorem 1 hold, and assume that* $\widehat{\boldsymbol{\Sigma}}_{S_0} = T^{-1}\widetilde{\boldsymbol{X}}'_{S_0}\widetilde{\boldsymbol{X}}_{S_0}$ *is invertible for* $s_0 < p$*. Then we have that*

$$\sqrt{T}\left\|\left(\ddot{\boldsymbol{\beta}}_{S_0}-\boldsymbol{\beta}_{S_0}\right)-\left(\widehat{\boldsymbol{\beta}}_{S_0}-\boldsymbol{\beta}_{S_0}\right)\right\|_1 = o_p(1). \tag{19}$$

Theorem 2 illustrates that performing least squares after model selection leads to inference that is asymptotically equivalent to inference based on least squares only including the covariates in S_0. However, it is important to note that such inference is of a point-wise nature and it is not uniformly valid. This non-uniformity is visible in the confidence intervals, that could occasionally undercover the true parameter. This remark serves as a warning in applying point-wise inference after ℓ_1-regularisation, similar to the discussion in Leeb and Pötscher (2005).

4 Uniformly valid inference based on the feasible GLS LASSO

In this section we extend the DEBIASED LASSO estimator, introduced by Van de Geer et al. (2014) to accommodate autocorrelated errors. With mild assumptions on $\{\boldsymbol{x}_t\}$ in place, we devise a penalised GLS-type estimator, introduced in (4) to enable valid inference in a high-dimensional setting. The construction of these inferential methods require solving the normal equations. The DEBIASED GLS LASSO is then formulated as

$$\widehat{\boldsymbol{b}} = \widehat{\boldsymbol{\beta}} + \boldsymbol{B}_{GLS}, \tag{20}$$

where $\widehat{\boldsymbol{\beta}}$ is the GLS LASSO estimator, $\boldsymbol{B}_{GLS} = T^{-1}\widehat{\boldsymbol{\Theta}}\widetilde{\boldsymbol{X}}'\left(\widetilde{\boldsymbol{y}} - \widetilde{\boldsymbol{X}}\widehat{\boldsymbol{\beta}}\right)$ is its bias correction, and $\widehat{\boldsymbol{\Theta}} = \widehat{\boldsymbol{\Sigma}}^{-1}$ is the approximation of the precision matrix $\boldsymbol{\Sigma}^{-1}$ using the nodewise GLS LASSO regressions. We define the nodewise estimator in the following section. Then, to obtain a scaled pivotal quantity for inference we substitute $\widetilde{\boldsymbol{y}}$ into (20), such that

$$\sqrt{T}\left(\widehat{\boldsymbol{b}} - \boldsymbol{\beta}\right) = \frac{1}{\sqrt{T}}\widehat{\boldsymbol{\Theta}}\widetilde{\boldsymbol{X}}'\widehat{\boldsymbol{L}}\boldsymbol{u} + \boldsymbol{\delta},$$

where $\boldsymbol{\delta} = \sqrt{T}(\widehat{\boldsymbol{\Theta}}\widehat{\boldsymbol{\Sigma}} - \boldsymbol{I}_{(p\times p)})(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})$ is the bias coming from the approximation of $\widehat{\boldsymbol{\Sigma}}^{-1}$ and the GLS LASSO estimator, and $\boldsymbol{\delta}$ is asymptotically negligible, while $\widehat{\boldsymbol{\Theta}}\widetilde{\boldsymbol{X}}'\widehat{\boldsymbol{L}}\boldsymbol{u}/\sqrt{T}$ is an asymptotically Gaussian quantity. These statements are formalised in Section 4.2, where we introduce the central limit theorem (CLT) based on the DEBIASED GLS LASSO estimator.

4.1 Construction of $\widehat{\boldsymbol{\Theta}}$

We consider the population nodewise regressions defined by the infeasible[1] GLS linear projections, where for each $i = 1, \dots, p$

$$x^*_{t,i} = \boldsymbol{x}^{*}_{t,-i}{}'\boldsymbol{\gamma}_i^{(*,0)} + \upsilon^*_{t,i}. \tag{21}$$

Let $\boldsymbol{\gamma}_i^{(*,0)} = (\gamma^{*,0}_{i,1}, \dots, \gamma^{*,p}_{i,1})'$ be the solution of

$$\boldsymbol{\gamma}_i^{(*,0)} = \arg\min_{\boldsymbol{\gamma}} \left\{ E\left[T^{-1}\sum_{t=1}^{T}\left(x^*_{t,i} - \boldsymbol{x}^{*}_{t,-i}{}'\boldsymbol{\gamma}\right)\right]\right\}, \tag{22}$$

and $\tau_i^{*2} = T^{-1}\sum_{t=1}^{T} E[\upsilon^{*}_{t,i}{}^2]$ is the population regression error of (21). Note that by construction of the nodewise linear projections, it holds that $E[\upsilon^*_{t,i}] = 0$, for all t, i and $E[\upsilon^*_{t,i}x^*_{t,k}] = 0$, for every $k \neq i$. We make the following assumption on the errors of (21).

ASSUMPTION 8. *For every $i = 1, \dots, p$, the process $\{\upsilon_{t,i}\}$ is α-mixing, with properties outlined in Definition 1. Depending on the tail assumptions for the distribution of $\{\upsilon_{t,i}\}$, one of the two following conditions hold, (i) $E[\exp(d|\upsilon^*_{t,i}|^r)] < \infty$, for some $d, r > 0$ or (ii) $E[|\upsilon^*_{t,i}|^\theta] < \infty$, $\theta > 4$.*

[1] The feasible GLS-based estimators are shown to be asymptotically equivalent to their infeasible GLS counterparts. A detailed analysis is provided in the Supplement.

REMARK 6. Assumption 8 requires that the errors of the nodewise linear projections are α-mixing processes and have bounded moments. The order of these moments is determined by whether the error distribution is thin- or heavy-tailed; see Definition 1. By the properties of α-mixing processes, see, for example, Theorem 14.1 of Davidson (1994) and using Assumption 3, we establish α-mixing properties for the product $(v_{t,i}^* x_{t,k}^*)$, for $k \neq i$, which we use extensively in the derivation of the asymptotic distribution of the DEBIASED GLS LASSO.

We now shift our focus to a feasible way of approximate $\mathbf{\Sigma}^{-1}$, using nodewise linear projections based on feasible transformations of the data. For each $i = 1, \ldots, p$, let $\widehat{\boldsymbol{\gamma}}_i = (\widehat{\gamma}_{i,1}, \ldots, \widehat{\gamma}_{i,p})'$ be the solution of

$$\arg\min_{\boldsymbol{\gamma} \in \mathbb{R}^{p-1}} \left\| \widetilde{\boldsymbol{x}}_i - \widetilde{\boldsymbol{X}}_{-i} \boldsymbol{\gamma} \right\|_2^2 + 2\lambda_i \left\| \boldsymbol{\gamma} \right\|_1,$$

where $\widetilde{\boldsymbol{x}}_i$ is a $(T-q) \times 1$ vector of observations and $\widetilde{\boldsymbol{X}}_{-i}$ a design sub-matrix, missing the i^{th} column. Consider the approximation error, calculated by the following nodewise regressions:

$$\widehat{\tau}_i = \left\| \widetilde{\boldsymbol{x}}_i - \widetilde{\boldsymbol{X}}_{-i} \widehat{\boldsymbol{\gamma}}_i \right\|_2^2 + \lambda_i \left\| \widehat{\boldsymbol{\gamma}}_i \right\|_1,$$

and the nodewise estimator for the precision matrix, $\widehat{\mathbf{\Theta}} = \widehat{\boldsymbol{M}}^{-\mathbf{1}} \widehat{\boldsymbol{C}}$, where $\widehat{\boldsymbol{M}} = \operatorname{diag}\left(\widehat{\tau}_1^2, \ldots, \widehat{\tau}_p^2\right)$ and

$$\widehat{\boldsymbol{C}} := \begin{pmatrix} 1 & -\widehat{\gamma}_{1,2} & \cdots & -\widehat{\gamma}_{1,p} \\ -\widehat{\gamma}_{2,1} & 1 & \cdots & -\widehat{\gamma}_{2,p} \\ \vdots & \vdots & \ddots & \vdots \\ -\widehat{\gamma}_{p,1} & -\widehat{\gamma}_{p,2} & \cdots & 1 \end{pmatrix}.$$

The following assumption introduces an additional condition on the required for the debiased central limit theorem.

ASSUMPTION 9. *(i)* $\|\mathbf{\Theta}\|_\infty = O(1)$; *(ii) For* $s_i = |\{i \neq j : \ \Theta_{i,j} \neq 0\}|$ *the sparsity of the precision matrix* $\mathbf{\Theta}$, *and* $\theta > 4$ $s_i = o(\sqrt{T}/\log p) \vee o(T^{\frac{\theta+1}{2}} p^{-\frac{1}{\theta}})$.

REMARK 7. Assumption 9 (i) requires the rows of the approximate inverse to be bounded in terms of ℓ_1-norm, which is a reasonable assumption, since the precision matrix can have sparse entries, in high dimensional settings. Assumption 9 (ii) introduces the a requirement on the sparse elements of each row of the precision matrix. Note that s_i depends on the tail behaviour of the data, see Assumption 3 and 8. We make use of these claims in the Supplement.

We discuss the asymptotic properties of $\widehat{\mathbf{\Theta}}$ in the Supplement.

4.2 Feasible Debiased CLT

We explore the asymptotic normality of the DEBIASED GLS LASSO and establish uniformly valid confidence intervals based on the theoretical results considered throughout Section 2.

THEOREM 3. *Consider the linear model in* (1), *the* GLS LASSO *estimator in* (4), *let* $\widehat{\mathbf{\Theta}}$ *be a suitable approximation of* $\widehat{\mathbf{\Sigma}}^{-1}$, *for a selection of* $\lambda_i \asymp \sqrt{\log p/T} \vee p^{\frac{1}{\theta}} T^{-(\frac{\theta}{2}+1)}$. *Further, let Assumptions 1, 3, and Assumptions 4 – 9 hold, and* $\boldsymbol{a} \in \mathbb{R}^{\tilde{p}}$, *with* $\|\boldsymbol{a}\|_2 = 1$, *and* $\tilde{p} \in \{1, \ldots, h\}$, $h < p$, *we have*

$$\sqrt{T}\boldsymbol{a}'\left(\widehat{\boldsymbol{b}}-\boldsymbol{\beta}\right)=\boldsymbol{a}'\boldsymbol{z}+\boldsymbol{a}'\boldsymbol{\delta},\quad \boldsymbol{a}'\boldsymbol{z}=\frac{1}{\sqrt{T}}\boldsymbol{a}'\widehat{\boldsymbol{\Theta}}\widetilde{\boldsymbol{X}}'\widehat{\boldsymbol{L}}\boldsymbol{u}\sim\mathcal{N}\left(0,\sigma^2\boldsymbol{V}\right), \tag{23}$$

$$\boldsymbol{a}'\boldsymbol{\delta}=o_P(1), \tag{24}$$

where $\boldsymbol{V}=\boldsymbol{a}'\boldsymbol{\Theta}'\boldsymbol{\Sigma}_{xu}\boldsymbol{\Theta}\boldsymbol{a}$, $\widehat{\boldsymbol{\Theta}}$ *is defined in Section 4.1,* $\boldsymbol{z}$ *is the bias correction of the* DEBIASED GLS LASSO, $\boldsymbol{\delta}:=\sqrt{T}(\widehat{\boldsymbol{\Theta}}\widehat{\boldsymbol{\Sigma}}-\boldsymbol{I}_{(p\times p)})(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta})$, *and* $\boldsymbol{\Sigma}_{xu}=E[(\boldsymbol{X}'\boldsymbol{L}'\boldsymbol{L}\boldsymbol{u})(\boldsymbol{X}'\boldsymbol{L}'\boldsymbol{L}\boldsymbol{u})']$.

Next $\boldsymbol{a}'\widehat{\boldsymbol{\Theta}}\widehat{\boldsymbol{\Sigma}}_{xu}'\widehat{\boldsymbol{\Theta}}\boldsymbol{a}$ converges uniformly to the asymptotic variance of $T^{-1/2}\boldsymbol{a}'\left(\widehat{\boldsymbol{b}}-\boldsymbol{\beta}\right)$, such that

$$\sup_{\boldsymbol{\beta},\|\boldsymbol{\beta}\|_0\leq s_0}\left|\boldsymbol{a}'\widehat{\boldsymbol{\Theta}}\widehat{\boldsymbol{\Sigma}}_{xu}'\widehat{\boldsymbol{\Theta}}\boldsymbol{a}-E\left[\boldsymbol{a}'\boldsymbol{\Theta}\boldsymbol{\Sigma}_{xu}\boldsymbol{\Theta}\boldsymbol{a}\right]\right|=o_P(1),$$

for all $\boldsymbol{a}\in\mathbb{R}^{\tilde{p}}$, with $\|\boldsymbol{a}\|_2=1$, and $\tilde{p}\in\{1,\ldots,h\},\ h<p$. Following Theorem 3, we introduce asymptotic point-wise confidence intervals for the true parameter, β_i, $1\leq i\leq p$, given by

$$\mathrm{CI}\left(\alpha\right)=\left[\widehat{b}_i\pm z_{\alpha/2}\widehat{\sigma}\left(T^{-1}\widehat{\boldsymbol{\Theta}}'\widehat{\boldsymbol{\Sigma}}_{xu}\widehat{\boldsymbol{\Theta}}\right)_{i,i}^{1/2}\right], \tag{25}$$

such that

$$\inf_{\boldsymbol{\beta},\|\boldsymbol{\beta}\|_0\leq s_0}P\left(\beta_i\in\left[\widehat{b}_i\pm z_{\alpha/2}\widehat{\sigma}\left(T^{-1}\widehat{\boldsymbol{\Theta}}'\widehat{\boldsymbol{\Sigma}}_{xu}\widehat{\boldsymbol{\Theta}}\right)_{i,i}^{1/2}\right]\right)=1-\alpha$$

where $\widehat{\boldsymbol{\Sigma}}_{xu}=T^{-1}(\boldsymbol{X}'\widehat{\boldsymbol{L}}'\widehat{\boldsymbol{L}}\boldsymbol{u})(\boldsymbol{X}'\widehat{\boldsymbol{L}}'\widehat{\boldsymbol{L}}\boldsymbol{u})'$, $z_{\alpha/2}:=\Phi^{-1}\left(1-\alpha/2\right)$, α is the confidence level, and $\Phi(\cdot)$ is the standard normal cumulative density function (CDF), such that

$$\sup_{\boldsymbol{\beta}\in\mathcal{B}(s)}\left|P\left(\boldsymbol{a}'\boldsymbol{\beta}\in\mathrm{CI}\left(\alpha\right)\right)-(1-\alpha)\right|=o(1),\ \forall\ i=1,\ldots,p \tag{26}$$

and $\widehat{\sigma}^2$ is a consistent estimate of the variance of $\widehat{u}_t$. Thus, one can perform inference on β_i, for some $z\in\mathbb{R}$, using the following form

$$\sup_{z\in\mathbb{R}}\sup_{\boldsymbol{\beta},\|\boldsymbol{\beta}\|_0\leq s_0}\left|P\left(\frac{\sqrt{T}\boldsymbol{a}'\left(\widehat{\boldsymbol{b}}-\boldsymbol{\beta}\right)}{\widehat{\sigma}\sqrt{\boldsymbol{a}'\left(\widehat{\boldsymbol{\Theta}}'\widehat{\boldsymbol{\Sigma}}_{xu}\widehat{\boldsymbol{\Theta}}\right)\boldsymbol{a}}}\leq z\right)-\Phi(z)\right|=o(1). \tag{27}$$

Notice that Theorem 3 along with (27) show that $\frac{\sqrt{T}\boldsymbol{a}'(\widehat{\boldsymbol{b}}-\boldsymbol{\beta})}{\widehat{\sigma}\sqrt{\boldsymbol{a}'(\widehat{\boldsymbol{\Theta}}'\widehat{\boldsymbol{\Sigma}}_{xu}\widehat{\boldsymbol{\Theta}})\boldsymbol{a}}}$ converges to the standard normal distribution uniformly over $\boldsymbol{\beta}\in\mathcal{B}_{s_0}$, for $\mathcal{B}_{s_0}=\{\|\boldsymbol{\beta}\|_0\leq s_0\}$, for some $s_0=|S_0|$. Then, the joint asymptotic normality in (27) allows the construction of Wald tests, such that, for any $S\subseteq\{1,\ldots,p\},\ s=|S|$

$$\left\|\left(\boldsymbol{a}'\widehat{\boldsymbol{\Theta}}\widehat{\boldsymbol{\Sigma}}_{xu}'\widehat{\boldsymbol{\Theta}}\boldsymbol{a}\right)_S^{-1/2}\sqrt{T}\boldsymbol{a}'\left(\widehat{\boldsymbol{b}}_S-\boldsymbol{\beta}_S\right)\right\|_2^2\to_d\chi(s)^2,\ \|\boldsymbol{a}\|_2=\sqrt{s}.$$

Wald tests can be constructed using the delta method. Therefore, (26) and (27), are uniformly valid.

REMARK 8. It is reasonable to consider the question of the relative efficiency of GLS LASSO compared to the LASSO. Given that both have DEBIASED versions with associated normal approximations and asymptotic variances, a comparison of those variances is a valid focus. Unfortunately the use of nodewise regressions, to approximate the precision matrix, towards feasible inference, lead to a situation that is much more complex than the low dimensional problem. This is primarily caused by the fact that different nodewise regressions are used for the approximation of the precision matrix of the DEBIASED versions of the LASSO and GLS, thus making comparisons prohibitively difficult. Therefore, we rely on our simulation study to provide some information on the relative efficiency of the two methods. Small sample evidence verify that DEBIASED GLS appears more efficient compared to the DEBIASED LASSO, as both p, T increase.

5 Selection of the optimal regularisation parameter

In this section we describe the cross validation method used to select the optimal regularisation coefficient required to obtain $\widehat{\boldsymbol{\beta}}$, $\widetilde{\boldsymbol{\beta}}$ and $\widehat{\boldsymbol{\gamma}}$, defined in (3), (13) and (31) respectively. Due to the autocorrelation present in the errors, we choose to use the hv-block cross validation method proposed in Racine (2000), which is suitable for dependent data, instead of standard k-fold[2] cross validation.

We succinctly describe how hv-block cross validation works using the time series $\{y_t\}$ and $\{\boldsymbol{x}_t\}$ from model (1) as an example, where $t = 1, \dots, T$. At a given t, which corresponds to a specific validation run, the data is divided into three parts: the training set, the test set (or validation set) and the gap. The test set, in each validation run is constructed as a contiguous set of size $2v+1$ that is centred at t. Then the gap set is obtained from removing equally $2 \times h$ observations from each side of the test set. This is the key difference of hv-block with k-fold cross validation.

In Figure 1 we illustrate two cases where the test set is far away from the boundary and very near the boundary T. For the latter, since there are less than h observations to remove, the gap is of smaller dimensions. Crucially, given the requirement that test sets need to be contiguous, there is a total of $T - 2v$ test sets that can be considered.

Then the optimal regularisation coefficient is selected as the one that minimises the following cross validation function:

$$\mathcal{L}_{\text{CV}}(\lambda) = \frac{1}{T-2v} \sum_{t=v+1}^{T-v} \left\| y_{(t:v)} - \boldsymbol{x}'_{(t:v)} \widehat{\boldsymbol{\beta}}_{(-t:h,v)} \right\|_2^2, \tag{28}$$

which is the average obtained from all $T-2v$ test sets and $\widehat{\boldsymbol{\beta}}_{(-t:h,v)}$ is estimated on each training set. The important parameters in hv-block cross validation are h and v. Regarding h, it can either be small as discussed in Györfi et al. (2013), or large, i.e. $h/T \in (0, 1/2)$, as required in Burman et al. (1994).

Different choices can also be used for v: if $v = 0$, then test sets are one-dimensional (this

[2] In previous versions of this paper we considered a k-fold cross validation method where we did not shuffle the data within the blocks, and they were taken to be non-overlapping and consequent to each other. We find that overall our Monte Carlo results do not change significantly, but we do observe a slight reduction in the MSE of the estimators under hv-block cross validation.

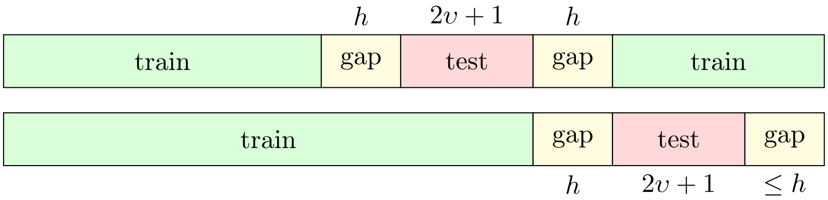


Figure 1: Example of $h\upsilon$-block cross-validation sample construction. The upper test set is far away from the boundary, whereas the lower test set is near the boundary.

is the h-block cross validation scheme in Burman et al. (1994)) and a version of this cross validation has been used in Babii et al. (2024), whereas Racine (2000) requires υ to be larger than the size of the training sample. Both of these choices however do not provide consistency in the sense of Shao (1993), as discussed in Zheng (2019), which while important is outside the scope of our paper.

In the spirit of k-fold cross validation and in order to reduce the computational cost of the $h\upsilon$-block cross validation, we construct only $K = 10$ contiguous test samples, hence we choose $2\upsilon + 1 = T/K$. Regarding our choice for the gap, we select $h = \lfloor T/4 \rfloor$, hence $h/T \approx 1/4 \in (0, 1/2)$, as required in Burman et al. (1994).

6 Simulation study

We use simulations to verify the theoretical properties of the proposed methodology in finite samples. We assess the performance of our method against the DEBIASED LASSO. We generate an array of samples from the following model

$$y_t = \boldsymbol{x}_t' \boldsymbol{\beta} + u_t, \quad u_t = \phi u_{t-1} + \varepsilon_t, \tag{29}$$

where $\boldsymbol{x}_t,\ \varepsilon_t \sim \text{i.i.d.}\,\mathcal{N}(0,1)$ and the sample sizes $T \in \{200, 500, 1000\}$ and $p \in \{100, 200, 500\}$. The active set has cardinality $s_0 = |S_0| = 3$, where $S_0 = \{i : \beta_i \neq 0,\ i = 1, \ldots, p\}$, while $S_0^c = \{i : \beta_i = 0,\ i = s_0 + 1, \ldots, p\}$. We choose the sparsity level s_0, according to the results of Van de Geer et al. (2014), which are indicative of good performance for the DEBIASED LASSO. Simulation results for $s_0 = 7$ are relegated to the Supplement. Each of the sets, S_0, S_0^c assume the form: $S_0 = \{1, 2, \ldots, s_0\} \equiv \{v_1, \ldots, v_{s_0}\}$, where $v_1, \ldots, v_{s_0}$ is a realisation of random draws of S_0 without replacement from $\{1, \ldots, p\}$. The parameters, $\beta_i \in S_0$ are simulated from the $U[0,1]$ distribution at each replication, while the autoregressive parameter ϕ used for the simulation of u_t, takes values $\phi = [0, 0.5, 0.8, 0.9, 0.95]$, where $\phi = 0$ indicates that $u_t \sim \text{i.i.d.}\,\mathcal{N}(0,1)$.

We base our findings on 1000 Monte Carlo simulations of (29). Further, we test the null hypothesis $H_0 : \beta_i = 0$, using the following t-statistic

$$\widehat{S}_i = T^{1/2} \frac{\widehat{b}_i - \beta_{i|H_0}}{\widehat{\sigma}\sqrt{\left(\widehat{\boldsymbol{\Theta}}\widehat{\boldsymbol{\Sigma}}\widehat{\boldsymbol{\Theta}}\right)_{i,i}}}, \tag{30}$$

where, for all $i = 1, \dots, p$, $\widehat{S}_i \sim t^{(T-1)}$, and $t^{(T-1)}$ the student's-t distribution with $T-1$ degrees of freedom, $\beta_{i|H_0} = 0$ under the null, $\widehat{b}_i$ is defined in (20), $\widehat{\mathbf{\Sigma}}$ is the sample covariance matrix, and $\widehat{\mathbf{\Theta}}$ its approximate inverse, defined in Section 4.1. Note that in the case where $i \in S_0$, we report the empirical power of the test based on (30). All tests are carried out at $\alpha = 5\%$ significance level.

Prior to the estimation of $\widehat{b}_i$, the estimate of $\widehat{\boldsymbol{\phi}}$, and a value for the regularisation parameter are required. In Section 2, we describe the method used for the estimation of $\widehat{\boldsymbol{\phi}}$, along with the testing scheme used for the selection of the optimal lag order, q, of the $AR(q)$ model. Further, we use the time-series cross-validation scheme described in Section 5, where $K = 10$, towards the optimal selection of the regularisation parameter, λ to obtain $\widehat{\boldsymbol{\beta}}$, $\widetilde{\boldsymbol{\beta}}$ and $\widehat{\boldsymbol{\gamma}}$. The results are displayed in Tables 1–2.

We further examine the performance of the LASSO, GLS LASSO, DEBIASED LASSO and DEBIASED GLS LASSO estimators in terms of average *root mean squared error* (RMSE) throughout 1000 replications of model (29). The results are displayed in Table 1.

The first panel of Table 1 reports the ratio of the average RMSE of the LASSO estimator over the RMSE of the GLS LASSO. The second panel of Table 1, reports the ratio of the average RMSE of the DEBIASED LASSO estimator over the RMSE of DEBIASED GLS LASSO. Entries larger than 1 indicate superiority of the competing model (GLS LASSO). In the parenthesis, the reported entries correspond to the RMSE of the GLS LASSO in Panel I and the DEBIASED GLS LASSO in Panel II.

The evidence is compelling, since both the GLS LASSO and DEBIASED GLS LASSO have, in general, the smallest RMSE compared with their counterpart respectively. The most pronounced cases of improvement are when the autocorrelation is strong, i.e. $\phi \in \{0.9, 0.95\}$ for all sample sizes, where the RMSE of the proposed method is almost (and /or more) two times smaller than the one recorded by the benchmark, LASSO. Notice that in the case of $\phi = 0$, GLS LASSO and LASSO perform equally well.

In the second panel of Table 2 we facilitate the comparison of the two methods by reporting the size adjusted power of the test and the standard errors, as seen in (30), reported in parenthesis. By size-adjusted power, we mean that if the test is found to reject at a rate $\widehat{\alpha} > 0.05$ under the null, then the power of the test is adjusted by finding the 0.95 empirical quantile of the statistic in (30). We then report the rate that exceeds this quantile. Alternatively, if $\widehat{\alpha} \leq 0.05$, then no adjustment is needed.

It is noticeable that when autocorrelation is present in the error term, DEBIASED LASSO underperforms severely in terms of all measures reported. More specifically, as T increases the benefits of DEBIASED GLS LASSO are more clear, the size adjusted power reported for our method is approaching the desirable rate of 0.95, while for the DEBIASED LASSO there are noticeable deviations from the nominal rate. Further, the standard errors of the GLS LASSO appear significantly smaller, indicative of narrower confidence intervals, therefore efficient estimation. On the other hand, LASSO reports standard errors almost twice the as much as the proposed estimator, coupled with the large deviations from 95% in terms of size-adjusted power reported, LASSO appears inefficient as ϕ approaches 0.9. The latter becomes severe for higher serial autocorrelation reported, e.g. $\phi = [0.8, 0.9]$ and more profoundly when $\phi = 0.95$. In Table 3 we evaluate the following performance measures across 1000 replications of model (29):

$$\text{AvgCov} = \frac{1}{z} \sum_{i \in S} P\left(\beta_i \in \text{CI}_i\right), \quad \text{AvgLength} = \frac{1}{z} \sum_{i \in S} \text{length}\left(\text{CI}_i\right), \tag{31}$$

where $z = s_0$, $S = S_0$ when $\beta_i \in S_0$, $z = p - s_0$, $S = S_0^c$, when $\beta_i \in S_0^c$, $\forall\, i = 1, \ldots, p$ and CI_i is a two-sided confidence interval for either $\beta_i \in S_0$ or $\beta_i \in S_0^c$, denoted in (25). The findings of Table 3 can be summarised as follows: For the case of $\phi = 0$, both performance measures for DEBIASED LASSO and DEBIASED GLS LASSO are closely related, and as p, T increase, the two methods report narrow confidence intervals with high average coverage rate (that in many instances approaches 95%), as our theoretical findings suggest. In cases where $\phi > 0$, we observe that the average coverage rate, of both methods, remains equivalently high in S_0 and S_0^c, where the DEBIASED LASSO reports higher average coverage rates compared to the DEBIASED GLS LASSO in certain cases. However, emphasis should be placed on the average confidence interval length. Specifically, the average length of the DEBIASED LASSO estimator increases sharply as ϕ increases, indicative of inefficient estimation. Contrary to the former, DEBIASED GLS LASSO reports stable and narrow interval lengths as ϕ increases, which decreases significantly as T increases, which is in line with the findings of Table 2.

Table 1: Relative to GLS LASSO average RMSE, $|S_0| = 3$. The values in the parenthesis report the average RMSE of GLS LASSO

Panel I																
	p/T			200					500					1000		
ϕ		0	0.5	0.8	0.9	0.95	0	0.5	0.8	0.9	0.95	0	0.5	0.8	0.9	0.95
LASSO/GLS LASSO	100	0.995	1.251	2.006	2.728	3.436	0.998	1.291	1.932	2.631	3.723	0.999	1.316	2.290	2.934	3.942
GLS LASSO		(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
LASSO/GLS LASSO	200	1.003	1.270	2.124	2.716	3.353	0.999	1.283	1.956	2.835	3.724	0.999	1.317	2.175	2.610	3.811
GLS LASSO		(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
LASSO/ GLS LASSO	500	0.996	1.300	2.089	2.675	3.455	1.000	1.247	2.172	3.154	3.681	1.000	1.313	1.898	2.639	3.821
GLS LASSO		(0.02)	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Panel II																
DEBIASED GLS	100	1.00	1.29	2.07	2.83	3.68	1.00	1.29	2.11	2.97	4.05	1.00	1.30	2.13	3.04	4.25
DEBIASED GLS		(0.07)	(0.06)	(0.05)	(0.05)	(0.05)	(0.04)	(0.04)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.02)	(0.02)	(0.02)
DEBIASED LASSO/ DEBIASED GLS	200	1.01	1.30	2.08	2.89	3.76	1.00	1.30	2.13	2.98	4.07	1.00	1.30	2.14	3.06	4.23
DEBIASED GLS		(0.07)	(0.06)	(0.05)	(0.05)	(0.05)	(0.04)	(0.04)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.02)	(0.02)	(0.02)
LASSO/ DEBIASED GLS	500	1.01	1.28	2.10	2.93	3.69	1.00	1.32	2.13	2.95	4.09	1.00	1.31	2.19	3.10	4.23
DEBIASED GLS		(0.06)	(0.05)	(0.05)	(0.05)	(0.05)	(0.04)	(0.04)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.02)	(0.02)	(0.02)

In Figure 2 we graph the absolute bias of GLS LASSO, defined in (16) and (15) respectively, of LASSO, defined in (13) and (14), and of the *sub-optimal GLS Lasso* considering a λ, in the preliminary estimate $\widetilde{\boldsymbol{\beta}}$ which maximizes rather than minimising the following test-loss function w.r.t. the regularisation parameter, λ, for sample sizes $T = [500, 1000]$, $p = [100, 200, 500]$ and $u_t = 0.9u_{t-1} + \varepsilon_t$. Notice that the absolute bias for the GLS LASSO are bounded from below as Theorem 1 indicates, while a lower bound for the LASSO is provided in Lemma 1. In empirical problems, cross-validation does not always yield an optimal selection of λ, because one has to select between true model recovery (interpretable model) and a parsimonious (highly regularised) model. The latter can potentially be an issue because the selection of λ controls the lower bound of both the prediction and estimation error, see, for example, (15)–(16). To

Table 2: Average Size-adjusted power, for $|S_0| = 3$, the reported values in the parenthesis indicate the standard errors of each debiased estimator, $\widehat{\sigma}_u^2 \left[(\widehat{\boldsymbol{\Theta}} \widehat{\boldsymbol{\Sigma}} \widehat{\boldsymbol{\Theta}})_{i,i} \right]^{1/2}$, reported as (s.e.)

Size adjusted power and standard errors																
	p/T			200					500					1000		
ϕ		0	0.5	0.8	0.9	0.95	0	0.5	0.8	0.9	0.95	0	0.5	0.8	0.9	0.95
GLS LASSO	100	0.84	0.85	0.87	0.87	0.86	0.90	0.91	0.92	0.92	0.93	0.92	0.94	0.94	0.94	0.95
(s.e.)		(0.07)	(0.06)	(0.06)	(0.06)	(0.06)	(0.04)	(0.04)	(0.04)	(0.03)	(0.04)	(0.03)	(0.03)	(0.02)	(0.02)	(0.02)
LASSO		0.84	0.81	0.71	0.66	0.57	0.90	0.87	0.83	0.77	0.71	0.92	0.92	0.89	0.84	0.79
(s.e.)		(0.07)	(0.08)	(0.11)	(0.15)	(0.20)	(0.04)	(0.05)	(0.07)	(0.10)	(0.14)	(0.03)	(0.04)	(0.05)	(0.07)	(0.10)
GLS LASSO	200	0.85	0.87	0.89	0.89	0.90	0.91	0.92	0.93	0.94	0.94	0.94	0.95	0.96	0.96	0.96
(s.e.)		(0.07)	(0.06)	(0.06)	(0.06)	(0.06)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.03)	(0.03)	(0.02)	(0.02)	(0.02)
LASSO		0.85	0.83	0.76	0.69	0.61	0.92	0.89	0.85	0.80	0.73	0.94	0.93	0.90	0.86	0.79
(s.e.)		(0.07)	(0.08)	(0.11)	(0.15)	(0.20)	(0.04)	(0.05)	(0.07)	(0.10)	(0.14)	(0.03)	(0.04)	(0.05)	(0.07)	(0.10)
GLS LASSO	500	0.85	0.86	0.88	0.89	0.88	0.91	0.92	0.93	0.93	0.93	0.93	0.94	0.95	0.95	0.95
(s.e.)		(0.07)	(0.07)	(0.06)	(0.06)	(0.06)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.03)	(0.03)	(0.02)	(0.02)	(0.03)
LASSO		0.85	0.83	0.77	0.70	0.59	0.91	0.90	0.85	0.80	0.74	0.93	0.93	0.90	0.85	0.80
(s.e.)		(0.07)	(0.08)	(0.12)	(0.15)	(0.20)	(0.04)	(0.05)	(0.07)	(0.10)	(0.13)	(0.03)	(0.04)	(0.05)	(0.07)	(0.10)

Table 3: Average coverage Rates, and lengths of CI of the debiased estimators, for $|S_0| = 3$.

		p/T			200					500					1000		
ϕ			0	0.5	0.8	0.9	0.95	0	0.5	0.8	0.9	0.95	0	0.5	0.8	0.9	0.95
DEBIASED LASSO	AvgCov S_0	100	0.912	0.915	0.924	0.934	0.943	0.928	0.930	0.938	0.943	0.943	0.926	0.928	0.936	0.941	0.941
	AvgCov S_0^c		0.954	0.955	0.955	0.954	0.952	0.953	0.953	0.953	0.953	0.953	0.951	0.952	0.952	0.952	0.952
	AvgLength		0.277	0.316	0.446	0.592	0.770	0.175	0.201	0.288	0.391	0.531	0.124	0.143	0.205	0.280	0.385
	AvgLength^c		0.277	0.316	0.446	0.592	0.769	0.175	0.201	0.288	0.391	0.531	0.124	0.143	0.205	0.280	0.385
DEBIASED GLS	AvgCov S_0		0.907	0.903	0.898	0.908	0.917	0.925	0.923	0.917	0.918	0.926	0.926	0.919	0.913	0.917	0.920
	AvgCov S_0^c		0.953	0.956	0.959	0.962	0.967	0.952	0.954	0.954	0.956	0.963	0.951	0.953	0.953	0.955	0.959
	AvgLength		0.275	0.250	0.224	0.224	0.233	0.175	0.157	0.138	0.135	0.139	0.124	0.111	0.097	0.094	0.094
	AvgLength^c		0.275	0.250	0.223	0.224	0.232	0.175	0.157	0.138	0.134	0.139	0.124	0.111	0.097	0.094	0.094
DEBIASED LASSO	AvgCov S_0	200	0.911	0.917	0.926	0.936	0.941	0.926	0.928	0.937	0.936	0.944	0.925	0.931	0.942	0.948	0.945
	AvgCov S_0^c		0.963	0.962	0.962	0.959	0.957	0.956	0.955	0.955	0.956	0.954	0.953	0.953	0.952	0.953	0.953
	AvgLength		0.279	0.319	0.449	0.594	0.772	0.176	0.202	0.289	0.391	0.530	0.124	0.143	0.205	0.280	0.385
	AvgLength^c		0.279	0.319	0.448	0.594	0.772	0.176	0.202	0.288	0.391	0.530	0.124	0.143	0.205	0.280	0.385
DEBIASED GLS	AvgCov S_0		0.907	0.893	0.890	0.896	0.905	0.923	0.917	0.911	0.914	0.927	0.923	0.926	0.921	0.919	0.927
	AvgCov S_0^c		0.962	0.965	0.970	0.972	0.975	0.955	0.957	0.958	0.963	0.968	0.952	0.954	0.956	0.957	0.963
	AvgLength		0.277	0.252	0.229	0.229	0.237	0.175	0.158	0.139	0.138	0.143	0.124	0.111	0.098	0.094	0.096
	AvgLength^c		0.277	0.251	0.228	0.228	0.237	0.175	0.158	0.139	0.138	0.142	0.124	0.111	0.098	0.094	0.096
DEBIASED LASSO	AvgCov S_0	500	0.896	0.906	0.923	0.928	0.931	0.917	0.928	0.930	0.936	0.939	0.926	0.927	0.941	0.945	0.943
	AvgCov S_0^c		0.974	0.974	0.971	0.967	0.965	0.963	0.961	0.963	0.962	0.958	0.958	0.958	0.955	0.957	0.956
	AvgLength		0.282	0.322	0.452	0.597	0.775	0.176	0.203	0.289	0.390	0.527	0.124	0.143	0.206	0.281	0.385
	AvgLength^c		0.282	0.322	0.452	0.597	0.775	0.176	0.203	0.289	0.390	0.527	0.124	0.143	0.206	0.281	0.385
DEBIASED GLS	AvgCov S_0		0.887	0.883	0.872	0.881	0.900	0.914	0.911	0.903	0.905	0.913	0.928	0.917	0.908	0.909	0.922
	AvgCov S_0^c		0.974	0.976	0.979	0.981	0.983	0.962	0.964	0.966	0.972	0.974	0.958	0.960	0.960	0.964	0.970
	AvgLength		0.280	0.257	0.236	0.237	0.250	0.176	0.159	0.140	0.142	0.145	0.124	0.112	0.098	0.096	0.100
	AvgLength^c		0.280	0.257	0.237	0.236	0.249	0.176	0.159	0.140	0.142	0.145	0.124	0.112	0.098	0.096	0.100

that end, GLS LASSO provides an asymptotic guarantee that even when the selection of λ is sub-optimal, GLS LASSO is preferable to using LASSO. Indicative cases of such behaviour are $T = 1000, p \in \{100, 200\}$, and $T = 500, p \in \{200, 500\}$. Furthermore, in the remaining sets of cases, of a sub-optimal selection of λ, GLS LASSO still reports smaller errors than LASSO but by a smaller margin. It is useful to note that as λ increases, the parameter estimation both by using GLS LASSO or LASSO become almost equivalent, indicative of $\widehat{s_0} \equiv 0$. The latter is expected, for values of $\lambda \to \|\widetilde{\boldsymbol{X}}' \widetilde{\boldsymbol{y}}/T\|_\infty$.

In Figure 3 we graph the density estimates of $\sqrt{T}(\widehat{\boldsymbol{b}} - \boldsymbol{\beta})$ corresponding to DEBIASED GLS LASSO starting from the left panel, the second panel corresponds to DEBIASED LASSO and the third panel illustrates the quantile-quantile (QQ) plots of the aforementioned. Figure 3 corresponds to the case where the sample size is the largest, i.e. $T = 1000$ and the number of

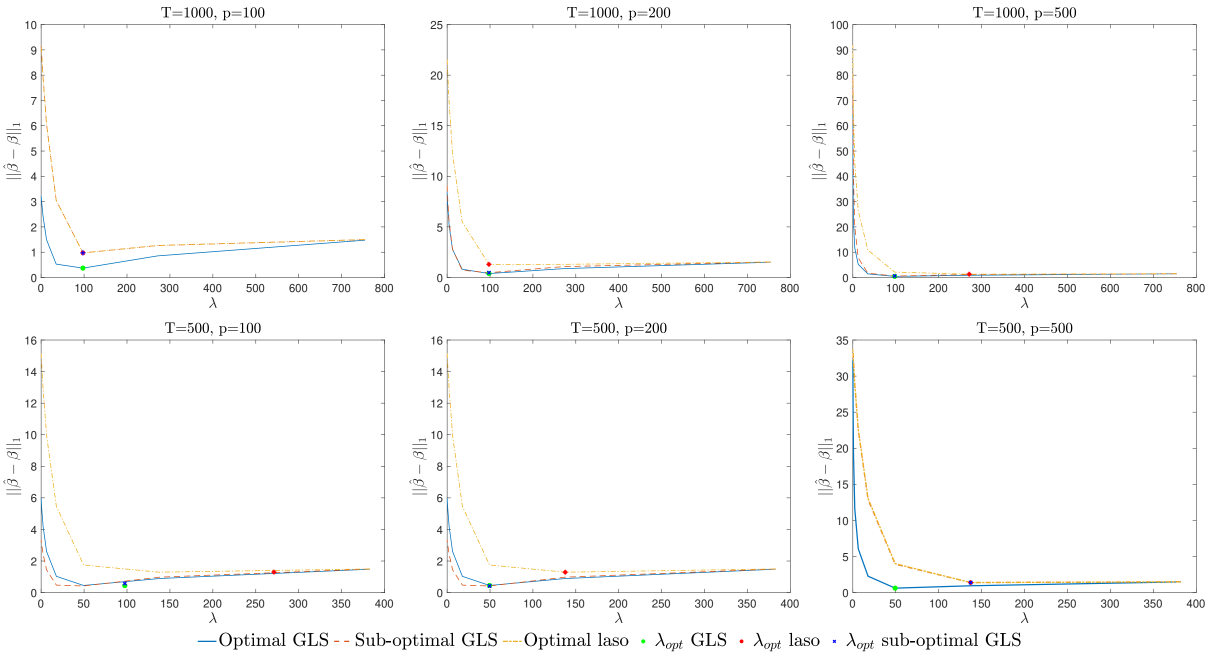


Figure 2: Estimation loss, $\|\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}\|_1$. Optimal GLS corresponds the GLS LASSO defined in (4) when the preliminary LASSO estimator, $\widetilde{\boldsymbol{\beta}}$ in (11) is obtained according to the optimal selection of the regularisation parameter via the cross-validation scheme highlighted in Section 5, while *Sub-optimal GLS* corresponds the GLS LASSO as in (4) when the preliminary LASSO estimator, $\widetilde{\boldsymbol{\beta}}$ is not obtained according to the maximisation of the CV loss, see e.g. (28). Further, *Optimal Lasso* corresponds to (11) selecting the optimal λ through the cross-validation scheme highlighted in Section 5, and λ_{opt} corresponds to the optimal selection of λ for each of the methods.

regressors is $p = 500$ for all $\phi = [0, 0.5, 0.8, 0.9, 0.95]$. A full set of graphs, for all the cases of the experimental design considered in this section, is available upon request.

In the first row of graphs in Figure 3 where $\phi = 0.9$, it is observed that extreme dependence between u_t and u_{t-1} is related to overdispersion of the distribution of $\sqrt{T}(\widehat{\boldsymbol{b}}-\boldsymbol{\beta})$, corresponding to the DEBIASED LASSO estimator, while in the same case, departures from the standard normal are observed, with the empirical data showing evidence of heavy tails. On the contrary, the empirical distribution of the DEBIASED GLS LASSO estimator is closer to the Standard Normal, as it is indicative from the QQ-plot of the same row. However, this effect fades as ϕ takes values closer to 0, indicative of the case where $u_t \sim$ i.i.d.$\mathcal{N}(0, 1)$. In the first row of Figure 3, DEBIASED LASSO appears slightly skewed left with heavy tails to persist indicative of departures from the standard normal distribution, while DEBIASED GLS LASSO appears to approach the standard normal distribution, similar behaviour for both methods is reported in the third row. Finally, in the case where $\phi = 0$ the two methods coincide supporting the evidence in Tables 2 and 3.

7 Forecasting Example

In this section we explore the practical applicability of the proposed methodology. Hence, our primary interest is to examine whether the GLSLASSO method outperforms in forecasting ability terms the standard LASSO estimator, and to what extent. More generally, we use this

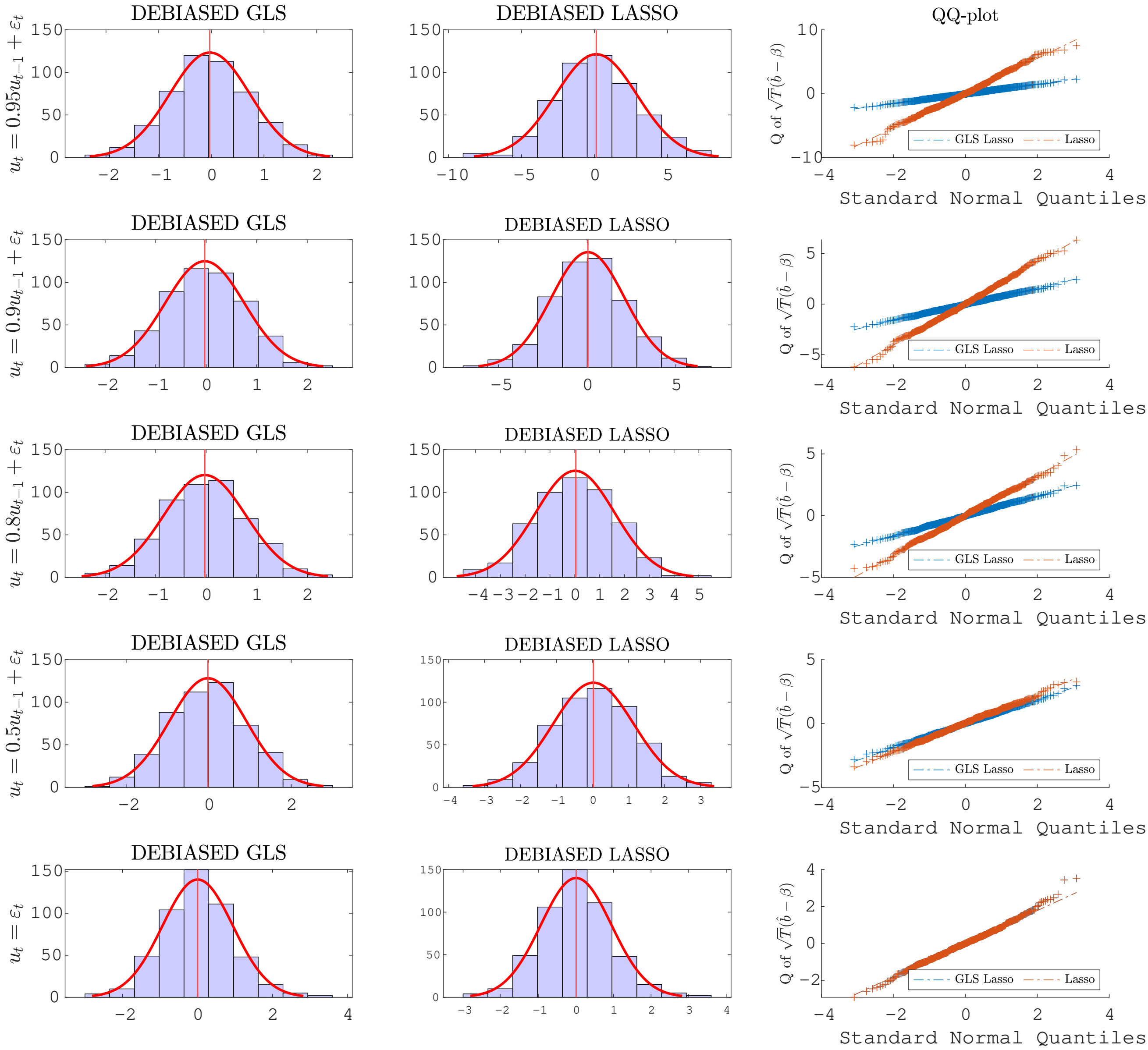


Figure 3: Histogram and QQ-plot of $\sqrt{T}(\widehat{\boldsymbol{b}} - \boldsymbol{\beta})$, where $\widehat{\boldsymbol{b}}$ corresponds to the DEBIASED GLS LASSO estimator defined in (20), in the first column of Figure 3 and to the DEBIASED LASSO estimator in the second column of Figure 3, for the case of $T = 1000$, $p = 500$ and all cases of ϕ and degree of sparsity $s_0 = 3$.

empirical example to assess whether high-dimensional forecasting methods that ignore serial error auto-correlation, report, on average, inferior forecasting ability. Our application overall consists of forecasting US industrial production using a large set of macroeconomic variables.

We use the FRED-MD dataset[3], a large collection of monthly macroeconomic data, updated in real-time and reported at different vintages. This dataset is widely used in macro-econometric applications, where the focus is forecasting with high-dimensional methods. We follow Fan et al.

[3]The dataset is available here https://research.stlouisfed.org/econ/mccracken/fred-databases/; for further details see McCracken and Ng (2016)

(2023) and use the series reported in the August 2022 vintage. Our sample extends from January 1960 to December 2019 (720 observations) and we use only those series that have observations in the above time span (122 variables). To proceed with our forecasting exercise, we use the transformations in McCracken and Ng (2016) to make all variables stationary.

In order to examine both the gains of exploring all relevant information in the dataset and also the potential impact of serial error auto-correlation, we construct one-step ahead forecasts for the first-order difference of the logarithm of the monthly industrial production index (IP, growth rate): $y_{\text{IP},t}$.

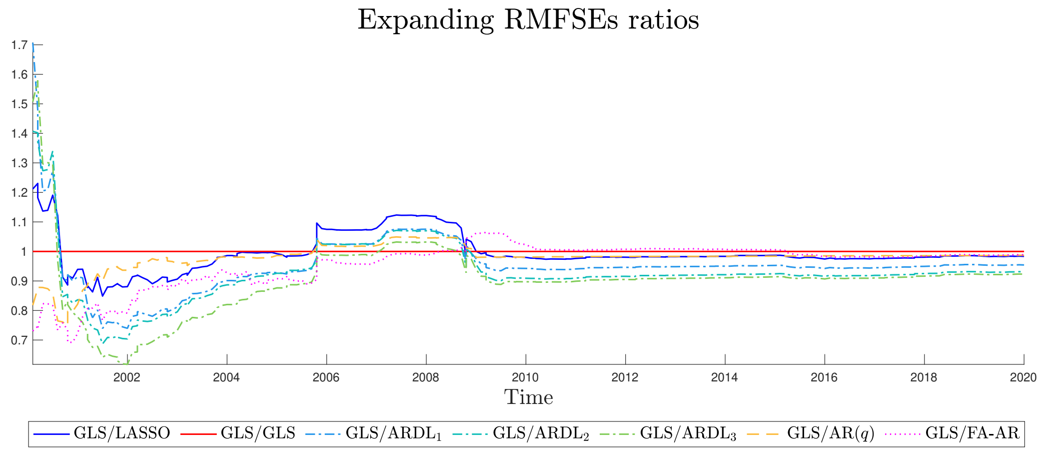
We compare the proposed GLS LASSO method with four different models, that are commonly used in high-dimensional applications: LASSO; Autoregressive Distributed Lag Models (ARDL (q, r)) models, an autoregressive model ($AR(q)$; typically used as the most parsimonious model) and finally, a factor augmented autoregressive model (FA-AR (q)). A detailed description of the models considered is relegated to the Supplement.

The forecasts are based on a rolling-window recursive out-of-the-sample exercise, where the window is of a fixed length of 480 observations, starting in January 1960. Therefore, the one-step ahead forecasts start on January 1990 and the last one is obtained in January 2020, for a total of 240 one-step ahead forecasts. Note that the AR(q) model only considers information concerning the own past of the variable of interest. Contrary to the latter, LASSO uses a sparse combination of the set of variables, GLS LASSO in addition to a sparse combination of the set of variables, uses the fit of an approximated AR(q) model on the estimated residuals. ARDL (q, r), $q = r = 1, 2, 3$ augments both the set of variables and the variable of interest with one, two and three lags respectively, and similarly to LASSO, it utilises a sparse combination of the augmented set. Finally, FA-AR(q) utilises a reduced representation of $(y_{IP,t}, \boldsymbol{x}_t)$ through principal components, in addition to information concerning the own past of the variable of interest.

In Figure 3 we report the ratios of expanding root mean squared forecasting errors (RMSFE) of the GLS LASSO model against the expanding RMSFE of the competing methods over the forecasting period, computed as:

$$\text{RMFSE}_j = \sqrt{\frac{1}{j-480}\sum_j \left(y_{IP,j+1} - \widehat{y}^{(k)}_{IP,j+1|j}\right)^2}, \quad \text{rRMFSE}_j = \frac{\text{RMFSE}_{\text{GLS}}}{\text{RMFSE}_{\text{k}}},$$

where $j = 481, \ldots, 720$ and $k = 2, 3, 4, 5$ the number of competing models considered. Several conclusions emerge from the plot. First, the GLS LASSO is, in general, superior in terms of forecasting ability across the different competing methods. Specifically, GLS LASSO over-performs considerably ARDL $(1, 1)$, ARDL $(2, 2)$ and by the largest margin ARDL $(3, 3)$ across the majority of the periods, with some exceptions, where ARDL methods appear to over-perform GLS LASSO in the initial periods of estimation and the years preceding the financial crisis of 2008. This can be a surprising finding given that both methods consider approximation of the autocorrelation structure in the residuals and given the efficiency of ARDL models in low dimensions, see Baillie et al. (2022), for an excellent discussion. In our high-dimensional example, the above



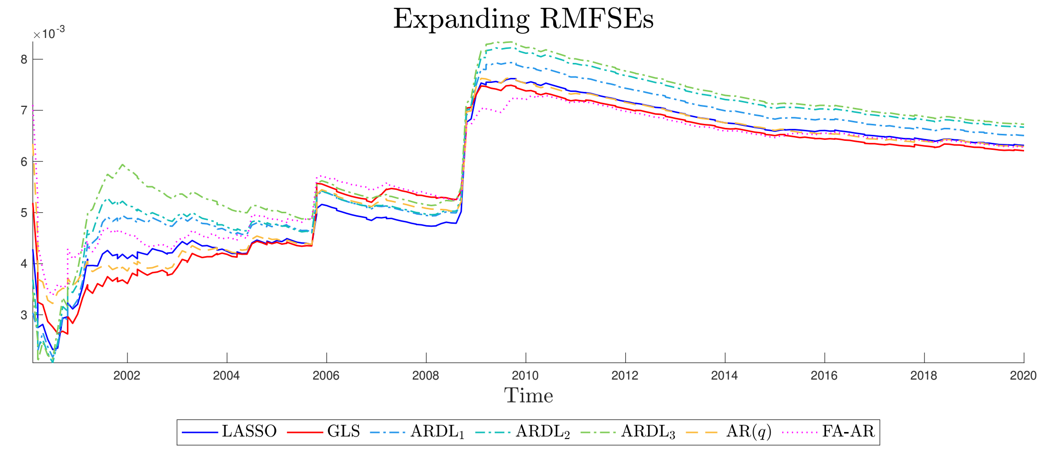


Figure 4: The upper panel of this Figure presents the relative RMFSE and the lower part presents the expanding RMFSE_k, where k is indicative of model number.

finding can be potentially explained from the significant increase in dimensionality an ARDL (q, r) model requires; a dimensionality increase from p to $(q \times p) + r$. A proper examination of the two methods in high-dimensions would be of great merit, especially given the increased interest in high-dimensional methods for time series. We leave this for future work.

Further, we find that GLS LASSO outperforms both the FA-AR(q), and AR(q) models in the majority of the periods. At the initial forecasting period we find that the forecasting performance of FA-AR(q), and AR(q) is dominated by the other models. We further find that in periods of financial crisis such as in 2008, GLS LASSO slightly under-performs compared to the AR(q) and methods such as ARDL, potentially, because direct additions of lags from the dependent variable capture the idiosyncrasies of the periods more efficiently, and for the case of ARDL, albeit the extra cost of increased dimensionality. The inclusion of an unobserved factor appears to provide some forecasting benefit in the period directly after the global financial crisis and for a limited time only.

Lastly, we find that GLS LASSO performs better than the LASSO in the majority of periods excluding the period from 2006 up to 2009 that includes the global financial crisis. At this stage we emphasize that our primary interest in this empirical example was to verify that the efficiency gains observed from our results in Section 6, also translate to an increased forecasting ability.

In Table 4, we further consider the average RMSFE across periods:

$$\text{RMSFE}_k = \sqrt{\frac{1}{720-480}\sum_{t=481}^{720}\left(y_{IP,t+1} - \widehat{y}^{(k)}_{IP,t+1|t}\right)^2},$$

where $k = 1, \ldots, 5$ indicates the model which was used to forecast $\widehat{y}^{(k)}_{IP,T+1|T}$. We find that across the whole forecasting period, GLS LASSO outperforms all other alternative models.

GLS/	LASSO	ARDL(1,1)	ARDL(2, 2)	ARDL (3,3)	AR(q)	FA-AR(q)
0.0062	0.983	0.954	0.931	0.923	0.987	0.988

Table 4: This Table reports the ratios of the average RMSFE of the GLS LASSO model against the average RMSFE of the other models produced across the forecasting period. The left cell reports the average RMSFE of the GLS LASSO. Entries (< 1) indicate better forecasting ability of GLS LASSO.

In summary our results and the previous discussion suggest that first GLS LASSO indeed outperforms the standard LASSO in the majority of the period examined and more generally, that, ignoring serial error correlation can have a negative impact in the precision of forecasting.

8 Discussion

This paper provides a complete inferential methodology for high dimensional linear regressions with serially correlated error terms. We propose GLS-type of estimators in the high-dimensional framework, derive a DEBIASED GLS LASSO estimator, and establish its asymptotic normality. Theoretical results lead to valid confidence intervals regardless of the dimensionality of the parameters and the degree of autocorrelation, under stationarity of the error and covariate processes. We derive non-asymptotic bounds using Bernstein-type of inequalities derived in Dendramis et al. (2021), allowing for a degree of flexibility in our model compared to what classical assumptions typically allow. Finally, our method explores the underlying structure of the process in the error term, introducing a flexible way of dealing with autocorrelation in high-dimensional models, without a prior knowledge of neither its existence nor its structure. We support this statement both theoretically and with our simulation results, where we show that even when autocorrelation is absent the DEBIASED LASSO and DEBIASED GLS LASSO are asymptotically equivalent.

There are a number of interesting avenues for future work. For example, a natural extension would be to generalise our method using not only ℓ_1-regularised models or ℓ_2 (e.g. Zhang and Politis (2023)), but a combination of the two, e.g. the elastic net, covering a wider class of penalised models. Further, since our assumptions allow for a wide class of dependent stochastic processes, one could consider heteroskedastic error processes, see, for example, the framework in Chronopoulos et al. (2022), where

$$u_t = h_t \epsilon_t, \quad t = 1, \ldots, T, \tag{32}$$

where h_t is a persistent scaling factor, smoothly varying in time, while $\epsilon_t \sim \ i.i.d.$

Acknowledgements: For discussions and/or comments we thank Jianqing Fan, Dacheng Xiu, Zacharias Psaradakis, Yunyi Zhang, Takashi Yamagata.

References

Adamek, R., S. Smeekes, and I. Wilms (2023): "Lasso inference for high-dimensional time series," *Journal of Econometrics*, 235(2), 1114–1143.

Amemiya, T. (1973): "Generalized least squares with an estimated autocovariance matrix," *Econometrica*, 41, 723–732.

Andrews, D. W. (1984): "Non-strong mixing autoregressive processes," *Journal of Applied Probability*, 21(4), 930–934.

Babii, A., R. T. Ball, E. Ghysels, and J. Striaukas (2023): "Machine learning panel data regressions with heavy-tailed dependent data: Theory and application," *Journal of Econometrics*, 237(2), 105315.

Babii, A., E. Ghysels, and J. Striaukas (2020): "Inference for High-Dimensional Regressions With Heteroskedasticity and Auto-correlation," *Available at SSRN 3615718*.

——— (2022): "Machine learning time series regressions with an application to nowcasting," *Journal of Business & Economic Statistics*, 40(3), 1094–1106.

——— (2024): "High-dimensional Granger causality tests with an application to VIX and news," *Journal of Financial Econometrics*, 22(3), 605–635.

Baillie, R. T., F. X. Diebold, G. Kapetanios, and K. H. Kim (2022): "On robust inference in time series regression," *arXiv preprint arXiv:2203.04080*.

Basu, S., and G. Michailidis (2015): "Regularized estimation in sparse high-dimensional time series models," *The Annals of Statistics*, 43(4), 1535–1567.

Berk, R., L. Brown, A. Buja, K. Zhang, and L. Zhao (2013): "Valid post-selection inference," *The Annals of Statistics*, 41(2), 802–837.

Bickel, P. J., Y. Ritov, and A. B. Tsybakov (2009): "Simultaneous analysis of Lasso and Dantzig selector," *The Annals of Statistics*, 37(4), 1705–1732.

Brockwell, P. J., R. Dahlhaus, and A. A. Trindade (2005): "Modified Burg algorithms for multivariate subset autoregression," *Statistica Sinica*, 15, 197–213.

Bühlmann, P., and S. Van De Geer (2011): *Statistics for high-dimensional data: methods, theory and applications*. Springer Science & Business Media.

Burg, J. P. (1968): "A new analysis technique for time series data," *Paper presented at NATO Advanced Study Institute on Signal Processing, Enschede, Netherlands, 1968*.

Burman, P., E. Chow, and D. Nolan (1994): "A cross-validatory method for dependent data," *Biometrika*, 81(2), 351–358.

Chernozhukov, V., W. K. Härdle, C. Huang, and W. Wang (2021): "Lasso-driven inference in time and space," *The Annals of Statistics*, 49(3), 1702–1735.

Chronopoulos, I., L. Giraitis, and G. Kapetanios (2022): "Choosing between persistent and stationary volatility," *The Annals of Statistics*, 50(6), 3466 – 3483.

Chudik, A., G. Kapetanios, and M. H. Pesaran (2018): "A one covariate at a time, multiple testing approach to variable selection in high-dimensional linear regression models," *Econometrica*, 86(4), 1479–1512.

Davidson, J. (1994): *Stochastic limit theory: An introduction for econometricians*. OUP Oxford.

Dendramis, Y., L. Giraitis, and G. Kapetanios (2021): "Estimation of time-varying covariance matrices for large datasets," *Econometric Theory*, 37(6), 1100–1134.

Fan, J., R. P. Masini, and M. C. Medeiros (2023): "Bridging factor and sparse models," *The Annals of Statistics*, 51(4), 1692–1717.

Györfi, L., W. Härdle, P. Sarda, and P. Vieu (2013): *Nonparametric curve estimation from time series*, vol. 60. Springer.

Hamilton, J. D. (1994): *Time series analysis*. Princeton university press.

Hsu, N.-J., H.-L. Hung, and Y.-M. Chang (2008): "Subset selection for vector autoregressive processes using lasso," *Computational Statistics & Data Analysis*, 52(7), 3645–3657.

Javanmard, A., and A. Montanari (2014): "Confidence intervals and hypothesis testing for high-dimensional regression," *The Journal of Machine Learning Research*, 15(1), 2869–2909.

Kapetanios, G., and Z. Psaradakis (2016): "Semiparametric sieve-type generalized least squares inference," *Econometric Reviews*, 35(6), 951–985.

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

Kock, A. B. (2016a): "Consistent and conservative model selection with the adaptive lasso in stationary and nonstationary autoregressions," *Econometric Theory*, 32(1), 243–259.

——— (2016b): "Oracle inequalities, variable selection and uniform inference in high-dimensional correlated random effects panel data models," *Journal of Econometrics*, 195(1), 71–85.

Kock, A. B., and L. Callot (2015): "Oracle inequalities for high dimensional vector autoregressions," *Journal of Econometrics*, 186(2), 325–344.

Leeb, H., and B. M. Pötscher (2005): "Model selection and inference: Facts and fiction," *Econometric Theory*, 21(1), 21–59.

Lutkepohl, H. (1997): "Handbook of matrices.," *Computational statistics and Data analysis*, 2(25), 243.

Masini, R. P., M. C. Medeiros, and E. F. Mendes (2022): "Regularized estimation of high-dimensional vector autoregressions with weakly dependent innovations," *Journal of Time Series Analysis*, 43(4), 532–557.

McCracken, M. W., and S. Ng (2016): "FRED-MD: A monthly database for macroeconomic research," *Journal of Business & Economic Statistics*, 34(4), 574–589.

Meinshausen, N., and B. Yu (2009): "Lasso-type recovery of sparse representations for high-dimensional data," *The Annals of Statistics*, 37(1), 246–270.

Nardi, Y., and A. Rinaldo (2011): "Autoregressive process modeling via the lasso procedure," *Journal of Multivariate Analysis*, 102(3), 528–549.

Perron, P., and E. González-Coya (2022): "Feasible GLS for Time Series Regression," *Manuscript, Department of Economics, Boston University.*

Racine, J. (2000): "Consistent cross-validatory model-selection for dependent data: hv-block cross-validation," *Journal of Econometrics*, 99(1), 39–61.

Raskutti, G., M. J. Wainwright, and B. Yu (2010): "Restricted eigenvalue properties for correlated Gaussian designs," *The Journal of Machine Learning Research*, 11, 2241–2259.

Shao, J. (1993): "Linear model selection by cross-validation," *Journal of the American Statistical Association*, 88(422), 486–494.

Taylor, J., and R. J. Tibshirani (2015): "Statistical learning and selective inference," *Proceedings of the National Academy of Sciences*, 112(25), 7629–7634.

Tian, X., and J. Taylor (2017): "Asymptotics of selective inference," *Scandinavian Journal of Statistics*, 44(2), 480–499.

Tibshirani, R. (1996): "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 58(1), 267–288.

Tibshirani, R. J., A. Rinaldo, R. Tibshirani, and L. Wasserman (2018): "Uniform asymptotic inference and the bootstrap after model selection," *The Annals of Statistics*, 46(3), 1255–1287.

Van de Geer, S., P. Bühlmann, 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*, 36(2), 614–645.

Van De Geer, S. A., and P. Bühlmann (2009): "On the conditions used to prove oracle results for the Lasso," *Electronic Journal of Statistics*, 3, 1360–1392.

Wang, H., G. Li, and C.-L. Tsai (2007): "Regression coefficient and autoregressive order shrinkage and selection via the lasso," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 69(1), 63–78.

Wong, K. C., Z. Li, and A. Tewari (2020): "Lasso guarantees for β-mixing heavy-tailed time series," *The Annals of Statistics*, 48(2), 1124 – 1142.

Zhang, C.-H., and J. Huang (2008): "The sparsity and bias of the lasso selection in high-dimensional linear regression," *The Annals of Statistics*, 36(4), 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.

Zhang, Y., and D. N. Politis (2022): "Ridge regression revisited: Debiasing, thresholding and bootstrap," *The Annals of Statistics*, 50(3), 1401–1422.

——— (2023): "Debiased and thresholded ridge regression for linear models with heteroskedastic and correlated errors," *Forthcoming in Journal of the Royal Statistical Society: Series B (Statistical Methodology)*.

Zhao, P., and B. Yu (2006): "On model selection consistency of Lasso," *The Journal of Machine Learning Research*, 7, 2541–2563.

Zheng, W. (2019): "*hv*-Block Cross Validation is not a BIBD: a Note on the Paper by Jeff Racine (2000)," *arXiv preprint arXiv:1910.08904*.

Supplement to "High Dimensional Generalised Penalised Least Squares"

Ilias Chronopoulos §1, Katerina Chrysikou ¶ 2 and George Kapetanios ‖3

[1]Essex Business School, University of Essex

[2]School of Business, University of Leicester

[3] King's Business School, King's College London

This Supplement provides proofs of the theoretical results given in the text of the main paper. It is organised as follows: Section A provides proofs of Lemma 1, Corollary 1, Theorem 1 and Corollary 2 of the main paper. Section B provides the proofs of Theorem 2 and Theorem 3 of the main paper. Section C provides auxiliary technical lemmas and D provides proofs of auxiliary technical lemmas. Section E contains supplementary simulation material. Section F contains supplementary material to Section 7 of the main paper.

Formula numbering in this supplement includes the section number, e.g. (8.1), and references to lemmas are signified as "Lemma 8.#", "Lemma 9.#", e.g. Lemma 8.1. Formula numbering in the Notation section is signified with the letter N.#, e.g. (N.1). Theorem references to the main paper are signified, e.g. as Theorem 1, while equation references are signified as, e.g. (1).

In the proofs, C, d, e stand for generic positive constants which may assume different values in different contexts.

Notation

For any vector $\boldsymbol{x} \in \mathbb{R}^n$, we denote the ℓ_p-, ℓ_∞- and ℓ_0- norms, as $\|\boldsymbol{x}\|_p = \left(\sum_{i=1}^n |x_i|^p\right)^{1/p}$, $\|\boldsymbol{x}\|_\infty = \max_{i=1,\dots,n} |x_i|$ and $\|\boldsymbol{x}\|_0 = \sum_{i=1}^n \mathbf{1}_{(x_i \neq 0)} = \text{supp}(\boldsymbol{x})$, respectively. Throughout this supplement, $\mathbb{R}$ denotes the set of real numbers. We denote the cardinality of a set S_0 by $s_0 = |S_0|$, while S_0^c denotes its complement. For any $\boldsymbol{x} \in \mathbb{R}^n$, sign(x) denotes the sign function applied to each component of $\boldsymbol{x}$, where $\text{sign}(x) = 0$ if $x = 0$. We use "$\rightarrow_P$" to denote convergence in probability. For two deterministic sequences a_n and b_n we define asymptotic proportionality, "$\asymp$", by writing $a_n \asymp b_n$, if there exist constants $0 < a_1 \leq a_2$ such that $a_1 b_n \leq a_n \leq a_2 b_n$ for all $n \geq 1$. For a real number a, $\lfloor a \rfloor$ denotes the largest integer no greater than a. For a symmetric matrix $\boldsymbol{B}$, we denote its minimum and maximum eigenvalues by $\Lambda_{\min}(\boldsymbol{B})$ and $\Lambda_{\max}(\boldsymbol{B})$ respectively.

§ilias.chronopoulos@essex.ac.uk

¶ac988@leicester.ac.uk . Corresponding author

‖george.kapetanios@kcl.ac.uk

For a general (not necessarily square) $n \times m$ matrix $\boldsymbol{B} = (b_{ij})$, $\|\boldsymbol{B}\|_1$, $\|\boldsymbol{B}\|_F$, $\|\boldsymbol{B}\|_\infty$ are its ℓ_1-, ℓ_F-, ℓ_∞-norm respectively. In particular, $\|\boldsymbol{B}\|_1 = \max_j \sum_i |b_{ij}|$, $\|\boldsymbol{B}\|_F = [\text{tr}(\boldsymbol{B}\boldsymbol{B}')]^{1/2}$, $\|\boldsymbol{B}\|_\infty = \max_i \sum_j |b_{ij}|$ and tr$(\boldsymbol{B})$= $\sum_{i=1}^n b_{ii}$. We further write that, $\widetilde{\boldsymbol{\beta}}$ is the LASSO estimator obtained from the solution of (11) in the main paper, $\widehat{\boldsymbol{\beta}}$ is the *GLS LASSO* estimator obtained from the solution (4), given some estimates of the autoregressive parameters $\widehat{\boldsymbol{\phi}} = (\widehat{\phi}_1, \dots, \widehat{\phi}_q)'$ and $\boldsymbol{\beta}$ is a $p \times 1$ parameter vector containing the true parameters.

We define $\widetilde{\boldsymbol{x}}_t = \boldsymbol{x}_t - \sum_{j=1}^q \widehat{\phi}_j \boldsymbol{x}_{t-j}$, and its matrix counterpart, $\widetilde{\boldsymbol{X}} = \widehat{\boldsymbol{L}}\boldsymbol{X}$, where $\widehat{\boldsymbol{L}}$ is a $(T-q) \times T$ matrix defined as:

$$\widehat{\boldsymbol{L}} = \begin{pmatrix} -\widehat{\phi}_q & -\widehat{\phi}_{q-1} & \cdots & \cdots & -\widehat{\phi}_1 & 1 & \cdots & 0 & 0 \\ 0 & -\widehat{\phi}_q & -\widehat{\phi}_{q-1} & \cdots & \cdots & -\widehat{\phi}_1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & -\widehat{\phi}_q & -\widehat{\phi}_{q-1} & \cdots & -\widehat{\phi}_1 & 1 \end{pmatrix}, \quad \widehat{\boldsymbol{\phi}} = \begin{pmatrix} \widehat{\phi}_1 \\ \vdots \\ \widehat{\phi}_q \end{pmatrix}. \tag{N.1}$$

Further, following the same argument that we used to define $\widehat{\boldsymbol{L}}$, we define $\boldsymbol{L}$, a $(T-q) \times T$ matrix, with $\widehat{\phi}_j$ replaced by ϕ_j, denoting the autoregressive parameters, $\forall\ j = 1, \dots, q$. As a consequence, we define $\boldsymbol{X}^* = \boldsymbol{L}\boldsymbol{X}$.

A Proofs of Lemma 1, Corollary 1 & 2, and Theorem 1 of the main paper

This section contains the proofs of the results of Section 2 of the main paper.

Proof of Lemma 1

In Lemma 1 we analyse the asymptotic properties of the LASSO estimator under mixing assumptions, e.g. Assumption 1 and offer non-asymptotic bounds for the empirical process and prediction errors, as seen in (13)–(14) of the main paper. Re-arranging the basic inequality, as in Lemma 6.1 of Bühlmann and Van De Geer (2011), we obtain

$$\frac{1}{T}\left\|\boldsymbol{X}\left(\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta}\right)\right\|_2^2 \leq \frac{2}{T}\boldsymbol{u}'\boldsymbol{X}\left(\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta}\right) + \lambda(\|\boldsymbol{\beta}\| - \|\widetilde{\boldsymbol{\beta}}\|). \tag{A.1}$$

The "empirical process" part of the right hand side of (A.1), $2\boldsymbol{u}'\boldsymbol{X}\left(\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta}\right)/T$, can be bounded further in terms of the ℓ_1-norm, such that,

$$\frac{1}{T}\left|\boldsymbol{u}'\boldsymbol{X}\left(\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta}\right)\right| \leq \frac{2}{T}\|\boldsymbol{u}'\boldsymbol{X}\|_\infty \left\|\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta}\right\|_1. \tag{A.2}$$

The regularisation parameter λ is chosen such that $2\|\boldsymbol{u}'\boldsymbol{X}\|_\infty /T \leq \lambda$. Hence, we introduce the event

$$\mathcal{E} := \left\{T^{-1}\|\boldsymbol{u}'\boldsymbol{X}\|_\infty \leq \frac{\lambda_0}{2}\right\}, \tag{A.3}$$

which needs to hold with high probability, where $\lambda_0 \leq \lambda/2$. We proceed to illustrate the former. Define $\psi_{t,i} = \{x_{t,i}u_t\}$, using the identity $P(\mathcal{E}^c) = 1 - P(\mathcal{E})$ and the union bound we obtain

$$P(\mathcal{E}^c) = 1 - P\left(\max_i \left|\frac{1}{T}\sum_{t=1}^{T}\psi_{t,i} - E(\psi_{t,i})\right| \leq \frac{\lambda_0}{2}\right) \leq \sum_i P\left(\frac{1}{\sqrt{T}}\left|\sum_{t=1}^{T}\psi_{t,i} - E(\psi_{t,i})\right| > \frac{\lambda_0\sqrt{T}}{2}\right).$$

Notice that $\{\psi_{t,i}\}$ is α-mixing as a product of two α-mixing sequences, $\forall\ j = 1,\ldots,q$, $i = 1,\ldots,p$. Then by direct application of Lemma 1 of Dendramis et al. (2021), we obtain the following results for two cases of probability tails:

Case 1: The process $\psi_{t,i} \in \mathcal{T}(r)$, $r > 0$, then for $\lambda = \sqrt{\log p/T}$ we have

$$\sum_i P\left(\frac{1}{\sqrt{T}}\left|\sum_{t=1}^{T}\psi_{t,i} - E(\psi_{t,i})\right| > \frac{\lambda_0\sqrt{T}}{2}\right) \leq pc\left\{\exp\left(-c_2\left(\frac{\sqrt{T}\lambda_0}{4}\right)^2\right)\right.$$
$$\left. + \exp\left(-c_3\frac{\sqrt{T\log(p)}}{4\log^2 T}\right)^{\zeta}\right\} = A_1 + A_2, \tag{A.4}$$

for some c_2, c_3, $\zeta > 0$. It is sufficient to bound A_1, then for a proper selection of $c_2 > 0$, we have that $(c_2/2)^2 > 1 + \epsilon$, $\epsilon > 0$, and $c > 0$ we have that

$$A_1 = pc\exp\left(-c_2\left(\frac{\sqrt{\log p}}{4\sqrt{T}}\sqrt{T}\right)^2\right) = pc\left(\frac{1}{\exp(\log(p)(1+\epsilon))}\right) = cp^{-\epsilon}. \tag{A.5}$$

Case 2: The process $\psi_{t,i} \in \mathcal{H}(\theta)$, $\theta > 4$, then for $\lambda_0 = \frac{p^{\frac{1}{\theta}}}{T^{(\frac{\theta}{2}+1)}}$ we have

$$\sum_i P\left(\frac{1}{\sqrt{T}}\left|\sum_{t=1}^{T}\psi_{t,i} - E(\psi_{t,i})\right| > \frac{\lambda_0\sqrt{T}}{2}\right) \leq pc\left[\exp\left(-c_2\left(\frac{\sqrt{T}\lambda_0}{4}\right)^2\right) + \left(\frac{\lambda_0\sqrt{T}}{2}\right)^{-\theta}T^{-\frac{\theta}{2}-1}\right]$$
$$\leq pc\left(\frac{\lambda_0\sqrt{T}}{2}\right)^{-\theta}T^{-\frac{\theta}{2}-1} \to 0,$$

at a slower rate than A_1. Let Assumptions 1–3 hold, then Theorem 6.1 of Bühlmann and Van De Geer (2011) implies

$$\left\|\boldsymbol{X}\left(\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta}\right)\right\|_2^2 \leq 4\lambda^2 s_0\phi_0^{-2}, \quad \left\|\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta}\right\|_1 \leq 4\lambda s_0\phi_0^{-2},$$

with probability approaching 1 at different rates, depending on the probability tail case (Case 1 or Case 2), for a compatibility constant $\phi_0^{-2} \leq \frac{s_0\boldsymbol{\beta}' E\left(T^{-1}\sum_{t=1}^{T}\boldsymbol{x}_t\boldsymbol{x}_t'\right)\boldsymbol{\beta}}{\left\|\boldsymbol{\beta}_{s_0}\right\|_1^2}$. □

Proof of Corollary 1

In Corollary 1 we provide an asymptotic rate of convergence for the error between the estimated autoregressive parameters, $\widehat{\boldsymbol{\phi}}$ utilising $\widetilde{\boldsymbol{\beta}}$ and the true autoregressive parameters, $\boldsymbol{\phi}$.

First we define the following

$$\boldsymbol{u}=\left(u_{q+1},\ldots,u_T\right)', \quad \boldsymbol{U}=\begin{bmatrix} u_q & u_{q-1} & \cdots & u_1 \\ u_{q+1} & u_q & \cdots & u_2 \\ \vdots & \vdots & \vdots & \vdots \\ u_{T-1} & u_{T-2} & \cdots & u_{T-q} \end{bmatrix}, \tag{A.6}$$

where $\boldsymbol{u}$ is a $(T-q)\times 1$ vector and $\boldsymbol{U}$ is a $(T-q)\times q$ design matrix. Note that $\widetilde{\boldsymbol{\phi}}=(\widetilde{\phi}_1,\ldots,\widetilde{\phi}_q)'$ denotes the OLS estimate of the regression coefficients in the following regression:

$$\boldsymbol{u}=\boldsymbol{U}\widetilde{\boldsymbol{\phi}}+\boldsymbol{\varepsilon}, \tag{A.7}$$

where $\boldsymbol{\varepsilon}=(\varepsilon_1,\ldots,\varepsilon_{T-q})'$, follows Assumption 1 and $q<\infty$. Similarly to (A.6), we define $\widehat{\boldsymbol{u}}$ and $\widehat{\boldsymbol{U}}$, and consider $\widehat{\boldsymbol{\phi}}=(\widehat{\phi}_1,\ldots,\widehat{\phi}_q)'$, the OLS estimates of an AR(q) regression, using $(\widehat{\boldsymbol{u}},\widehat{\boldsymbol{U}})$ instead of $(\boldsymbol{u},\boldsymbol{U})$ and $\widehat{\boldsymbol{\phi}}$ instead of $\widetilde{\boldsymbol{\phi}}$. Then, we can write that

$$\widehat{\boldsymbol{\phi}}=\left(\widehat{\boldsymbol{U}}'\widehat{\boldsymbol{U}}\right)^{-1}\widehat{\boldsymbol{U}}'\widehat{\boldsymbol{u}}, \quad \widetilde{\boldsymbol{\phi}}=\left(\boldsymbol{U}'\boldsymbol{U}\right)^{-1}\boldsymbol{U}'\boldsymbol{u}, \quad \boldsymbol{\phi}=E\left[\left(\boldsymbol{U}'\boldsymbol{U}\right)^{-1}\boldsymbol{U}'\boldsymbol{u}\right], \tag{A.8}$$

and

$$\widehat{\sigma}^2=\left(\widehat{\boldsymbol{u}}-\widehat{\boldsymbol{U}}\widehat{\boldsymbol{\phi}}\right)'\left(\widehat{\boldsymbol{u}}-\widehat{\boldsymbol{U}}\widehat{\boldsymbol{\phi}}\right), \quad \widetilde{\sigma}^2=\left(\boldsymbol{u}-\boldsymbol{U}\widetilde{\boldsymbol{\phi}}\right)'\left(\boldsymbol{u}-\boldsymbol{U}\widetilde{\boldsymbol{\phi}}\right),$$
$$\sigma^2=E\left[\left(\boldsymbol{u}-\boldsymbol{U}\boldsymbol{\phi}\right)'\left(\boldsymbol{u}-\boldsymbol{U}\boldsymbol{\phi}\right)\right].$$

Notice that $\left\|\widehat{\boldsymbol{\phi}}-\boldsymbol{\phi}\right\|_1=\left\|\widehat{\boldsymbol{\phi}}-\widetilde{\boldsymbol{\phi}}\right\|_1+\left\|\widetilde{\boldsymbol{\phi}}-\boldsymbol{\phi}\right\|_1$. We are interested to show that the following holds

$$\left\|\widehat{\boldsymbol{\phi}}-\boldsymbol{\phi}\right\|_1=O_P\left(s_0^2\lambda^2\vee s_0\lambda^2\right). \tag{A.9}$$

To show (A.9), it is sufficient to show that

$$\left\|\widehat{\boldsymbol{\phi}}-\widetilde{\boldsymbol{\phi}}\right\|_1+\left\|\widetilde{\boldsymbol{\phi}}-\boldsymbol{\phi}\right\|_1=O_P\left(s_0^2\lambda^2\right)+O_P\left(s_0\lambda^2\right), \tag{A.10}$$
$$T^{-1}\widehat{\sigma}^2\to\sigma^2, \quad T^{-1}\widetilde{\sigma}^2\to\sigma^2. \tag{A.11}$$

Proof of (A.10). To prove the first two statements in (A.10) it is sufficient to show that

$$T^{-1}\left(\sum_{t=j+1}^{T}\left(\widehat{u}_{t-j}\widehat{u}_t-u_{t-j}u_t\right)\right)=O_P\left(s_0^2\lambda^2\vee s_0\lambda^2\right), \tag{A.12}$$
$$T^{-1}\left(\sum_{t=j+1}^{T}\left(\widehat{u}_{t-j}^2-u_{t-j}^2\right)\right)=O_P\left(s_0^2\lambda^2\vee s_0\lambda^2\right). \tag{A.13}$$

Proof of (A.12). (A.12) is bounded by $s_1+s_2+s_3$, where

$$s_1=T^{-1}\sum_{t=j+1}^{T}\left(\widehat{u}_t-u_t\right)\left(\widehat{u}_{t-j}-u_{t-j}\right), \tag{A.14}$$

$$s_2 = T^{-1} \sum_{t=j+1}^{T} (\widehat{u}_t - u_t)\, u_{t-j}, \quad s_3 = T^{-1} \sum_{t=j+1}^{T} u_t\, (\widehat{u}_{t-j} - u_{t-j})\,. \tag{A.15}$$

Notice that $\widehat{u}_t = u_t + \boldsymbol{x}_t' \left(\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta}\right)$, and by Assumptions 1–3, $\{\boldsymbol{x}_t\}$, $\{u_t\}$ are stationary, ergodic and mutually uncorrelated, α-mixing series. By Assumptions 1–4 and by using the exponential inequalities in Theorem 1 of Dendramis et al. (2021), the following hold

$$\left\| \frac{1}{T} \sum_{t=j+1}^{T} \boldsymbol{x}_t' \boldsymbol{x}_{t-j} \right\|_\infty = O_P(1), \quad \left\| \frac{1}{T} \sum_{t=j+1}^{T} u_t \boldsymbol{x}_{t-j} \right\|_\infty = O_P(\lambda), \quad \left\| \frac{1}{T} \sum_{t=j+1}^{T} \boldsymbol{x}_t u_{t-j} \right\|_\infty = O_P(\lambda), \tag{A.16}$$

where $j = 1, \ldots, q$. Further, by Lemma 1, we have that

$$\left\| \widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta} \right\|_1 = O_P\left(s_0 \lambda\right). \tag{A.17}$$

Then by the Cauchy-Schwartz inequality and substituting (A.16) and (A.17) in (A.12), we obtain

$$\begin{aligned} s_1 &\le \left\| T^{-1} \sum_{t=j+1}^{T} \boldsymbol{x}_t' \boldsymbol{x}_{t-j} \right\|_\infty \left\| \widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta} \right\|_1^2 = O_P\left(s_0^2 \lambda^2\right), \\ s_2 &\le \left\| T^{-1} \sum_{t=j+1}^{T} \boldsymbol{x}_t u_{t-j} \right\|_\infty \left\| \widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta} \right\|_1 = O_P\left(s_0 \lambda^2\right). \end{aligned}$$

Similarly, s_3 is bounded using the same arguments with s_2, which completes the proof of the first part of (A.10). In view of the results in (A.12) – (A.13) and the analysis in [8.3.17]– [8.3.19] of Hamilton (1994), we can conclude that the second part of (A.10) holds.

Proof of (A.13). Using (A.16)–(A.17), obtain

$$\begin{aligned} T^{-1} \sum_{t=j+1}^{T} \left(\widehat{u}_{t-j}^2 - u_{t-j}^2\right) &\le 2 \left\| T^{-1} \sum_{t=j+1}^{T} u_{t-j} \boldsymbol{x}_{t-j} \right\|_\infty \left\| \widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta} \right\|_1 + \left\| T^{-1} \sum_{t=j+1}^{T} \boldsymbol{x}_{t-j}' \boldsymbol{x}_{t-j} \right\|_\infty \left\| \widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta} \right\|_2^2 \\ &= O_P\left(s_0 \lambda^2 \vee s_0^2 \lambda^2\right), \end{aligned} \tag{A.18}$$

where λ satisfies Assumption 6. Together with (A.12)–(A.13), we obtain

$$\left\| \widehat{\boldsymbol{\phi}} - \widetilde{\boldsymbol{\phi}} \right\|_1 = O_P\left(s_0 \lambda^2 \vee s_0^2 \lambda^2\right). \tag{A.19}$$

Proof of (A.11). The convergence of both statements in (A.11) follows directly from the definitions of $\widehat{\sigma}^2$ and $\widetilde{\sigma}^2$ through statements (A.12), (A.13) and the definitions of $\widehat{\boldsymbol{\phi}}$, $\widetilde{\boldsymbol{\phi}}$, and $\boldsymbol{\phi}$.

Following a similar analysis to show the first statement of (A.10), one can show that $\|\widetilde{\boldsymbol{\phi}} - \boldsymbol{\phi}\|_1 = O_P\left(s_0^2 \lambda^2\right) + O_P\left(s_0 \lambda^2\right)$. Together with (A.10) and (A.11), we obtain

$$\left\| \widehat{\boldsymbol{\phi}} - \boldsymbol{\phi} \right\|_1 = O_P\left(s_0^2 \lambda^2 \vee s_0 \lambda^2\right).$$

Further, in **Case 1**: $(u_{t-j}\boldsymbol{x}_{t-j}) \in \mathcal{T}(r), r > 0$, $\lambda = O(\sqrt{\log p/T})$, for $s_0 = o(\sqrt{T}/\log p)$,

$$\|\widetilde{\boldsymbol{\phi}} - \boldsymbol{\phi}\|_1 = O_P\left(s_0^2 T^{-1}\log p \vee s_0 T^{-1}\log p\right) \asymp O_P(T^{-1/2}).$$

In **Case 2**: $(u_{t-j}\boldsymbol{x}_{t-j}) \in \mathcal{H}(\theta), \theta > 4$, $\lambda = O(p^{\frac{1}{\theta}}T^{-(\frac{\theta}{2}+1)})$, for $s_0 = o(T^{(\theta-1)/2}p^{-\frac{1}{\theta}})$

$$\|\widetilde{\boldsymbol{\phi}} - \boldsymbol{\phi}\|_1 = O_P\left(s_0^2 p^{\frac{1}{\theta^2}}T^{-(\frac{\theta}{2}+1)^2} \vee s_0 p^{\frac{1}{\theta^2}}T^{-(\frac{\theta}{2}+1)^2}\right) \asymp O_P(T^{-1/2}). \quad \square$$

Proof of Theorem 1

In this Theorem, we illustrate that the feasible GLS LASSO estimator, attains similar non-asymptotic bounds to the LASSO, both in terms of prediction and estimation errors. The proof follows closely the steps in Chapter 6 of Bühlmann and Van De Geer (2011). We consider the feasible GLS corrected model:

$$\widetilde{y}_t = \widetilde{\boldsymbol{x}}_t'\boldsymbol{\beta} + \widehat{\varepsilon}_t, \quad \text{where}$$

$$\widetilde{y}_t = y_t - \sum_{j=1}^{q}\widehat{\phi}_j y_{t-j}, \quad \widetilde{\boldsymbol{x}}_t = \boldsymbol{x}_t - \sum_{j=1}^{q}\widehat{\phi}_j \boldsymbol{x}_{t-j}, \quad \widehat{\varepsilon}_t = u_t - \sum_{j=1}^{q}\widehat{\phi}_j u_{t-j}, \quad t = q+1,\dots,T,$$

where $\widehat{\phi}_j$ is the OLS estimate obtained from an $AR(q)$ regression on the residuals $\widehat{u}_t$, where $\widehat{u}_t = y_t - \boldsymbol{x}_t'\widetilde{\boldsymbol{\beta}}$, $\widetilde{\boldsymbol{\beta}}$ the solution to the original LASSO problem, using $(\boldsymbol{y}, \boldsymbol{X})$. Re-arranging the basic inequality, as in Lemma 6.1 of Bühlmann and Van De Geer (2011), we obtain

$$\frac{1}{T}\left\|\widetilde{\boldsymbol{X}}\left(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right)\right\|_2^2 \leq \frac{2}{T}\widehat{\boldsymbol{\varepsilon}}'\widetilde{\boldsymbol{X}}\left(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right) + \lambda\left(\|\boldsymbol{\beta}\|_1 - \|\widehat{\boldsymbol{\beta}}\|_1\right). \tag{A.20}$$

Define $2\widehat{\boldsymbol{\varepsilon}}'\widetilde{\boldsymbol{X}}\left(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right)/T$ as the "empirical process". Notice that the latter can be bounded further in terms of the ℓ_1-norm, such that,

$$\frac{1}{T}\left|\widehat{\boldsymbol{\varepsilon}}'\widetilde{\boldsymbol{X}}\left(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right)\right| \leq \frac{2}{T}\left\|\widehat{\boldsymbol{\varepsilon}}'\widetilde{\boldsymbol{X}}\right\|_\infty \left\|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right\|_1. \tag{A.21}$$

The regularisation parameter, λ is chosen such that $T^{-1}\left\|\widehat{\boldsymbol{\varepsilon}}'\widetilde{\boldsymbol{X}}\right\|_\infty \leq \lambda$. Hence, we introduce the following event

$$\widetilde{\mathcal{E}} := \left\{T^{-1}\left\|\widehat{\boldsymbol{\varepsilon}}'\widetilde{\boldsymbol{X}}\right\|_\infty \leq \frac{\lambda_0}{2}\right\}, \tag{A.22}$$

which needs to hold with high probability. Further, recall under Assumption 5, $\lambda_0 = \sqrt{\log p/T}$ or $\lambda_0 = p^{\frac{1}{\theta}}T^{-(\frac{\theta}{2}+1)}$, depending on the tail behaviour of $\{\boldsymbol{x}_t\}$ and $\{\varepsilon_t\}$, and $\lambda_0 < \lambda/2$. We use the vector representation of the process $\widehat{\boldsymbol{\varepsilon}}'\widetilde{\boldsymbol{X}}$, and proceed with the following steps:

$$\begin{aligned}
\frac{2}{T}\sum_{t=1}^{T}\widehat{\varepsilon}_t\widetilde{\boldsymbol{x}}_t'\left(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right) &= \frac{2}{T}\sum_{t=1}^{T}\left(u_t - \sum_{j=1}^{q}\widehat{\phi}_j u_{t-j}\right)\widetilde{\boldsymbol{x}}_t'\left(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right) \\
&= \frac{2}{T}\sum_{t=1}^{T}\left[u_t\widetilde{\boldsymbol{x}}_t'\left(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right) + \widetilde{\boldsymbol{x}}_t'\left(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right)\sum_{j=1}^{q}\widehat{\phi}_j u_{t-j}\right] \\
&= \frac{2}{T}\sum_{t=1}^{T}\left[(\varepsilon_t + \sum_{j=1}^{q}\phi_j u_{t-j})\widetilde{\boldsymbol{x}}_t'\left(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right) + \widetilde{\boldsymbol{x}}_t'\left(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right)\sum_{j=1}^{q}\widehat{\phi}_j u_{t-j}\right]
\end{aligned}$$

$$= \frac{2}{T}\sum_{t=1}^{T}\left[\left(\widetilde{\boldsymbol{x}}_t\sum_{j=1}^{q}\phi_j u_{t-j} + \varepsilon_t\widetilde{\boldsymbol{x}}_t + \widetilde{\boldsymbol{x}}_t\sum_{j=1}^{q}\widehat{\phi}_j u_{t-j}\right)\left(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right)\right]$$

$$\leq \max_{i=1,\dots,p}\left|\frac{2}{T}\left[\sum_{t=1}^{T}\left(\widetilde{x}_{t,i}\sum_{j=1}^{q}\phi_j u_{t-j} + \varepsilon_t\widetilde{x}_{t,i} + \widetilde{x}_{t,i}\sum_{j=1}^{q}\widehat{\phi}_j u_{t-j}\right)\right]\right|\left\|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right\|_1. \tag{A.23}$$

Then, we create the set $\widetilde{\mathcal{E}}$, under which (A.21) holds.

$$\widetilde{\mathcal{E}} = \left\{\max_{i=1,\dots,p}\left|\frac{1}{T}\sum_{t=1}^{T}\left[\widetilde{x}_{t,i}\sum_{j=1}^{q}\phi_j u_{t-j} + \varepsilon_t\widetilde{x}_{t,i} + \widetilde{x}_{t,i}\sum_{j=1}^{q}\widehat{\phi}_j u_{t-j}\right]\right| \leq \frac{\lambda_0}{2}\right\} \tag{A.24}$$

$$= \left\{\bigcup_{i=1}^{p}\left|\frac{1}{T}\left[\sum_{t=1}^{T}\widetilde{x}_{t,i}\sum_{j=1}^{q}\phi_j u_{t-j} + \sum_{t=1}^{T}\varepsilon_t\widetilde{x}_{t,i} + \sum_{t=1}^{T}\widetilde{x}_{t,i}\sum_{j=1}^{q}\widehat{\phi}_j u_{t-j}\right]\right| \leq \frac{\lambda_0}{2}\right\}. \tag{A.25}$$

Using the identity $P(\widetilde{\mathcal{E}}^c) = 1 - P(\widetilde{\mathcal{E}})$ and the union bound, we get

$$P(\widetilde{\mathcal{E}}^c) = 1 - P\left(\bigcup_{i=1}^{p}\left|\frac{1}{T}\left[\sum_{t=1}^{T}\widetilde{x}_{t,i}\sum_{j=1}^{q}\phi_j u_{t-j} + \sum_{t=1}^{T}\varepsilon_t\widetilde{x}_{t,i} + \widetilde{x}_{t,i}\sum_{j=1}^{q}\widehat{\phi}_j u_{t-j}\right]\right| \leq \frac{\lambda_0}{2}\right)$$

$$\leq 1 - \sum_i P\left(\left|\frac{1}{T}\left[\sum_{t=1}^{T}\widetilde{x}_{t,i}\sum_{j=1}^{q}\phi_j u_{t-j} + \sum_{t=1}^{T}\varepsilon_t\widetilde{x}_{t,i} + \sum_{t=1}^{T}\widetilde{x}_{t,i}\sum_{j=1}^{q}\widehat{\phi}_j u_{t-j}\right]\right| \leq \frac{\lambda_0}{2}\right)$$

$$\leq \sum_i\left\{P\left(\left|\frac{1}{T}\sum_{t=1}^{T}x_{t,i}\sum_{j=1}^{q}\left(\widehat{\phi}_j - \phi_j\right)u_{t-j}\right| > \frac{\lambda_0}{6}\right) + P\left(\frac{1}{T}\left|\sum_{t=1}^{T}\varepsilon_t x_{t,i}\right| > \frac{\lambda_0}{6}\right)\right.$$

$$\left. + P\left(\left|\frac{1}{T}\sum_{t=1}^{T}\sum_{j=1}^{q}\widehat{\phi}_j x_{t-j,i}\left(\widehat{\phi}_j - \phi_j\right)u_{t-j}\right| > \frac{\lambda_0}{6}\right)\right\} = (I) + (II) + (III). \tag{A.26}$$

We proceed to analyse each term. Define $\{\boldsymbol{z}_t\} = \{\boldsymbol{x}_t\sum_{j=1}^{q}u_{t-j}\}$ which is a p-dimensional, zero-mean, stationary and ergodic α-mixing series, as a product of two α-mixing series by Assumptions 1–3 and using Theorem 14.1 of Davidson (1994), with properties similar to the ones in Assumption 3, for (I) and $i = 1,\dots,p$, we have:

$$\sum_i P\left(\left|\frac{1}{T}\sum_{t=1}^{T}x_{t,i}\sum_{j=1}^{q}\left(\widehat{\phi}_j - \phi_j\right)u_{t-j}\right| > \frac{\lambda_0}{6}\right)$$

$$\leq \sum_i P\left(\left|\frac{1}{\sqrt{T}}\sum_{t=1}^{T}z_{t,i} - E(z_{t,i})\right| > \frac{\lambda_0\sqrt{T}}{6C}\right) + P\left(\left|\sum_{j=1}^{q}\left(\widehat{\phi}_j - \phi_j\right)\right| > C\right)$$

$$\leq \sum_i P\left(\left|\frac{1}{\sqrt{T}}\sum_{t=1}^{T}z_{t,i} - E(z_{t,i})\right| > \frac{\sqrt{\log p}}{6C}\right) + o(1).$$

Case 1: The process $\zeta_{t,i} \in \mathcal{T}(r),\ r > 0$, then

$$\sum_i P\left(\left|\frac{1}{\sqrt{T}}\sum_{t=1}^{T}z_{t,i} - E(z_{t,i})\right| > \frac{\sqrt{\log p}}{6C}\right) \leq pc\left\{\exp\left(-c_2\frac{\log p}{36C^2}\right) + \exp\left(-c_3\frac{\sqrt{\log p}}{6C\log^2 T}\right)^{\zeta}\right\}$$

$$= I_1 + I_2, \tag{A.27}$$

where by Lemma 1 of Dendramis et al. (2021) to obtain (A.27). It is then sufficient to bound I_1:

$$I_1 = pc \exp\left(-c_2 \frac{\log p}{36C^2}\right) = cp^{1-\left(\frac{c_2}{36C^2}\right)}, \tag{A.28}$$

for some large enough constant, $c_2 > 0$, and positive constants, c, C. For $i = 1, \ldots, p$, we have,

$$\sum_i P\left(\left|\frac{1}{\sqrt{T}}\sum_{t=1}^{T} z_{t,i} - E(z_{t,i})\right| > \frac{\sqrt{\log p}}{6C}\right) \leq cp^{1-\left(\frac{c_2}{36C^2}\right)} = o\left(p^{1-\left(\frac{c_2}{36C^2}\right)}\right).$$

Case 2: The process $\zeta_{t,i} \in \mathcal{H}(\theta),\ \theta > 4$, then

$$\sum_i P\left(\left|\frac{1}{\sqrt{T}}\sum_{t=1}^{T} z_{t,i} - E(z_{t,i})\right| > \frac{\lambda_0\sqrt{T}}{6C}\right) \leq cp\left(\frac{\lambda_0\sqrt{T}}{6C}\right)^{-\theta} T^{-\frac{\theta}{2}-1} \to 0 \tag{A.29}$$

We proceed to analyse (II). Define $\{\boldsymbol{v}_t\} = \{\boldsymbol{x}_t\varepsilon_t\}$, then by Assumption 5, Theorem 14.1 of Davidson (1994) and by using the results of Lemma 1 of Dendramis et al. (2021), the probability inequalities for $\{\boldsymbol{v}_t\}$ follow similarly to (A.27)–(A.29), for both cases of $v_{t,i} \in \mathcal{T}(r),\ r > 0$ or $v_{t,i} \in \mathcal{H}(\theta),\ \theta > 4$.

To analyse (III), recall that $\{\sum_{j=1}^{q} \phi_j \boldsymbol{x}_{t-j} u_{t-j}\}$ is α-mixing as a product of two α-mixing series, by Theorem 14.1 of Davidson (1994), then define $\{\boldsymbol{\zeta}_t\} = \{\sum_{j=1}^{q} \boldsymbol{x}_{t-j} u_{t-j}\}$, $\{\boldsymbol{\eta}_t\} = \{\sum_{j=1}^{q} \phi_j \boldsymbol{x}_{t-j} u_{t-j}\}$. Then:

$$\begin{aligned}
&\sum_i P\left(\left|\frac{1}{T}\sum_{t=1}^{T}\sum_{j=1}^{q} \widehat{\phi}_j x_{t-j,i}\left(\widehat{\phi}_j - \phi_j\right) u_{t-j}\right| > \frac{\lambda_0}{6}\right) \\
&= \sum_i P\left(\left|\frac{1}{T}\sum_{t=1}^{T}\sum_{j=1}^{q}\left[\left(\widehat{\phi}_j - \phi_j\right) x_{t-j,i}\left(\widehat{\phi}_j - \phi_j\right) u_{t-j} + \phi_j x_{t-j,i}\left(\widehat{\phi}_j - \phi_j\right) u_{t-j}\right]\right| > \frac{\lambda_0}{6}\right) \\
&= \sum_i P\left(\left|\frac{1}{T}\sum_{t=1}^{T}\sum_{j=1}^{q}\left[\left(\widehat{\phi}_j - \phi_j\right) x_{t-j,i}\left(\widehat{\phi}_j - \phi_j\right) u_{t-j}\right]\right| > \frac{\lambda_0}{12}\right) \\
&\quad + \sum_i P\left(\left|\frac{1}{T}\sum_{t=1}^{T}\sum_{j=1}^{q}\left[\phi_j x_{t-j,i}\left(\widehat{\phi}_j - \phi_j\right) u_{t-j}\right]\right| > \frac{\lambda_0}{12}\right) \\
&\underset{(1)}{\leq} \sum_i \left[P\left(\left|\frac{1}{T}\sum_{t=1}^{T} \zeta_{t,i} - E(\zeta_{t,i})\right| > \frac{\lambda_0}{12C_0}\right) P\left(\left|\frac{1}{T}\sum_{t=1}^{T} \eta_{t,i} - E(\eta_{t,i})\right| > \frac{\lambda_0}{12C_1}\right)\right] \\
&\quad + P\left(\left|\sum_{j=1}^{q}\left(\widehat{\phi}_j - \phi_j\right)^2\right| > C_0\right) + P\left(\left|\sum_{j=1}^{q}\left(\widehat{\phi}_j - \phi_j\right)\right| > C_1\right) = B_1 \times B_2 + B_3 + B_4.
\end{aligned} \tag{A.30}$$

Note that (1) results by direct application of Lemma A.11, equation (B.60) of Chudik et al. (2018). Notice that by implication of Corollary 1 $B_3 = o(1)$, $B_4 = o(1)$, B_1 and B_2 follow

similar line of arguments with term (I) of (A.26), specifically:
Case 1: The process $\zeta_{t,i} \in \mathcal{T}(r),\ r>0$, then

$$
\begin{aligned}
B_1 \leq \sum_i P\left(\left|T^{-1/2}\sum_{t=1}^{T}[\zeta_{t,i}-E(\zeta_{t,i})]\right| > T^{1/2}\lambda_0/12C_0\right) &\leq pc\left(\exp(-c_1(T^{1/2}\lambda_0/12C_0)^{\gamma_1})\right.\\
&\quad \left.+\exp(-c_2((T^{1/2}\lambda_0/12C_0)T^{1/2}\log^{-2}T)^{\gamma_2})\right)\\
&\underset{\gamma_1=2}{\leq} cp^{1-c_1/(144C_0^2)} \to 0, \qquad (A.31)
\end{aligned}
$$

where $c, c_1, C_0 > 0$, for a large enough p.
Case 2: The process $\zeta_{t,i} \in \mathcal{H}(\theta),\ \theta > 4$, then

$$
B_1 \leq \sum_i P\left(\left|T^{-1/2}\sum_{t=1}^{T}[\zeta_{t,i}-E(\zeta_{t,i})]\right| > T^{1/2}\lambda_0/12C_0\right) \leq pc(T^{1/2}\lambda_0/12C_0)^{-\theta}T^{-\frac{\theta}{2}-1} \to 0, \qquad (A.32)
$$

at a slower rate than in Case 1. For B_2, the analysis is similar to B_1, then for Case 1 where $\eta_{t,i} \in \mathcal{T}(r),\ r>0,\ B_2 \leq c_* p^{1-c_2/144C_0^2} \to 0$, for some positive constants, c_*, c_2, C_0, and for Case 2: $\eta_{t,i} \in \mathcal{H}(\theta),\ \theta > 4,\ \lambda_0 = \frac{p^{\frac{1}{\theta}}}{T^{(\frac{\theta}{2}+1)}}$, we have that $B_2 \leq pc_*(T^{1/2}\lambda_0/12C_0)^{-\theta}T^{-\frac{\theta}{2}-1} \to 0$, for some positive constants, c_*, c_2, C_0 for $\lambda_0 = p^{-\frac{1}{\theta}}T^{-(\frac{\theta}{2})+1}$.

Up to this point in the analysis, we bound the RHS of (A.20) in probability, so that the LHS is bounded by a dual bound. We continue the analysis, towards showing the latter. Note that the "prediction error", $\left\|\widetilde{\boldsymbol{X}}\left(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}\right)\right\|_2^2$ is ℓ_2-bounded and $\left\|\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}\right\|_1$ is ℓ_1-bounded, then by Corollary 6.2 of Bühlmann and Van De Geer (2011)

$$
\begin{aligned}
\frac{1}{T}\left\|\widetilde{\boldsymbol{X}}(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta})\right\|_2^2 &\leq \frac{2}{T}\widehat{\boldsymbol{\varepsilon}}'\widetilde{\boldsymbol{X}}\left(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}\right)+\lambda\left(\|\boldsymbol{\beta}\|_1-\|\widehat{\boldsymbol{\beta}}\|_1\right), \qquad (A.33)\\
&\leq \left\|\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}\right\|_1+\lambda\left(\|\boldsymbol{\beta}\|_1-\|\widehat{\boldsymbol{\beta}}\|_1\right). \qquad (A.34)
\end{aligned}
$$

Since $\lambda \geq 2\lambda_0$ under $\widetilde{\mathcal{E}}$, and by Assumption 7,

$$
\begin{aligned}
&\left\|\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}\right\|_1 = \left\|\widehat{\boldsymbol{\beta}}_{S_0}-\boldsymbol{\beta}_{S_0}\right\|_1+\left\|\widehat{\boldsymbol{\beta}}_{S_0^c}\right\|_1, \quad \left\|\widehat{\boldsymbol{\beta}}_{S_0^c}\right\|_1 \leq 3\left\|\widehat{\boldsymbol{\beta}}_{S_0}-\boldsymbol{\beta}_{S_0}\right\|_1 \qquad (A.35)\\
&\left\|\widehat{\boldsymbol{\beta}}_{S_0}-\boldsymbol{\beta}_{S_0}\right\|_1 \leq \sqrt{s_0}\left\|\widehat{\boldsymbol{\beta}}_{S_0}-\boldsymbol{\beta}_{S_0}\right\|_2, \quad \left\|\widehat{\boldsymbol{\beta}}\right\|_1 = \left\|\widehat{\boldsymbol{\beta}}_{S_0}\right\|_1+\left\|\widehat{\boldsymbol{\beta}}_{S_0^c}\right\|_1 \qquad (A.36)\\
&\left\|\boldsymbol{\beta}_{S_0}\right\|_1^2 \leq \left(\boldsymbol{\beta}'\boldsymbol{\Sigma}\boldsymbol{\beta}\right)s_0\zeta_*^{-2}\left(s_0,\boldsymbol{\phi}\right). \qquad (A.37)
\end{aligned}
$$

Then, substituting (A.35)–(A.37) in (A.34), we obtain the following dual bound

$$
\begin{aligned}
\left\|\widetilde{\boldsymbol{X}}\left(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}\right)\right\|_2^2+\left\|\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}\right\|_1 &\leq \left\|\widetilde{\boldsymbol{X}}\left(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}\right)\right\|_2^2+\left\|\widehat{\boldsymbol{\beta}}_{S_0}-\boldsymbol{\beta}_{S_0}\right\|_1+3\left\|\widehat{\boldsymbol{\beta}}_{S_0}-\boldsymbol{\beta}_{S_0}\right\|_1\\
&\leq \left\|(\widetilde{\boldsymbol{X}}-\boldsymbol{X}^*)\left(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}\right)\right\|_2^2+\left\|\boldsymbol{X}^*\left(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}\right)\right\|_2^2\\
&\quad +\left\|\widehat{\boldsymbol{\beta}}_{S_0}-\boldsymbol{\beta}_{S_0}\right\|_1+3\left\|\widehat{\boldsymbol{\beta}}_{S_0}-\boldsymbol{\beta}_{S_0}\right\|_1\\
&\leq \left\|(\widehat{\boldsymbol{L}}-\boldsymbol{L})\boldsymbol{X}\left(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}\right)\right\|_2^2+\left\|\boldsymbol{X}^*\left(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}\right)\right\|_2^2+4s_0\zeta_*^{-2}(s_0,\boldsymbol{\phi})\lambda,
\end{aligned}
$$

$$\leq \left\|(\widehat{\boldsymbol{L}}-\boldsymbol{L})\right\|_2^2 \left\|\boldsymbol{X}\left(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}\right)\right\|_2^2 + \left\|\boldsymbol{X}^*\left(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}\right)\right\|_2^2 + 4s_0\zeta_*^{-2}(s_0,\boldsymbol{\phi})\lambda, \tag{A.38}$$

leading to

$$\left\|\widetilde{\boldsymbol{X}}\left(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}\right)\right\|_2^2 - \left\|\boldsymbol{X}^*\left(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}\right)\right\|_2^2 + \left\|\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}\right\|_1 - \left\|(\widehat{\boldsymbol{L}}-\boldsymbol{L})\boldsymbol{X}\left(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}\right)\right\|_2^2 \leq 4s_0\zeta_*^{-2}(s_0,\boldsymbol{\phi})\lambda$$
$$A+B+C \leq 4s_0\zeta_*^{-2}(s_0,\boldsymbol{\phi})\lambda,$$

Consider $A = \left\|\widetilde{\boldsymbol{X}}\left(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}\right)\right\|_2^2 - \left\|\boldsymbol{X}^*\left(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}\right)\right\|_2^2$, then by the squared reverse triangle inequality

$$\begin{aligned} A &\geq -\left\|(\widehat{\boldsymbol{L}}-\boldsymbol{L})\right\|_2^2 \left\|\boldsymbol{X}\left(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}\right)\right\|_2^2 - 2\left\|(\widehat{\boldsymbol{L}}-\boldsymbol{L})\right\|_2 \|\boldsymbol{X}\|_2 \left\|\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}\right\|_2 \\ -A &\leq \left\|(\widehat{\boldsymbol{L}}-\boldsymbol{L})\right\|_2^2 \left\|\boldsymbol{X}\left(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}\right)\right\|_2^2 + 2\left\|(\widehat{\boldsymbol{L}}-\boldsymbol{L})\right\|_1 \|\boldsymbol{X}\|_2 \left\|\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}\right\|_1 \\ &= O_P\left(\frac{1}{T}\right)\left\|\boldsymbol{X}\left(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}\right)\right\|_2^2 + O_P\left(\frac{1}{\sqrt{T}}\right)\Lambda_{\max}^{1/2}\left\|\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}\right\|_1, \end{aligned} \tag{A.39}$$

where $\Lambda_{\max}^{1/2}$ the maximum eigenvalue of $(T^{-1}\boldsymbol{X'X})^{1/2}$. Then $B = \left\|\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}\right\|_1$, which will be used to obtain a dual bound below, and by implication of Corollary 1

$$C = \left\|(\widehat{\boldsymbol{L}}-\boldsymbol{L})\boldsymbol{X}\left(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}\right)\right\|_2^2 = O_P(T^{-1})\left\|\boldsymbol{X}\left(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}\right)\right\|_2^2.$$

Combining terms A, B, C and for some $\Lambda_{\max}^{1/2} > 0$, we have the following dual bound for the feasible GLS LASSO estimator

$$T^{-1}\left\|\widetilde{\boldsymbol{X}}\left(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}\right)\right\|_2^2 \leq 4\zeta_*^{-2}(s_0,\boldsymbol{\phi})s_0\lambda^2 + \frac{\lambda^2}{T}, \tag{A.40}$$

$$\left\|\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}\right\|_1 \leq 4\zeta_*^{-2}(s_0,\boldsymbol{\phi})s_0\lambda + \frac{\lambda}{\sqrt{T}}, \tag{A.41}$$

with probability approach 1 for $\{\lambda, s_0\}$ defined in (18) of the main paper and λ satisfies Assumption 6. □

Proof of Corollary 2

Following the analysis in (A.38)–(A.41) the results follows. □

B Proofs of Theorems 2, and 3

This Section provides proofs of Theorems 2, and 3 of the main paper.

Proof of Theorem 2

Consider Assumptions 1 – 7 to hold. Then, with a selection of λ_0, the weak irrepresentable condition proposed by Zhao and Yu (2006), holds, hence $P\left(\left|\widehat{S}_0 / \ S_0\right|\right) \to 1$, where $S_0 = \{i : \beta_i \neq 0\}$, $\widehat{S}_0 = \{i : \ \widehat{\beta}_i \neq 0\}$. This can be easily confirmed by directly applying Theorem 4 of Zhao and Yu (2006). For the sake of clarity we provide the proof below:

Let $\boldsymbol{\beta}_{s_0} := \beta_i \, \mathbf{1}\{i \in S_0;\ i = 1, \ldots, p\}$, and $\boldsymbol{\beta}_{s_0^c} := \beta_i \, \mathbf{1}\{i \notin S_0, i = 1, \ldots, p\}$. Denote $\widetilde{\boldsymbol{X}}(S_0)$, $\widetilde{\boldsymbol{X}}(S_0^c)$ as the first s_0 and last $p - s_0$ columns of $\widetilde{\boldsymbol{X}}$ respectively and let $\widehat{\boldsymbol{\Sigma}} = \widetilde{\boldsymbol{X}}'\widetilde{\boldsymbol{X}}/T$. By setting $\widehat{\boldsymbol{\Sigma}}_{11} = \widetilde{\boldsymbol{X}}(S_0)'\widetilde{\boldsymbol{X}}(S_0)/T, \widehat{\boldsymbol{\Sigma}}_{22} = \widetilde{\boldsymbol{X}}(S_0^c)'\widetilde{\boldsymbol{X}}(S_0^c)/T, \widehat{\boldsymbol{\Sigma}}_{12} = \widetilde{\boldsymbol{X}}(S_0)'\widetilde{\boldsymbol{X}}(S_0^c)/T$, and $\widehat{\boldsymbol{\Sigma}}_{21} = \widetilde{\boldsymbol{X}}(S_0^c)'\widetilde{\boldsymbol{X}}(S_0)/T$. $\widehat{\boldsymbol{\Sigma}}$ can then be expressed in a block-wise form as follows:

$$\widehat{\boldsymbol{\Sigma}} = \begin{bmatrix} \widehat{\boldsymbol{\Sigma}}_{11} & \widehat{\boldsymbol{\Sigma}}_{12} \\ \widehat{\boldsymbol{\Sigma}}_{21} & \widehat{\boldsymbol{\Sigma}}_{22} \end{bmatrix}.$$

We introduce the following statements

$$\boldsymbol{\omega} = \left(\widehat{\boldsymbol{\Sigma}}_{21}\left(\widehat{\boldsymbol{\Sigma}}_{11}\right)^{-1}\widetilde{\boldsymbol{X}}(S_0)'\widehat{\boldsymbol{\varepsilon}}\right), \quad W(S_0) = \frac{1}{\sqrt{T}}\widetilde{\boldsymbol{X}}(S_0)'\widehat{\boldsymbol{\varepsilon}}, \quad W(S_0^c) = \frac{1}{\sqrt{T}}\widetilde{\boldsymbol{X}}(S_0^c)'\widehat{\boldsymbol{\varepsilon}},$$

to define two distinct events,

$$A_T = \left\{\left|\left(\widehat{\Sigma}_{11,i}\right)^{-1} W_i(S_0)\right| < \sqrt{T}\left(|\beta_i| - \frac{\lambda_0}{2T}\left|\left(\widehat{\Sigma}_{11,i}\right)^{-1}\operatorname{sign}(\beta_i)\right|\right)\right\}, \quad i = 1, \ldots, s_0,$$
$$B_T = \left\{|\omega_i - W_i(S_0^c)| \le \frac{\lambda_0}{2\sqrt{T}}\right\}, \quad i = s_0 + 1, \ldots, p,$$

where $W_i(S_0)$, $W_i(S_0^c)$ is an element of $W(S_0)$, and $W(S_0^c)$ respectively, similarly, ω_i is an element of $\boldsymbol{\omega}$. Event A_T implies that the signs of the active set, S_0, are correctly estimated, while A_T, B_T together imply that the signs of the non-active set, S_0^c, are estimated consistently.

To show $P\left(\left|\widehat{S}_0 /\ S_0\right|\right) \to 1$, it is sufficient to show that

$$P\left(\left|\widehat{S}_0 /\ S_0\right|\right) \ge P\left(A_T \cap B_T\right). \tag{B.1}$$

Using the identity of $1 - P\left(A_T \cap B_T\right) \le P\left(A_T\right)^c + P\left(B_T\right)^c$ we have that

$$P\left(A_T\right)^c + P\left(B_T\right)^c \le \sum_{i \in S_0} P\left(\frac{1}{\sqrt{T}}\left|\left(\widehat{\Sigma}_{11,i}\right)^{-1}\widetilde{\boldsymbol{x}}_i'\widehat{\boldsymbol{\varepsilon}}\right| \ge \sqrt{T}\,|\beta_i| - \frac{\lambda_0}{2T}\left|\left(\widehat{\Sigma}_{11,i}\right)^{-1}\operatorname{sign}(\beta_{\mathrm{i}})\right|\right)$$
$$+ \sum_{i \in S_0^c} P\left(\frac{1}{\sqrt{T}}\left|\omega_i - \widetilde{\boldsymbol{x}}_i'\widehat{\boldsymbol{\varepsilon}}\right| \le \frac{\lambda_0}{2\sqrt{T}}\right) = A + B, \tag{B.2}$$

where $\boldsymbol{\omega}$ is a $(p - s_0) \times 1$ vector. Notice that for $i, j = 1, 2$ and by implication of Assumption 7, $0 < \lambda_{\min}\left(\widehat{\boldsymbol{\Sigma}}_{ij}\right) \le \lambda_{\max}\left(\widehat{\boldsymbol{\Sigma}}_{ij}\right)$ holds , hence

$$\frac{\lambda_0}{2T}\left|\left(\widehat{\Sigma}_{11,i}\right)^{-1}\operatorname{sign}(\beta_i)\right| \le \frac{\lambda_0}{2c_0T}\left\|\operatorname{sign}(\beta_{\mathrm{i}})\right\|_2 \le \sqrt{s_0}\frac{\lambda_0}{2c_0T},$$

for some positive constant c_0. Denote $\boldsymbol{\Sigma}^* = \boldsymbol{X}'\boldsymbol{L}'\boldsymbol{L}\boldsymbol{X}$, $\widehat{\boldsymbol{\varepsilon}} = \widehat{\boldsymbol{L}}\boldsymbol{u}$, and $\boldsymbol{\varepsilon}^* = \boldsymbol{L}\boldsymbol{u}$ where $\boldsymbol{L}$ is defined in (N.1), then A of (B.2) becomes:

$$A \le \sum_{i \in S_0} P\left(\frac{1}{\sqrt{T}}\left|\left(\widehat{\Sigma}_{11,i}\right)^{-1}\widetilde{\boldsymbol{x}}_i'\widehat{\boldsymbol{\varepsilon}} - \left(\widehat{\Sigma}_{11,i}\right)^{-1}\widetilde{\boldsymbol{x}}_i'\boldsymbol{\varepsilon}^*\right| \ge \sqrt{T}\,|\beta_i| - \frac{\lambda_0\sqrt{s_0}}{8c_0T}\right)$$

$$+\sum_{i\in S_0} P\left(\frac{1}{\sqrt{T}}\left|\left(\widehat{\Sigma}_{11,i}\right)^{-1}\widetilde{\boldsymbol{x}}_i'\boldsymbol{\varepsilon}-\left(\Sigma^*_{11,i}\right)^{-1}\widetilde{\boldsymbol{x}}_i'\boldsymbol{\varepsilon}^*\right|\geq\sqrt{T}\left|\beta_i\right|-\frac{\lambda_0\sqrt{s_0}}{8c_0T}\right)$$
$$+\sum_{i\in S_0} P\left(\frac{1}{\sqrt{T}}\left|\left(\Sigma^*_{11,i}\right)^{-1}\widetilde{\boldsymbol{x}}_i'\boldsymbol{\varepsilon}^*-\left(\Sigma^*_{11,i}\right)^{-1}\boldsymbol{x}_i^{*\prime}\boldsymbol{\varepsilon}^*\right|\geq\sqrt{T}\left|\beta_i\right|-\frac{\lambda_0\sqrt{s_0}}{8c_0T}\right)$$
$$+\sum_{i\in S_0} P\left(\frac{1}{\sqrt{T}}\left|\left(\Sigma^*_{11,i}\right)^{-1}\boldsymbol{x}_i^{*\prime}\boldsymbol{\varepsilon}^*\right|\geq\sqrt{T}\left|\beta_i\right|-\frac{\lambda_0\sqrt{s_0}}{8c_0T}\right)=A_1+A_2+A_3+A_4.$$

Terms $A_1,\ldots,A_4$ are bounded following similar analysis as in Lemma C.4, hence $A\leq A_4$. Further, notice that similarly with $\widehat{\boldsymbol{\Sigma}}_{ij}$, $\forall\ i,j=1,2,\ 0<\Lambda_{\min}\left(\boldsymbol{\Sigma}^*_{ij}\right)\leq\Lambda_{\max}\left(\boldsymbol{\Sigma}^*_{ij}\right)$. Therefore by Assumption 5 and using Lemma A3 of Chudik et al. (2018) we have that,

$$A_4\leq\sum_{i\in s_0}P\left(\frac{1}{\sqrt{T}}\left|c_{t,i}\right|\geq\frac{\left[\sqrt{T}\left|\beta_i\right|-\frac{\lambda_0\sqrt{s_0}}{16c_0T}\right]}{C}\right)\rightarrow 0$$

for some large enough constants $d,C>0$, and

$$\lambda_0=\begin{cases}\sqrt{\log p/T}, & \boldsymbol{c}_t\in\mathcal{T}(r),r>0\\ p^{\frac{1}{\theta}}T^{-\left(\frac{\theta}{2}+1\right)}, & \boldsymbol{c}_t\in\mathcal{H}(\theta),\theta>4.\end{cases}$$

Similar analysis is conducted for B. The result follows. □

Proof of Theorem 3

We show that

$$\widehat{t}_s=\frac{\sqrt{T}\boldsymbol{a}'\left(\widehat{\boldsymbol{b}}-\boldsymbol{\beta}\right)}{\sqrt{\boldsymbol{a}'\widehat{\boldsymbol{\Theta}}\widehat{\boldsymbol{\Sigma}}_{xu}\widehat{\boldsymbol{\Theta}}'\boldsymbol{a}}}\sim\mathcal{N}(0,1),\tag{B.3}$$

where $\widehat{\boldsymbol{b}}=(\widehat{b}_1,\ldots,\widehat{b}_p)'$ a vector of DEBIASED GLS estimates of $\boldsymbol{\beta}$, defined in (20) and $\widehat{\boldsymbol{\Theta}}=\widehat{\boldsymbol{M}}^{-1}\widehat{\boldsymbol{C}}$ is a $p\times p$ matrix, where $\widehat{\boldsymbol{M}}^{-1}=\mathrm{diag}(\tau_1^2,\ldots,\tau_p^2)'$ and $\widehat{C}$ defined in Section 4.1 of the main paper. By (23) we can write

$$\widehat{t}_s=A+B,\quad\text{where}\tag{B.4}$$

$$A=\frac{\boldsymbol{a}'\widehat{\boldsymbol{\Theta}}'\widetilde{\boldsymbol{X}}'\widehat{\boldsymbol{L}}\boldsymbol{u}/T}{\sqrt{\boldsymbol{a}'\widehat{\boldsymbol{\Theta}}\widehat{\boldsymbol{\Sigma}}_{xu}\widehat{\boldsymbol{\Theta}}'\boldsymbol{a}}},\quad B=-\frac{\boldsymbol{a}'\boldsymbol{\delta}}{\sqrt{\boldsymbol{a}'\widehat{\boldsymbol{\Theta}}\widehat{\boldsymbol{\Sigma}}_{xu}\widehat{\boldsymbol{\Theta}}'\boldsymbol{a}}},\quad\boldsymbol{\delta}=\sqrt{T}\left[\widehat{\boldsymbol{\Theta}}\widehat{\boldsymbol{\Sigma}}-\boldsymbol{I}_{p\times p}\right]\left(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}\right),$$

where $\widehat{\boldsymbol{L}}$ is defined similarly to (N.1). It is sufficient to show that $A\sim\mathcal{N}\left(0,1\right)$ and $B=o_P\left(1\right)$. We start by showing that $A\sim\mathcal{N}\left(0,1\right)$. Denote

$$\widehat{\boldsymbol{\Sigma}}_{xu}=T^{-1}\left(\widetilde{\boldsymbol{X}}'\widehat{\boldsymbol{L}}\boldsymbol{u}\right)\left(\widetilde{\boldsymbol{X}}'\widehat{\boldsymbol{L}}\boldsymbol{u}\right)',\quad\boldsymbol{\Sigma}^*_{xu}=T^{-1}\left(\boldsymbol{X}'\boldsymbol{L}\boldsymbol{u}\right)\left(\boldsymbol{X}'\boldsymbol{L}\boldsymbol{u}\right)'.$$

To show that $A \sim \mathcal{N}(0,1)$, we first need to show that

$$A' = \frac{\boldsymbol{a}'\boldsymbol{\Theta}^*\boldsymbol{X}^{*\prime}\boldsymbol{\varepsilon}^*/T}{\sqrt{T\boldsymbol{a}'\boldsymbol{\Theta}^*\boldsymbol{\Sigma}^*_{xu}\boldsymbol{\Theta}^{*\prime}\boldsymbol{a}}} \sim \mathcal{N}(0,1)\,, \text{ and } A - A' = o_P(1), \tag{B.5}$$

where $\boldsymbol{X}^{*\prime}\boldsymbol{\varepsilon}^* = \boldsymbol{X}'\boldsymbol{L}'\boldsymbol{L}\boldsymbol{u}$.

Step 1: We show that the nominator and denominator of A is asymptotically equivalent to their corresponding quantities of A'. Starting with the nominator, by Lemma C.4.

$$\frac{1}{\sqrt{T}}\left\|\boldsymbol{a}'\widehat{\boldsymbol{\Theta}}'\widetilde{\boldsymbol{X}}'\widehat{\boldsymbol{\varepsilon}} - \boldsymbol{a}'\boldsymbol{\Theta}^{*\prime}\boldsymbol{X}^{*\prime}\boldsymbol{\varepsilon}^*\right\|_1 = o_P(1), \tag{B.6}$$

where $\widehat{\boldsymbol{\varepsilon}} = \widehat{\boldsymbol{L}}\boldsymbol{u}$, and $\boldsymbol{\varepsilon}^* = \boldsymbol{L}\boldsymbol{u}$ where $\boldsymbol{L}$ is defined similarly to (N.1). Further, for the denominator, by Lemma C.5

$$\left|\boldsymbol{a}'\widehat{\boldsymbol{\Theta}}\widehat{\boldsymbol{\Sigma}}_{xu}\widehat{\boldsymbol{\Theta}}'\boldsymbol{a} - \boldsymbol{a}'\boldsymbol{\Theta}^*\boldsymbol{\Sigma}^*_{xu}\boldsymbol{\Theta}^{*\prime}\boldsymbol{a}\right| = o_P(1). \tag{B.7}$$

Hence, by (B.6)–(B.7), $A - A' = o_P(1)$.

We now show that $A' \sim \mathcal{N}(0,1)$. We remark that $\boldsymbol{a}'\boldsymbol{\Theta}^*\boldsymbol{\Sigma}^*_{xu}\boldsymbol{\Theta}^*\boldsymbol{a}$ is asymptotically bounded away from zero (positive definite), such that the following statement holds:

$$\boldsymbol{a}'\boldsymbol{\Theta}^*\boldsymbol{\Sigma}^*_{xu}\boldsymbol{\Theta}^*\boldsymbol{a} \geq \Lambda_{\min}(\boldsymbol{\Sigma}^*_{xu})\|\boldsymbol{\Theta}'\boldsymbol{a}\|_2^2 \geq \Lambda_{\min}(\boldsymbol{\Sigma}^*_{xu})\Lambda^2_{\min}(\boldsymbol{\Theta}^*)\|\boldsymbol{a}\|_2^2 \geq \Lambda_{\min}(\boldsymbol{\Sigma}^*_{xu})\Lambda^{-2}_{\min}(\boldsymbol{\Theta}^*), \tag{B.8}$$

where $\|\boldsymbol{a}\|_2 = 1$, $\Lambda_{\min}(\boldsymbol{\Sigma}^*_{xu})$ the smallest eigenvalue of $\boldsymbol{\Sigma}^*_{xu}$, and $\Lambda_{\min}(\boldsymbol{\Theta}^*)$ the smallest eigenvalue of $\boldsymbol{\Theta}$, which obey $0 < \Lambda_{\min}(\boldsymbol{\Theta}^*) \leq \Lambda_{\max}(\boldsymbol{\Theta}^*) < \infty$. Then, by Assumption 5, $\{\boldsymbol{x}^{*\prime}_t\varepsilon^*_t\}$ is a zero-mean stationary process, by consequence $E\left[(\boldsymbol{x}^{*\prime}_1\varepsilon^*_1)(\boldsymbol{x}^{*\prime}_1\varepsilon^*_1)'\right] = \text{Var}[\boldsymbol{x}^{*\prime}_1\varepsilon^*_1] > 0$, $E\left[\boldsymbol{x}^{*\prime}_1\varepsilon^*_1\right] = 0$. Taking the expectation of A'_1 and $(A'_1)^2$ we obtain

$$E\left[\frac{\boldsymbol{a}'\boldsymbol{\Theta}^*_i\boldsymbol{x}^{*\prime}_t\varepsilon^*_t}{\sqrt{T\boldsymbol{a}'\boldsymbol{\Theta}^*\boldsymbol{\Sigma}^*_{xu}\boldsymbol{\Theta}^{*\prime}\boldsymbol{a}}}\right] = 0, \tag{B.9}$$

$$E\left[\frac{\boldsymbol{a}'\boldsymbol{\Theta}^*\boldsymbol{x}^{*\prime}_t\varepsilon^*_t}{\sqrt{T\boldsymbol{a}'\boldsymbol{\Theta}^*\boldsymbol{\Sigma}^*_{xu}\boldsymbol{\Theta}^*\boldsymbol{a}'}}\right]^2 = E\left[\frac{T^{-1}\boldsymbol{a}'\left(\boldsymbol{\Theta}^*\boldsymbol{x}^{*\prime}_t\varepsilon^*_t\right)'\boldsymbol{a}'\left(\boldsymbol{\Theta}^*\boldsymbol{x}^{*\prime}_t\varepsilon^*_t\right)}{T^{-1}\boldsymbol{a}'\left(\boldsymbol{\Theta}^*_i\boldsymbol{x}^{*\prime}_t\varepsilon^*_t\right)'\boldsymbol{a}'\left(\boldsymbol{\Theta}^*_i\boldsymbol{x}^{*\prime}_t\varepsilon^*_t\right)}\right] = 1. \tag{B.10}$$

In view of Theorem 24.6 of Davidson (1994), A' is asymptotically standard normal, across $t = 1, \ldots, T$.

Step 2: It remains to show that $B = o_P(1)$. The denominators of A, B are identical, so by (B.8), the denominator of B is asymptotically positive definite. It suffices to show that $\boldsymbol{\delta} = \sqrt{T}(\widehat{\boldsymbol{\Theta}}\widehat{\boldsymbol{\Sigma}} - \boldsymbol{I})(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})$ is asymptotically negligible. As an implication of Proposition C.1, $\|\widehat{\boldsymbol{\Theta}}_i\widehat{\boldsymbol{\Sigma}} - \boldsymbol{e}_i\|_\infty = o_P(\lambda_i)$ and by Theorem 1, $\|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}\|_1 = O_P(s_0\lambda)$, where the order of λ can be defined according to (18), so then

$$\begin{aligned}\|\boldsymbol{a}'\boldsymbol{\delta}\|_1 &\leq \left\|\sqrt{T}\left(\widehat{\boldsymbol{\Theta}}\widehat{\boldsymbol{\Sigma}} - \boldsymbol{I}\right)\left(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right)\right\|_\infty \|\boldsymbol{a}\|_1 \\ &\leq \sqrt{T}\max_{i=1,\ldots,p}\left\|\widehat{\boldsymbol{\Theta}}_i\widehat{\boldsymbol{\Sigma}} - \boldsymbol{e}_i\right\|_\infty\left\|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right\|_1\|\boldsymbol{a}\|_1 = o_P(1),\end{aligned}$$

showing that $B = o_P(1)$. $\square$

C Auxiliary results

PROPOSITION C.1. *Let Assumption 2 hold and let* $\widehat{\boldsymbol{\Sigma}} = T^{-1}\widetilde{\boldsymbol{X}}'\widetilde{\boldsymbol{X}}$. *Then* $\widehat{\boldsymbol{\Theta}}$ *is a good approximation of* $\widehat{\boldsymbol{\Sigma}}$ *uniformly for all* $i = 1, \dots, p$ *if*

$$\frac{\left\|\widetilde{\boldsymbol{X}}'_{-i}\widetilde{\boldsymbol{X}}\widehat{\boldsymbol{\Theta}}_i\right\|_\infty}{T} = O_P\left(\lambda\right), \tag{C.1}$$

where the order of λ *is defined in* (18), $\widehat{\boldsymbol{\Theta}} = \widehat{\boldsymbol{T}}^{-2}\widehat{\boldsymbol{C}}$, $\widehat{\boldsymbol{T}}^2 = \operatorname{diag}\left(\widehat{\tau}_1^2, \dots, \widehat{\tau}_{\mathrm{p}}^2\right)$ *and* $\widehat{\boldsymbol{C}}$ *as defined in Section 4.1 of the main paper.*

LEMMA C.1. *Let Assumption 3 hold. If further,* $\lambda_i \asymp \sqrt{\frac{\log p}{T}}$ *and for* $p >> T$, *we have*

$$\left\|\widetilde{\boldsymbol{X}}_{-i}\left(\widehat{\boldsymbol{\gamma}}_i - \boldsymbol{\gamma}_i\right)\right\|_2^2 = O_P\left(s_i\lambda_i\right), \quad \left\|\boldsymbol{X}^*_{-i}\left(\widehat{\boldsymbol{\gamma}}^*_i - \boldsymbol{\gamma}_i\right)\right\|_2^2 = O_P\left(s_i\lambda_i\right), \tag{C.2}$$

$$\left\|\widehat{\boldsymbol{\gamma}}_i - \boldsymbol{\gamma}_i\right\|_1 = O_P\left(s_i\sqrt{\lambda_i}\right), \quad \left\|\widehat{\boldsymbol{\gamma}}^*_i - \boldsymbol{\gamma}_i\right\|_1 = O_P\left(s_i\sqrt{\lambda_i}\right), \tag{C.3}$$

where the order of λ *is defined in* (18).

COROLLARY C.1. *Under Assumptions 3, 6, 7 and 9 and for a suitable choice of the regularisation parameter, with order defined in* (18), *we have*

$$\max_i \left\{\frac{1}{\widehat{\tau}_i^2}\right\} = O_P(1), \quad i = 1, \dots, p. \tag{C.4}$$

LEMMA C.2. *Suppose Assumption 3 an 6 hold, then*

$$|\widehat{\tau}_i^2 - \tau_i^2| = O_P\left(\sqrt{s_i\lambda_i}\right), \quad \left|\widehat{\tau}_i^{-2} - \tau_i^{-2}\right| = O_P\left(\sqrt{s_i\lambda_i}\right), \tag{C.5}$$

for a suitable choice of the regularisation parameter, with order defined in (18)*,and* $\lambda_i s_i = o(1)$.

LEMMA C.3. *Suppose Assumptions 1–9 hold, then the following statements hold*

$$\|\widehat{\boldsymbol{\Theta}}_i - \boldsymbol{\Theta}_i\|_1 = O_P\left(s_i\lambda_i\right) \tag{C.6}$$

$$\|\widehat{\boldsymbol{\Theta}}_i - \boldsymbol{\Theta}_i\|_2 = O_P\left(\sqrt{s_i}\lambda_i\right) \tag{C.7}$$

$$\|\widehat{\boldsymbol{\Theta}}_i\|_1 = O_P\left(\sqrt{s_i}\right) \tag{C.8}$$

$$\|\boldsymbol{\Theta}_i\|_1 = O\left(\sqrt{s_i}\right) \tag{C.9}$$

for some suitable choice of λ_i, *with order defined in* (18) *and* $\lambda_i s_i = o(1)$. *and* $s_i := |\{j \neq i : \Theta_{i,j} \neq 0\}|$.

LEMMA C.4. *Under Assumptions 1–9, we have for* $i = 1, \dots, p$,

$$\frac{1}{\sqrt{T}}\left\|\boldsymbol{a}'\widehat{\boldsymbol{\Theta}}\widetilde{\boldsymbol{X}}'\widehat{\boldsymbol{L}}\boldsymbol{u} - \boldsymbol{a}'\boldsymbol{\Theta}\boldsymbol{X}'\boldsymbol{L}\boldsymbol{u}\right\|_1 = o_P\left(1\right), \tag{C.10}$$

where $\widehat{\mathbf{\Theta}}$ *is defined in Section 4.1 of the main paper, while* $\mathbf{\Theta}$ *is its population counterpart, and* $\widetilde{\boldsymbol{X}} = \widehat{\boldsymbol{L}}\boldsymbol{X}$, *where* $\widehat{\boldsymbol{L}}$ *is defined in* (N.1).

LEMMA C.5. *Under Assumption 3, we have* $\forall\, i \in \{1,\ldots,p\}$,

$$\left|\widehat{\mathbf{\Theta}}_i\widehat{\mathbf{\Sigma}}_{xu}\widehat{\mathbf{\Theta}}_i' - \mathbf{\Theta}_i\mathbf{\Sigma}_{xu}\mathbf{\Theta}_i'\right| = o_p(1), \tag{C.11}$$

where $\widehat{\mathbf{\Theta}}_i$, $\widehat{\mathbf{\Sigma}}_{xu}$ *are defined in Section 4 of the main paper.*

LEMMA C.6. *Under Assumptions 3 – 7,*

$$\mathrm{P}\left(\max_{0<i,k\leq p}\left|\widehat{\mathbf{\Sigma}}_{i,k} - \mathbf{\Sigma}_{i,k}\right| > \nu\right) \to 0, \quad \textit{for some} \quad \nu > 0. \tag{C.12}$$

where $\widehat{\mathbf{\Sigma}} = T^{-1}\widetilde{\boldsymbol{X}}'\widetilde{\boldsymbol{X}}$, *and* $\mathbf{\Sigma} = E(T^{-1}\boldsymbol{X}^{*\prime}\boldsymbol{X}^*)$, *for some large enough positive and finite constants,* $c_1, \nu > 0$ *and* $\delta_1, \delta_2 > 0$ *independent from* T *and* p.

D Proofs of auxiliary results

This section contains auxiliary technical lemmas in Section C.

Proof of Proposition C.1

We show that $\widehat{\mathbf{\Theta}}$ is a good approximation of $\widehat{\mathbf{\Sigma}}^{-1}$, for $\widehat{\mathbf{\Sigma}} = T^{-1}\left(\widetilde{\boldsymbol{X}}'\widetilde{\boldsymbol{X}}\right)$ the sample variance-covariance matrix. Immediate application of the KKT condition on $\widehat{\boldsymbol{\gamma}}_i^2$ gives

$$\frac{1}{T}\widetilde{\boldsymbol{X}}_{-i}'\left(\widetilde{\boldsymbol{x}}_i - \widetilde{\boldsymbol{X}}_{-i}\widehat{\boldsymbol{\gamma}}_i\right) = \lambda_i\widehat{\boldsymbol{\eta}}_i. \tag{D.1}$$

It suffices to show that

$$\frac{1}{T}\boldsymbol{X}_{-i}^{*}{}'\left(\boldsymbol{x}_i^* - \boldsymbol{X}_{-i}^*\widehat{\boldsymbol{\gamma}}_i^*\right) + o_P(1) = \left(\lambda_i\widehat{\boldsymbol{\eta}}_i - \lambda_i\widehat{\boldsymbol{\eta}}_i^*\right) + \lambda_i\widehat{\boldsymbol{\eta}}_i^*, \tag{D.2}$$

$$\frac{1}{T}\widetilde{\boldsymbol{X}}_{-i}'\left(\widetilde{\boldsymbol{x}}_i - \widetilde{\boldsymbol{X}}_{-i}\widehat{\boldsymbol{\gamma}}_i^*\right) - \frac{1}{T}\boldsymbol{X}_{-i}^{*}{}'\left(\boldsymbol{x}_i^* - \boldsymbol{X}_{-i}^*\widehat{\boldsymbol{\gamma}}_i^*\right) = o_P(1), \tag{D.3}$$

where (D.3), the remainder term on (D.2), and $\widehat{\boldsymbol{\eta}}_i = \mathrm{sign}\left(\widehat{\boldsymbol{\gamma}}_i\right)$, $\widehat{\boldsymbol{\eta}}_i^* = \mathrm{sign}\left(\widehat{\boldsymbol{\gamma}}_i^*\right)$. Notice that

$$\frac{\boldsymbol{X}_{-i}^{*}{}'\left(\boldsymbol{x}_i^* - \boldsymbol{X}_{-i}^*\widehat{\boldsymbol{\gamma}}_i^*\right)}{T} = \lambda_i\widehat{\boldsymbol{\eta}}_i^*,$$

are the KKT conditions of a node-wise regression, $\boldsymbol{x}_i^*|\boldsymbol{X}_{-i}^*$, and $\widehat{\boldsymbol{\gamma}}_i^*$ the corresponding estimates, $\forall\, i = 1,\ldots,p$. Following similar arguments to Theorem 1, $\|(\widehat{\boldsymbol{\gamma}}_i - \widehat{\boldsymbol{\gamma}}_i{}^*)\|_1 = O_P\left(s_0\sqrt{\lambda_i}\right)$, while $\|\widehat{\boldsymbol{\gamma}}_i^*\|_1$ will have the same properties as $\|\widehat{\boldsymbol{\gamma}}_i\|_1$. We proceed to show (D.3), using the scalar representations of the processes involved,

$$\begin{aligned}&\widetilde{\boldsymbol{X}}_{-i}'\left(\widetilde{\boldsymbol{x}}_i - \widetilde{\boldsymbol{X}}_{-i}\widehat{\boldsymbol{\gamma}}_i\right)/T - \boldsymbol{X}_{-i}^{*}{}'\left(\boldsymbol{x}_i^* - \boldsymbol{X}_{-i}^*\widehat{\boldsymbol{\gamma}}_i^*\right)/T\\ &= \left(\sum_{t=1}^{T}\boldsymbol{x}_{t,-i} - \sum_{j=1}^{q}\widehat{\phi}_j\boldsymbol{x}_{t-j,-i}\right)\left[\left(\sum_{t=1}^{T}x_{t,i} - \sum_{j=1}^{q}\widehat{\phi}_j x_{t-j,i}\right)\right.\end{aligned}$$

$$
\begin{aligned}
&\qquad\qquad - \left(\sum_{t=1}^{T} \boldsymbol{x}_{t,-i} - \sum_{j=1}^{q} \widehat{\phi}_j \boldsymbol{x}_{t-j,-i} \right) \widehat{\boldsymbol{\gamma}}_i \Bigg] / T \\
&- \left(\sum_{t=1}^{T} \boldsymbol{x}_{t,-i} - \sum_{j=1}^{q} \phi_j \boldsymbol{x}_{t-j,-i} \right) \Bigg[\left(\sum_{t=1}^{T} x_{t,i} - \sum_{j=1}^{q} \phi_j x_{t-j,i} \right) \\
&\qquad\qquad - \left(\sum_{t=1}^{T} \boldsymbol{x}_{t,-i} - \sum_{j=1}^{q} \phi_j \boldsymbol{x}_{t-j,-i} \right) \widehat{\boldsymbol{\gamma}}_i^* \Bigg] / T \\
\leq & \frac{1}{T} \sum_{t=1}^{T} \boldsymbol{x}_{t,-i} \sum_{j=1}^{q} \left(\widehat{\phi}_j - \phi_j \right) x_{t-j,i} + \frac{1}{T} \| \widehat{\boldsymbol{\gamma}}_i - \widehat{\boldsymbol{\gamma}}_i^* \|_1 \sum_{t=1}^{T} \boldsymbol{x}_{t,-i}' \boldsymbol{x}_{t,-i} \\
&+ \frac{1}{T} \| \widehat{\boldsymbol{\gamma}}_i - \widehat{\boldsymbol{\gamma}}^*{}_i \|_1 \sum_{t=1}^{T} \boldsymbol{x}_{t,-i} \sum_{j=1}^{q} \boldsymbol{x}_{t-j,-i} \left(\widehat{\phi}_j - \phi_j \right) + \frac{1}{T} \sum_{t=1}^{T} x_{t,i} \sum_{j=1}^{q} \boldsymbol{x}_{t-j,-i} \left(\widehat{\phi}_j - \phi_j \right) \\
&+ \frac{1}{T} \sum_{t=1}^{T} \sum_{j=1}^{q} \boldsymbol{x}_{t-j,-i} x_{t-j,i} \left[\left(\widehat{\phi}_j - \phi_j \right)^2 + \left(\widehat{\phi}_j - \phi_j \right) 2\phi_j \right] \\
&+ \frac{1}{T} \sum_{t=1}^{T} x_{t,i} \sum_{j=1}^{q} \boldsymbol{x}_{t-j,-i} \left(\widehat{\phi}_j - \phi_j \right) \\
&+ \frac{1}{T} \| \widehat{\boldsymbol{\gamma}}_i - \widehat{\boldsymbol{\gamma}}^*{}_i \|_1 \sum_{t=1}^{T} \sum_{j=1}^{q} \boldsymbol{x}_{t-j,-i} x_{t-j,-i}' \left[\left(\widehat{\phi}_j - \phi_j \right)^2 + \left(\widehat{\phi}_j - \phi_j \right) 2\phi_j \right] \\
= & O_P(s_i T^{-1/2}) + O_P(s_i \sqrt{\lambda_i}) + O_P(s_i \sqrt{\lambda_i / T}) + O_P(T^{-1/2}) \qquad \text{(D.4)} \\
&+ O_P(T^{-1} \vee T^{-1/2}) + O_P(T^{-1}) + O_P(s_i \sqrt{\lambda_i} / T \vee s_i \sqrt{\lambda_i / T}), \qquad \text{(D.5)}
\end{aligned}
$$

Notice that $\widehat{\boldsymbol{\gamma}}_i \lambda_i \widehat{\boldsymbol{\eta}}_i = \lambda_i \| \widehat{\boldsymbol{\gamma}}_i \|_1$, and $\widehat{\boldsymbol{\gamma}}_i^* \lambda_i \widehat{\boldsymbol{\eta}}_i^* = \lambda_i \| \widehat{\boldsymbol{\gamma}}_i^* \|_1$,

$$
(\widehat{\boldsymbol{\gamma}}_i - \widehat{\boldsymbol{\gamma}}^*{}_i) = (\widehat{\boldsymbol{\gamma}}_i - \boldsymbol{\gamma}_i) + (\boldsymbol{\gamma}_i - \widehat{\boldsymbol{\gamma}}^*{}_i) \leq \| \widehat{\boldsymbol{\gamma}}_i - \boldsymbol{\gamma}_i \|_1 + \| \boldsymbol{\gamma}_i - \widehat{\boldsymbol{\gamma}}^*{}_i \|_1 = O_P\left(s_i \lambda_i \right), \qquad \text{(D.6)}
$$

where the first term obtains its rate of convergence by Lemma C.1 and the second term has the same rate of convergence following similar arguments as in Lemma C.1. Completing the proof of (D.3). Then we can show that

$$
\begin{aligned}
| \widehat{\boldsymbol{\gamma}}_i \lambda_i \widehat{\boldsymbol{\eta}}_i - \widehat{\boldsymbol{\gamma}}_i^* \lambda_i \widehat{\boldsymbol{\eta}}_i^* | &\leq | \widehat{\boldsymbol{\gamma}}_i \lambda_i \widehat{\boldsymbol{\eta}}_i - \widehat{\boldsymbol{\gamma}}_i^* \lambda_i \widehat{\boldsymbol{\eta}}_i | + | \widehat{\boldsymbol{\gamma}}_i^* \lambda_i \widehat{\boldsymbol{\eta}}_i - \widehat{\boldsymbol{\gamma}}_i^* \lambda_i \widehat{\boldsymbol{\eta}}_i^* | \qquad \text{(D.7)} \\
&\leq \lambda_i \| \widehat{\boldsymbol{\eta}}_i \|_\infty \| \widehat{\boldsymbol{\gamma}}_i - \widehat{\boldsymbol{\gamma}}_i^* \|_1 + \lambda_i \| \widehat{\boldsymbol{\eta}}_i^* \|_\infty \| \widehat{\boldsymbol{\gamma}}_i^* \|_1 , \qquad \text{(D.8)}
\end{aligned}
$$

where $\| \widehat{\boldsymbol{\eta}}_i^* \|_\infty \leq 1$ as the sub-gradient of $\| \widehat{\boldsymbol{\gamma}}_i^* \|_1$, $\| \widehat{\boldsymbol{\eta}}_i \|_\infty \leq 1$ as the sub-gradient of $\| \widehat{\boldsymbol{\gamma}}_i \|_1$, and $\| \widehat{\boldsymbol{\gamma}}_i - \widehat{\boldsymbol{\gamma}}_i^* \|_1 = O_P\left(s_i \lambda_i^{1/2} \right)$, leading to $| \widehat{\boldsymbol{\gamma}}_i \lambda_i \widehat{\boldsymbol{\eta}}_i - \widehat{\boldsymbol{\gamma}}_i^* \lambda_i \widehat{\boldsymbol{\eta}}_i^* | = o_P(1)$ by (D.6). Considering the analysis on (D.3) and (D.8), (D.2) becomes

$$
\widehat{\boldsymbol{\gamma}}_i^* \boldsymbol{X}_{-i}^*{}' \left(\boldsymbol{x}_i^* - \boldsymbol{X}_{-i}^* \widehat{\boldsymbol{\gamma}}_i^* \right) / T + o_P(1) = \lambda_i \| \widehat{\boldsymbol{\gamma}}_i^* \|_1, \qquad \text{(D.9)}
$$

by plugging (D.9) into $(\widehat{\tau}_i^*)^2 = T^{-1} \left\| \boldsymbol{x}_i^* - \boldsymbol{X}_{-i}^* \widehat{\boldsymbol{\gamma}}_i^* \right\|_2^2 + \lambda_i \| \widehat{\boldsymbol{\gamma}}_i^* \|_1$, then,

$$
(\widehat{\tau}_i^*)^2 = T^{-1} \left(\left(\boldsymbol{x}_i^* - \boldsymbol{X}_{-i}^* \widehat{\boldsymbol{\gamma}}_i^* \right)' \left(\boldsymbol{x}_i^* - \boldsymbol{X}_{-i}^* \widehat{\boldsymbol{\gamma}}_i^* \right) + \widehat{\boldsymbol{\gamma}}_i \boldsymbol{X}_{-i}^*{}' \left(\boldsymbol{x}_i^* - \boldsymbol{X}_{-i}^* \widehat{\boldsymbol{\gamma}}_i^* \right) \right) \qquad \text{(D.10)}
$$

$$= T^{-1}\left(\boldsymbol{x}_i^* - \boldsymbol{X}_{-i}^*\widehat{\gamma}_i^*\right)\left(\left(\boldsymbol{x}_i^* - \boldsymbol{X}_{-i}^*\widehat{\gamma}_i^*\right) + \widehat{\gamma}_i^*\boldsymbol{X}_{-i}^{*}{}'\right) \tag{D.11}$$

$$= T^{-1}\left(\boldsymbol{x}_i^* - \widetilde{\boldsymbol{X}}_{-i}^*\widehat{\gamma}_i^*\right)\boldsymbol{x}_i^*. \tag{D.12}$$

Recall the definition of $\widehat{\boldsymbol{C}}$ in Section 4.1 of the main paper and notice that $\widehat{\boldsymbol{c}}_i$, is the i-the row of $\widehat{\boldsymbol{C}}$, we get that $\widetilde{\boldsymbol{X}}\widehat{\boldsymbol{c}}_i = \left(\widetilde{\boldsymbol{x}}_i - \widetilde{\boldsymbol{X}}_{-i}\widehat{\gamma}_i\right)$ and $\widehat{\boldsymbol{\Theta}}_i = \widehat{\boldsymbol{c}}_i/\widehat{\tau}_i^2$. Equation (D.12) then becomes

$$(\tau_i^*)^2 = T^{-1}\boldsymbol{X}_{-i}^{*}{}'\boldsymbol{X}^*\widehat{\boldsymbol{c}}_i^* \Rightarrow T^{-1}\boldsymbol{X}_{-i}^{*}{}'\boldsymbol{X}^*\boldsymbol{\Theta}_i^* = 1, \tag{D.13}$$

where $\boldsymbol{\Theta}_i^* = \widehat{\boldsymbol{c}}_i^*/(\tau_i^*)^2$. Then, (D.13) will hold if $\widehat{\boldsymbol{\Theta}}$ is a good approximation of $\widehat{\boldsymbol{\Sigma}}^{-1} = \left(\widetilde{\boldsymbol{X}}'\widetilde{\boldsymbol{X}}/T\right)^{-1}$, in the sense that the approximation error $\left\|\widehat{\boldsymbol{\Theta}}\widehat{\boldsymbol{\Sigma}} - I\right\|_\infty = o_P(1)$, which we now evaluate.

For (C.1) to hold, the following expression should hold as well.

$$\left\|\widehat{\boldsymbol{\Theta}}_i\widehat{\boldsymbol{\Sigma}} - \boldsymbol{e}_i\right\|_\infty \le \left\|\widehat{\boldsymbol{\Theta}}_i\widehat{\boldsymbol{\Sigma}} - \widehat{\boldsymbol{\Theta}}_i\widehat{\boldsymbol{\Sigma}}^*\right\|_\infty + \left\|\widehat{\boldsymbol{\Theta}}_i\widehat{\boldsymbol{\Sigma}}^* - \boldsymbol{\Theta}_i^*\widehat{\boldsymbol{\Sigma}}^*\right\|_\infty + \left\|\boldsymbol{\Theta}_i^*\widehat{\boldsymbol{\Sigma}}^* - \boldsymbol{e}_i\right\|_\infty \tag{D.14}$$

$$= B_1 + B_2 + B_3, \tag{D.15}$$

where $\widehat{\boldsymbol{\Sigma}} = \sum_{t=1}^T\left(\widetilde{\boldsymbol{x}}_t\widetilde{\boldsymbol{x}}_t'\right)/T$, $\widehat{\boldsymbol{\Sigma}}^* = \sum_{t=1}^T\left(\boldsymbol{x}_t^*\boldsymbol{x}_t^{*\prime}\right)/T$. We wish to show that $B_1, B_2 = o_P(1)$ and B_3 attains a bound such that the one in (C.1). For B_1: $B_1 \le \left\|\widehat{\boldsymbol{\Sigma}} - \widehat{\boldsymbol{\Sigma}}^*\right\|_\infty \|\widehat{\boldsymbol{\Theta}}_i\|_1$. Then, we have:

$$\left\|\widehat{\boldsymbol{\Sigma}} - \widehat{\boldsymbol{\Sigma}}^*\right\|_\infty \le \left\|\frac{1}{T}\left(\boldsymbol{X}'\widehat{\boldsymbol{L}}'\widehat{\boldsymbol{L}}\boldsymbol{X} - \boldsymbol{X}'\boldsymbol{L}'\boldsymbol{L}\boldsymbol{X}\right)\right\|_\infty$$

$$\underset{(1),(2)}{\le} \frac{1}{T}\left\|\boldsymbol{X}'\widehat{\boldsymbol{L}}'\left(\widehat{\boldsymbol{L}} - \boldsymbol{L}\right)\boldsymbol{X}\right\|_1 + \frac{1}{T}\left\|\boldsymbol{X}'\left(\widehat{\boldsymbol{L}} - \boldsymbol{L}\right)'\boldsymbol{L}\boldsymbol{X}\right\|_1 = B_{11} + B_{12}, \tag{D.16}$$

where (1) results from the norm inequality (8) of Chapter 8.5.2 of Lutkepohl (1997) and (2) results by applying the triangle inequality. For $i = 1, \dots, p,\ s = 1, \dots, T - q$, we analyse B_{12}:

$$B_{12} \le \frac{1}{T}\max_i\left|\sum_{s=1}^{T-q}\boldsymbol{x}_i'(\widehat{\boldsymbol{l}}_s - \boldsymbol{l}_s)'\boldsymbol{l}_s\boldsymbol{x}_i\right| \le \max_i\left|\frac{1}{T}\boldsymbol{x}_i'\boldsymbol{l}_s'\boldsymbol{x}_i\right|\left\|\sum_{s=1}^{T-q}(\widehat{\boldsymbol{l}}_s - \boldsymbol{l}_s)\right\|_1$$

$$= O_P(s_0^2\lambda^2 \vee T^{-1/2}s_0\lambda) \tag{D.17}$$

where $\boldsymbol{l}_s$, $\widehat{\boldsymbol{l}}_s$ two $1 \times T$ row vectors of the matrices $\boldsymbol{L}$, $\widehat{\boldsymbol{L}}$ respectively, defined in (N.1). The asymptotic rate of B_{12} results by directly applying Corollary 1. Following a similar analysis to B_{12}, the term $B_{11} = O_P(s_i\lambda_i)$. By Lemma C.3, we have that $\|\widehat{\boldsymbol{\Theta}}_i\|_1 = O_P\left(\sqrt{s_i}\right)$, hence $B_1 = O_P\left(qs_0^2\lambda_i\right) + O_P\left(qs_0\sqrt{\lambda_i}\right)$. We find that $B_2 \asymp B_1$. The asymptotic rate of B_2 is obtained by (D.28). Using $\widehat{\boldsymbol{\Sigma}}^* = T^{-1}\boldsymbol{X}^{*\prime}\boldsymbol{X}^*$ and expression (A.16), we get that $\|\widehat{\boldsymbol{\Sigma}}^*\|_\infty = O_P(1)$. We then continue the analysis for B_3, to derive the result. We examine the column norm of $\|\boldsymbol{\Theta}_i^*\widehat{\boldsymbol{\Sigma}}^* - \boldsymbol{e}_i\|_\infty$, where $\boldsymbol{e}_i$ is the i^{th} column of the identity matrix $\boldsymbol{I}_{p\times p}$. By (D.13), $\left\|T^{-1}\boldsymbol{X}^{*\prime}_{-i}\boldsymbol{X}^*\boldsymbol{\Theta}_i^*\right\|_\infty = \|\boldsymbol{\Theta}_i^*\widehat{\boldsymbol{\Sigma}}^* - \boldsymbol{e}_i\|_\infty$, incorporating $\boldsymbol{\Theta}_i^* = \widehat{\boldsymbol{c}}_i^*/(\tau_i^*)^2$ into $\left\|T^{-1}\boldsymbol{X}^{*\prime}_{-i}\boldsymbol{X}^*\boldsymbol{\Theta}_i^*\right\|_\infty$,

$$\left\|T^{-1}\boldsymbol{X}^{*\prime}_{-i}\boldsymbol{X}^*\boldsymbol{\Theta}_i^*\right\|_\infty = \left\|T^{-1}\boldsymbol{X}^{*\prime}_{-i}\boldsymbol{X}^*\widehat{\boldsymbol{c}}_i^*/(\tau_i^*)^2\right\|_\infty \le \lambda_i\frac{\|\widehat{\boldsymbol{\eta}}_i^*\|_\infty}{(\tau_i^*)^2} \le \frac{\lambda_i}{(\tau_i^*)^2}, \tag{D.18}$$

then (D.18) holds, since $\|\widehat{\boldsymbol{\eta}}_i^*\|_\infty := \sup_i \|\widehat{\boldsymbol{\eta}}_i^*\|_1 \leq 1$. Further, using (D.15) and since $\left\|T^{-1}\boldsymbol{X}^{*\prime}_{-i}\boldsymbol{X}^*\boldsymbol{\Theta}_i^*\right\|_\infty = \|\boldsymbol{\Theta}_i^*\widehat{\boldsymbol{\Sigma}}^* - \boldsymbol{e}_i\|_\infty$, we obtain

$$\left\|\boldsymbol{\Theta}_i^*\widehat{\boldsymbol{\Sigma}}^* - \boldsymbol{e}_i\right\|_\infty \leq \frac{\lambda_i}{(\tau_i^*)^2} + o_P(1), \tag{D.19}$$

where $(\tau_i^*)^2 = \boldsymbol{x}_i^{*\prime}\left(\boldsymbol{x}_i^* - \boldsymbol{X}^*_{-i}\boldsymbol{\gamma}_i^*\right)$. It holds that $\max_i\{1/(\tau_i^*)^2\} = O_P(1)$ by direct application of Corollary C.1, the result follows. □

Proof of Lemma C.1

The proof follows the same line of arguments as Theorem 1 and Corollary 2. □

Proof of Corollary C.1

We show that the errors of the node-wise regressions uniformly on i are small. First we make the following standard statements: The compatibility condition holds uniformly for all node-wise regressions, and the corresponding compatibility constant is bounded away from zero. Further,

$$1/\tau_i^2 = \boldsymbol{\Theta}_{i,i}^* \geq \Lambda_{\min}^2(\boldsymbol{\Theta}^*) > 0, \quad \forall\, i = 1, \dots, p,$$

and $\tau_i^2 \leq E\boldsymbol{x}_i^{*\prime}\boldsymbol{x}_i^* = \Sigma_{i,i}^* = O_P(1)$. Now, each node-wise regression has bounded prediction and estimation errors as in Lemma C.1, and we can then write

$$\begin{aligned}
&\left[\left\|\widetilde{\boldsymbol{x}}_i - \widetilde{\boldsymbol{X}}_{-i}\widehat{\gamma}_i\right\|_2^2/T - \left\|\boldsymbol{x}_i^* - \boldsymbol{X}^*_{-i}\widehat{\gamma}_i^*\right\|_2^2/T\right] + \left\|\boldsymbol{x}_i^* - \boldsymbol{X}^*_{-i}\widehat{\gamma}_i^*\right\|_2^2/T \\
&= A_1 + \left\|\boldsymbol{x}_i^* - \boldsymbol{X}^*_{-i}\widehat{\gamma}_i^*\right\|_2^2/T \\
&\underset{(1)}{=} o_p(1) + \left\|\boldsymbol{x}_i^* - \boldsymbol{X}^*_{-i}\widehat{\gamma}_i^*\right\|_2^2/T \\
&= o_p(1) + \left\|\boldsymbol{x}_i^* - \boldsymbol{X}^*_{-i}\boldsymbol{\gamma}_i^*\right\|_2^2/T + \left\|\boldsymbol{X}^*_{-i}(\widehat{\gamma}_i^* - \boldsymbol{\gamma}_i^*)\right\|_2^2/T + 2\left(\boldsymbol{x}_i^* - \boldsymbol{X}^*_{-i}\boldsymbol{\gamma}_i^*\right)'\boldsymbol{X}^*_{-i}(\widehat{\gamma}_i^* - \boldsymbol{\gamma}_i^*) \\
&= o_P(1) + \tau_i^2 + O_P(s_i\lambda_i) + O_P\left(\sqrt{s_i\lambda_i}\right) + O_P\left(\sqrt{s_i\lambda_i}\right) = o_p(1) + \tau_i^2,
\end{aligned}$$

where $\boldsymbol{\gamma}_i^* = \underset{\boldsymbol{\gamma}_i^* \in \mathbb{R}^p}{\arg\min}\left\{E\left[(x_{t,i}^* - \boldsymbol{x}_{-i,t}^*\boldsymbol{\gamma}_i^*)^2\right]\right\}$, consequently, $\widetilde{\boldsymbol{\gamma}}_i$ can be defined following the same logic. Above we used the fact that $A_1 = o_P(1)$, which we now prove.

Notice that (1) results from the following expressions:

$$\begin{aligned}
A_1 &= \left\|\widetilde{\boldsymbol{x}}_i - \widetilde{\boldsymbol{X}}_{-i}\widehat{\gamma}_i\right\|_2^2/T - \left\|\boldsymbol{x}_i^* - \boldsymbol{X}^*_{-i}\widehat{\gamma}_i^*\right\|_2^2/T \\
&= \widetilde{\boldsymbol{x}}_i'\widetilde{\boldsymbol{x}}_i/T - 2\widetilde{\boldsymbol{x}}_i'\widetilde{\boldsymbol{X}}_{-i}\widehat{\gamma}_i/T + \left\|\widetilde{\boldsymbol{X}}_{-i}\widehat{\gamma}_i\right\|_2^2/T - \boldsymbol{x}_i^{*\prime}\boldsymbol{x}_i^*/T \\
&\quad + 2\boldsymbol{x}_i^{*\prime}\boldsymbol{X}^*_{-i}\widehat{\gamma}_i^*/T - \left\|\boldsymbol{X}^*_{-i}\widehat{\gamma}_i^*\right\|_2^2/T \\
&\leq \left\|\widehat{\boldsymbol{\Sigma}} - \widehat{\boldsymbol{\Sigma}}^*\right\|_\infty + \frac{1}{T}\left(2\boldsymbol{\gamma}_i^{*\prime}\boldsymbol{X}^{*\prime}_{-i}\boldsymbol{X}^*_{-i}\widehat{\gamma}_i^* + 2\boldsymbol{v}_i^{*\prime}\boldsymbol{X}^*_{-i}\widehat{\gamma}_i^* - 2\widetilde{\gamma}_i'\widetilde{\boldsymbol{X}}'_{-i}\widetilde{\boldsymbol{X}}_{-i}\widehat{\gamma}_i' + 2\widetilde{\boldsymbol{v}}_i'\widetilde{\boldsymbol{X}}_{-i}\widehat{\gamma}_i\right) \\
&\quad + \left(\left\|\widetilde{\boldsymbol{X}}_{-i}\widehat{\gamma}_i\right\|_2^2/T - \left\|\boldsymbol{X}^*_{-i}\widehat{\gamma}_i^*\right\|_2^2/T\right) = a_1 + a_2 + a_3.
\end{aligned}$$

By (D.16), $a_1 = O_P(T^{-1/2})$. For $1 \leq i \leq p$ and $1 \leq j \leq T$, we have that

$$a_2 = \left(2\boldsymbol{\gamma}_i^{*\prime}\boldsymbol{X}^{*\prime}_{-i}\boldsymbol{X}^*_{-i}\widehat{\gamma}_i^* + 2\boldsymbol{v}_i^{*\prime}\boldsymbol{X}^*_{-i}\widehat{\gamma}_i^* - 2\widetilde{\gamma}_i'\widetilde{\boldsymbol{X}}'_{-i}\widetilde{\boldsymbol{X}}_{-i}\widehat{\gamma}_i' + 2\widetilde{\boldsymbol{v}}_i'\widetilde{\boldsymbol{X}}_{-i}\widehat{\gamma}_i\right)/T \tag{D.20}$$

$$
\begin{aligned}
&\le \frac{1}{T}\Big[\left\|\widetilde{\boldsymbol{X}}_{-i}\right\|_2^2 \|\widetilde{\gamma}_i\|_2 \|(\widehat{\gamma}_i-\widetilde{\gamma}_i)\|_2 + \left\|\widetilde{\boldsymbol{v}}_i'\widetilde{\boldsymbol{X}}_{-i}\right\|_2 \|\widetilde{\gamma}_i\|_2 \|(\widehat{\gamma}_i-\widetilde{\gamma}_i)\|_2 \\
&\quad - \left\|\boldsymbol{X}^*_{-i}\right\|_2^2 \|\boldsymbol{\gamma}^*_i\|_2 \|(\widehat{\gamma}^*_i-\boldsymbol{\gamma}^*_i)\|_2 + \left\|\boldsymbol{v}_i^{*\prime}\boldsymbol{X}^*{}_{-i}\right\|_2 \|\widetilde{\gamma}_i\|_2 \|(\widehat{\gamma}_i-\widetilde{\gamma}_i)\|_2 \\
&\quad - \left\|\boldsymbol{X}^*_{-i}\right\|_2^2 \|\boldsymbol{\gamma}^*_i\|_2 \|(\widehat{\gamma}^*_i-\boldsymbol{\gamma}^*_i)\|_2\Big] \\
&\le \frac{1}{T}\Big[\left\|\widetilde{\boldsymbol{X}}_{-i}\right\|_2^2 O_P(\sqrt{s_i})\, O_P\left(s_i\sqrt{\lambda_i}\right) + \left\|\widetilde{\boldsymbol{v}}_i{}'\widetilde{\boldsymbol{X}}_{-i}\right\|_2 O_P(\sqrt{s_i})\, O_P\left(s_i\sqrt{\lambda_i}\right) \\
&+ \left\|\boldsymbol{X}^*_{-i}\right\|_2^2 O_P(\sqrt{s_i})\, O_P\left(s_i\sqrt{\lambda_i}\right) + \left\|\boldsymbol{X}^{*\prime}{}_{-i}\boldsymbol{v}^*_i\right\|_2 O_P(\sqrt{s_i})\, O_P\left(s_i\sqrt{\lambda_i}\right)\Big] \\
&= a_{21}+a_{22}+a_{23}+a_{24}.
\end{aligned}
$$

Using the reverse triangle inequality we have,

$$
\begin{aligned}
\left|\left\|\widetilde{\boldsymbol{X}}_{-i}\right\|_2^2 - \left\|\boldsymbol{X}^*_{-i}\right\|_2^2\right| O_P\left(s_i^{3/2}\frac{\sqrt{\lambda_i}}{T}\right) &\le \left\|\widetilde{\boldsymbol{X}}_{-i} - \boldsymbol{X}^*_{-i}\right\|_2^2 O_P\left(s_i^{3/2}\frac{\sqrt{\lambda_i}}{T}\right) \\
&\le \left\|\widehat{\boldsymbol{L}}\boldsymbol{X}_{-i}/T - \boldsymbol{L}\boldsymbol{X}_{-i}/T\right\|_2^2 O_P\left(s_i^{3/2}\sqrt{\lambda_i}\right) \\
&\le \left\|\left(\widehat{\boldsymbol{L}}-\boldsymbol{L}\right)\boldsymbol{X}_{-i}/T\right\|_2^2 O_P\left(s_i^{3/2}\sqrt{\lambda_i}\right) \\
&\le \left\|\boldsymbol{X}'_{-i}\boldsymbol{X}_{-i}/T\right\|_\infty \left\|\widehat{\boldsymbol{L}}-\boldsymbol{L}\right\|_1 O_P\left(s_i^{3/2}\sqrt{\lambda_i}\right). \qquad \text{(D.21)}
\end{aligned}
$$

Using similar arguments to prove (D.16), we have that $a_{21} - a_{23} = o_P(1)$. Following a similar line of arguments and utilising the definitions of $\boldsymbol{v}^*_i = E[\boldsymbol{x}^*_i - \boldsymbol{X}^*_i\boldsymbol{\gamma}^*_i]$ and $\widetilde{\boldsymbol{v}}_i = E[\widetilde{\boldsymbol{x}}_i - \widetilde{\boldsymbol{X}}_i\widetilde{\gamma}_i]$, we conclude that $a_2 = O_P(s_i^{3/2}\sqrt{\lambda_i})$. For a_3 and using the reverse triangle inequality we have,

$$
\begin{aligned}
a_3 &\le \left|\left\|\widetilde{\boldsymbol{X}}_{-i}\widehat{\gamma}_i\right\|_2^2/T - \left\|\boldsymbol{X}^*_{-i}\widehat{\gamma}^*_i\right\|_2^2/T\right| \\
&\le \left\|\widetilde{\boldsymbol{X}}_{-i}\widehat{\gamma}_i/T - \boldsymbol{X}^*_{-i}\widehat{\gamma}^*_i/T\right\|_2^2 \\
&= \left\|\left(\widehat{\boldsymbol{L}}\boldsymbol{X}_{-i}\widehat{\gamma}_i - \widehat{\boldsymbol{L}}\boldsymbol{X}_{-i}\boldsymbol{\gamma}_i\right) + \left(\widehat{\boldsymbol{L}}\boldsymbol{X}_{-i}\boldsymbol{\gamma}_i - \boldsymbol{L}\boldsymbol{X}_{-i}\boldsymbol{\gamma}_i\right) + \left(\boldsymbol{L}\boldsymbol{X}_{-i}\boldsymbol{\gamma}_i - \boldsymbol{L}\boldsymbol{X}_{-i}\boldsymbol{\gamma}^*_i\right)\right\|_2^2/T \\
&\le \left(\left\|\widehat{\boldsymbol{L}}\boldsymbol{X}_{-i}\widehat{\gamma}_i - \widehat{\boldsymbol{L}}\boldsymbol{X}_{-i}\boldsymbol{\gamma}_i\right\| + \left\|\widehat{\boldsymbol{L}}\boldsymbol{X}_{-i}\boldsymbol{\gamma}_i - \boldsymbol{L}\boldsymbol{X}_{-i}\boldsymbol{\gamma}_i\right\| + \left\|\boldsymbol{L}\boldsymbol{X}_{-i}\boldsymbol{\gamma}_i - \boldsymbol{L}\boldsymbol{X}_{-i}\boldsymbol{\gamma}^*_i\right\|\right)^2/T \\
&= a_{31}+a_{32}+a_{33}. \qquad \text{(D.22)}
\end{aligned}
$$

To show that $a_3 = o_P(1)$, we first show that $a_{31}, a_{32}, a_{33} = o_P(1)$. We start with a_{31}:

$$
\begin{aligned}
a_{31} &= \left\|\widehat{\boldsymbol{L}}\boldsymbol{X}_{-i}\widehat{\gamma}_i - \widehat{\boldsymbol{L}}\boldsymbol{X}_{-i}\widetilde{\gamma}_i \pm \boldsymbol{L}\boldsymbol{X}_{-i}\widehat{\gamma}_i \pm \boldsymbol{L}\boldsymbol{X}_{-i}\widetilde{\gamma}_i\right\|/T \\
&\le \left\|\widehat{\boldsymbol{L}}-\boldsymbol{L}\right\|_1 \|\boldsymbol{X}_{-i}\|_2 \|\widetilde{\gamma}_i\|_2 \|(\widehat{\gamma}_i-\widetilde{\gamma}_i)\|_2/T + \left\|\widehat{\boldsymbol{L}}-\boldsymbol{L}\right\|_1 \|\boldsymbol{X}_{-i}\|_2 \|\widetilde{\gamma}_i\|_2/T \\
&+ \|\boldsymbol{L}\boldsymbol{X}_{-i}\|_2 \|\widetilde{\gamma}_i\|_2 \|(\widehat{\gamma}_i-\widetilde{\gamma}_i)\|_2/T \\
&\underset{(1)}{=} O_P(s_0^2\lambda_i)O(1)O_P\left(s_i\sqrt{\lambda_i}\right) \vee O_P(s_0^2\lambda_i)O(1)O_(\sqrt{s_i}) \\
&\quad \vee O(1)O(\sqrt{s_i})O_P\left(s_i\sqrt{\lambda_i}\right), \qquad \text{(D.23)}
\end{aligned}
$$

where (1) results by following similar arguments as the ones to show (D.16), further the rate of $\|\widetilde{\gamma}_i\|_2$ and $\|(\widehat{\gamma}_i - \widetilde{\gamma}_i)\|_2$, is obtained by implication of Lemma C.1. Regarding $\|\widetilde{\gamma}_i\|_1$, we can write

$$\|\widetilde{\gamma}_i\|_1 \leq \lambda_i \left(\|\widetilde{\gamma}_i\|_1 + \|\widehat{\gamma}_i - \widetilde{\gamma}_i\|_1\right) = \lambda_i O(\sqrt{s_i}) + \lambda_i O_P(\sqrt{s_i \lambda_i}). \tag{D.24}$$

Lastly a_{32}, a_{33} follow similar analysis to a_{31}. Hence, we conclude that $a_{31}, a_{32}, a_{33} = o_P(1)$, showing that $A_1 = o_P(1)$. Showing a similar result as in (D.24) for $\|\boldsymbol{\gamma}_i^*\|_1$, we proceed to show (C.4):

$$\begin{aligned} \max_i \left\{1/\widehat{\tau}^2\right\} &= \max_i \left\{\left[(\tau_i)^2 + O_P\left(\sqrt{s_i/T}\right) + O_P\left(s_i\lambda_i^2\right) + O_P\left(\sqrt{s_i\lambda_i}\right)\right]^{-1}\right\} \\ &\leq \max_i \left\{\left[(\tau_i)^2 + o_P\left(1\right)\right]^{-1}\right\}, \end{aligned}$$

where $(\tau_i)^2 \leq \Sigma^*_{i,i} = O_P(1)$, completing the proof. □

Proof of Lemma C.2.

Before we proceed to the analysis, it is useful to point out that, from the KKT conditions of the node-wise regressions of $\widetilde{\boldsymbol{x}}_i|\widetilde{\boldsymbol{X}}_{-i}$ we have $\widehat{\tau}_i^2 = \widetilde{\boldsymbol{x}}_i'(\widetilde{\boldsymbol{x}}_i - \widetilde{\boldsymbol{X}}_{-i}\widehat{\gamma}_i)$ with residuals $\widehat{\boldsymbol{v}}_i = (\widetilde{\boldsymbol{x}}_i - \widetilde{\boldsymbol{X}}_{-i}\widehat{\gamma}_i)$, further on the same note, the KKT conditions of the node-wise regressions of $\boldsymbol{x}_i^*|\boldsymbol{X}^*_{-i}$ we have $(\widehat{\tau}_i^*)^2 = \boldsymbol{x}_i^{*\prime}(\boldsymbol{x}_i^* - \boldsymbol{X}^*_{-i}\widehat{\gamma}_i{}^*)$ with residuals $\widehat{\boldsymbol{v}}_i^* = (\boldsymbol{x}_i^* - \boldsymbol{X}^*_{-i}\widehat{\gamma}_i{}^*)$ and $\tau_i^2 = E\left[\left(\boldsymbol{x}_i^* - \sum_{i\neq k}\gamma_{i,k}\boldsymbol{X}^*_{-i}\right)^2\right]$, with population residuals, $E\left(\upsilon_i^*\right) = E\left[\boldsymbol{x}_i^* - \boldsymbol{X}^*_{-i}\gamma_i^*\right]$.

We first consider $|\widehat{\tau}_i^2 - \tau_i^2|$. Then we have that

$$\left|\widehat{\tau}_i^2 - \tau_i^2\right| \leq \left|\widehat{\tau}_i^2 - (\widehat{\tau}_i^*)^2\right| + \left|(\widehat{\tau}_i^*)^2 - \tau_i^2\ \right| = B_1 + B_2.$$

The first part of the Lemma follows if $B_1,\ B_2 = o_P\left(1\right)$, we proceed to analyse B_1, where $1 \leq i \leq p,\ 1 \leq j \leq q$:

$$\begin{aligned} B_1 &= T^{-1}\left|\widetilde{\boldsymbol{x}}_i'\widehat{\boldsymbol{v}}_i - (\boldsymbol{x}_i^*)\prime(\widehat{\boldsymbol{v}}_i^*)\right| \\ &= T^{-1}\left|(\widehat{\boldsymbol{l}}_j\boldsymbol{x}_i)'\left(\widehat{\boldsymbol{l}}_j\boldsymbol{x}_i - \widehat{\boldsymbol{l}}_j\boldsymbol{X}_{-i}\widehat{\gamma}_i\right) - (\boldsymbol{l}_j\boldsymbol{x}_i)'\left(\boldsymbol{l}_j\boldsymbol{x}_i - \boldsymbol{l}_j\boldsymbol{X}_{-i}\widehat{\gamma}_i^*\right)\right| \\ &\leq T^{-1}\left|\boldsymbol{x}_i'\widehat{\boldsymbol{l}}_j'\widehat{\boldsymbol{l}}_j\boldsymbol{X}_{-i}\widehat{\gamma}_i - \boldsymbol{x}_i'\boldsymbol{l}_j{}'\boldsymbol{l}_j\boldsymbol{X}_{-i}\widehat{\gamma}_i^*\right| + T^{-1}\left\|\boldsymbol{x}_i'\left(\widehat{\boldsymbol{l}}_j'\widehat{\boldsymbol{l}}_j - \boldsymbol{l}_j{}'\boldsymbol{l}_j\right)\boldsymbol{x}_i\right\|_1 \\ &= b_1 + b_2. \end{aligned}$$

Following a similar analysis as in (D.22), (D.16), it can be shown that $b_2 = O_P(s_0^2 \log pT^{-1})$. We analyse b_1

$$\begin{aligned} b_1 &= \left|\boldsymbol{x}_i'\widehat{\boldsymbol{l}}_j'\widehat{\boldsymbol{l}}_j\boldsymbol{X}_{-i}\widehat{\gamma}_i' - \boldsymbol{x}_i'\boldsymbol{l}_j{}'\boldsymbol{l}_j\boldsymbol{X}_{-i}\widehat{\gamma}_i^* \pm \boldsymbol{x}_i'\boldsymbol{l}_j{}'\boldsymbol{l}_j\boldsymbol{X}_{-i}\widehat{\gamma}_i \pm \boldsymbol{x}_i'\boldsymbol{l}_j{}'\boldsymbol{l}_j\boldsymbol{X}_{-i}\gamma_i^*\right|/T \\ &= \left|\boldsymbol{x}_i'\left(\widehat{\boldsymbol{l}}_j'\widehat{\boldsymbol{l}}_j - \boldsymbol{l}_j{}'\boldsymbol{l}_j\right)\boldsymbol{X}_{-i}\widehat{\gamma}_i\right|/T + \left|\boldsymbol{x}_i'\boldsymbol{l}_j{}'\boldsymbol{l}_j\boldsymbol{X}_{-i}\left(\widehat{\gamma}_i^* - \boldsymbol{\gamma}_i\right)\right|/T \\ &\quad + \left|\boldsymbol{x}_i'\boldsymbol{l}_j{}'\boldsymbol{l}_j\boldsymbol{X}_{-i}\left(\widehat{\gamma}_i - \boldsymbol{\gamma}_i\right)\right|/T = b_{11} + b_{12} + b_{13}. \end{aligned}$$

Further, b_{11} can be analysed such that

$$\begin{aligned}
b_{11} &= \left|\boldsymbol{x}_i' \left(\widehat{\boldsymbol{l}}_j'\widehat{\boldsymbol{l}}_j - \boldsymbol{l}_j'\boldsymbol{l}_j\right) \boldsymbol{X}_{-i}\widehat{\boldsymbol{\gamma}}_i\right| /T \\
&\leq \frac{1}{T}\left\|\boldsymbol{x}_i' \left(\widehat{\boldsymbol{l}}_j'\widehat{\boldsymbol{l}}_j - \boldsymbol{l}_j'\boldsymbol{l}_j\right) \boldsymbol{X}_{-i}\right\|_1 \|\widehat{\boldsymbol{\gamma}}_i - \boldsymbol{\gamma}_i\|_1 \qquad \text{(D.25)}\\
&\quad + \frac{1}{T}\left\|\boldsymbol{x}_i' \left(\widehat{\boldsymbol{l}}_j'\widehat{\boldsymbol{l}}_j - \boldsymbol{l}_j'\boldsymbol{l}_j\right) \boldsymbol{X}_{-i}\right\|_1 \|\boldsymbol{\gamma}_i\|_1 \qquad \text{(D.26)}\\
&\underset{(1)}{=} O_P\left(s_i^2\lambda_i^2\right) O_P\left(s_i\sqrt{\lambda_i}\right) + O_P\left(s_i^2\lambda_i^2\right) O_P\left(\sqrt{s_i}\right).
\end{aligned}$$

The first and third term in (D.25)–(D.26) can be analysed following similar analysis as (D.16), the second term is obtained by directly applying Corollary C.1 and the last term is an implication of Corollary C.1, hence $b_{11} = O_P\left(s_i T^{-1/2}\lambda_i\right)$. Following the analysis in (D.23) and by direct application of Corollary C.1, we conclude that $b_{12}, b_{13} = O_P\left(s_i\lambda_i\right)$. Concluding, $B_1 = b_1 \vee b_2 = O_P\left(s_i\lambda_i\right)$, and B_2 follows the same analysis with B_1. Therefore the following holds,

$$\left|\widehat{\tau}_i^2 - \tau_i^2\right| = O_P\left(\sqrt{s_i}\lambda_i\right).$$

To show (C.5), note that $\tau_i^2 = \frac{1}{\Theta_{i,i}} \geq \frac{1}{\Lambda_{\max}(\boldsymbol{\Theta})} = \Lambda_{\min}\left(\boldsymbol{\Sigma}\right)$, for all $i = 1, \ldots, p$. Recall that $\Lambda_{min}\left(\boldsymbol{\Sigma}\right) > 0$, thus

$$\widehat{\tau}_i^2 \leq \left|\widehat{\tau}_i^2 - (\widehat{\tau}_i^*)^2\right| + \left|(\widehat{\tau}_i^*)^2 - \tau_i^2\right| + \tau_i^2 = \tau_i^2 + o_P(1) > 0,$$

which taken together with Corollary C.1, imply

$$\left|\frac{1}{\widehat{\tau}_i^2} - \frac{1}{{\tau_i}^2}\right| \leq \left|\frac{1}{\widehat{\tau}_i^2} - \frac{1}{(\widehat{\tau}_i^*)^2}\right| + \left|\frac{1}{(\widehat{\tau}_i^*)^2} - \frac{1}{{\tau_i}^2}\right| = \frac{\left|\widehat{\tau}_i^2 - (\widehat{\tau}_i^*)^2\right|}{\widehat{\tau}_i^2(\widehat{\tau}_i^*)^2} + \frac{\left|(\widehat{\tau}_i^*)^2 - \tau_i^2\right|}{(\widehat{\tau}_i^*)^2\tau_i^2} = O_P\left(\sqrt{s_i}\lambda_i\right),$$

completing the proof. □

Proof of Lemma C.3

In this proof, we largely follow Kock (2016b). By definition $\widehat{\boldsymbol{\Theta}}_i = \widehat{\boldsymbol{c}}_i/\widehat{\tau}_i^2$, similar arguments hold for $\boldsymbol{\Theta}_i^*$, then

$$\begin{aligned}
\|\widehat{\boldsymbol{\Theta}}'\boldsymbol{a} - \boldsymbol{\Theta}'\boldsymbol{a}\|_1 &\leq \|\widehat{\boldsymbol{\Theta}}'\boldsymbol{a} - \boldsymbol{\Theta}^{*\prime}\boldsymbol{a}\|_1 + \|\boldsymbol{\Theta}^{*\prime}\boldsymbol{a} - \boldsymbol{\Theta}'\boldsymbol{a}\|_1 \\
&\leq \|(\widehat{\boldsymbol{\Theta}} - \boldsymbol{\Theta}^*)'\|_1\|\boldsymbol{a}\|_1 + \|(\boldsymbol{\Theta}^* - \boldsymbol{\Theta})'\|_1\|\boldsymbol{a}\|_1 = [C_1 + C_2]\|\boldsymbol{a}\|_1. \qquad \text{(D.27)}
\end{aligned}$$

Note that $\|\boldsymbol{a}\|_2 = 1$ and $\|\boldsymbol{a}\|_2 \leq \|\boldsymbol{a}\|_1$, then for C_1, we have that

$$\begin{aligned}
C_1 &= \left\|\frac{\widehat{\boldsymbol{c}}_i}{\widehat{\tau}_i^2} - \frac{\boldsymbol{c}_i^*}{(\widehat{\tau}_i^*)^2}\right\|_1 = \left\|\frac{1 - \widehat{\boldsymbol{\gamma}}_i}{\widehat{\tau}_i^2} - \frac{1 - \widehat{\boldsymbol{\gamma}}_i^*}{(\widehat{\tau}_i^*)^2}\right\|_1 \\
&= \left\|\frac{1}{\widehat{\tau}_i^2} - \frac{1}{(\widehat{\tau}_i^*)^2} + \frac{\widehat{\boldsymbol{\gamma}}_i^*}{\widehat{\tau}_i^2} - \frac{\widehat{\boldsymbol{\gamma}}_i^*}{\widehat{\tau}_i^2} + \frac{\widehat{\boldsymbol{\gamma}}_i^*}{(\widehat{\tau}_i^*)^2} - \frac{\widehat{\boldsymbol{\gamma}}_i}{\widehat{\tau}_i^2}\right\|_1 \\
&\leq \left\|\frac{1}{\widehat{\tau}_i^2} - \frac{1}{(\widehat{\tau}_i^*)^2}\right\|_1 + \|\widehat{\boldsymbol{\gamma}}_i^*\|_1\left\|\frac{1}{\widehat{\tau}_i^2} - \frac{1}{(\widehat{\tau}_i^*)^2}\right\|_1 + \|\widehat{\boldsymbol{\gamma}}_i - \widehat{\boldsymbol{\gamma}}_i^*\|_1\left\|\frac{1}{\widehat{\tau}_i^2}\right\|_1 \\
&= O_P\left(\sqrt{s_i}\lambda_i\right) + O(\sqrt{s_i})\, O_P\left(\sqrt{s_i}\lambda_i\right) + O_P\left(s_i\lambda_i\right) O_P(1) \\
&= O_P\left(s_i\lambda_i\right). \qquad \text{(D.28)}
\end{aligned}$$

The first term of (D.28) is a direct result of Lemma C.2, the second results from (D.24), and Lemma C.2 and the third term results from (D.6), and direct application of Corollary D.12. Following similar arguments to (D.28), we can show that $\|\widehat{\mathbf{\Theta}}_i - \mathbf{\Theta}_i\|_2 = O_P(\sqrt{s_i\lambda_i})$. By definition of $\|\widehat{\mathbf{\Theta}}_i\|_1$, $\widehat{\mathbf{\Theta}}_i = \widehat{\tau}_i^{-2}\widehat{\boldsymbol{c}}_i$ (similar arguments hold for $\mathbf{\Theta}_i^*$), we have

$$\|\mathbf{\Theta}_i\|_1 = \|\tau_i{}^{-2}\boldsymbol{c}_i\|_1 = \left\|\frac{1-\boldsymbol{\gamma}_i}{\tau_i^2}\right\|_1 \leq \left\|\frac{1}{\tau_i^2}\right\|_1 - \|\boldsymbol{\gamma}_i\|_1\left\|\frac{1}{\tau_i^2}\right\|_1 = O(\sqrt{s_i})$$

$$\begin{aligned}\|\widehat{\mathbf{\Theta}}_i\|_1 &\leq \|\widehat{\mathbf{\Theta}}_i - \mathbf{\Theta}_i^*\|_1 + \|\mathbf{\Theta}_i^*\|_1 \\ &\leq O_P\left(s_i\lambda_i\right) + \left\|\frac{1}{(\widehat{\tau}_i^*)^2}\right\|_\infty - \|\widehat{\boldsymbol{\gamma}}_i^* - \boldsymbol{\gamma}_i\|_1\left\|\frac{1}{(\widehat{\tau}_i^*)^2}\right\|_\infty - \|\boldsymbol{\gamma}_i\|_1\left\|\frac{1}{(\widehat{\tau}_i^*)^2}\right\|_\infty \\ &= O_P(1) + O_P\left(s_i\lambda_i\right)O_P(1) + O(\sqrt{s_i})O_P(1) = O_P(\sqrt{s_i}).\end{aligned} \tag{D.29}$$

Further, note that $\left\|\mathbf{\Theta}^{*\prime}\boldsymbol{a}\right\|_2 \leq \left\|\mathbf{\Theta}^{*\prime}\boldsymbol{a}\right\|_1 \leq \sum_{i\in\{1,\dots\tilde{p}\}}|a_i|\|\mathbf{\Theta}_i^*\|_1$, Then,by direct application of Corollary C.1, from which we get that $\left\|\frac{1}{\tau_i{}^2}\right\|_\infty = O(1)$ and $\left\|\frac{1}{(\widehat{\tau}_i^*)^2}\right\|_\infty = O_P(1)$, from Lemma C.1 we get $\|\widehat{\boldsymbol{\gamma}}_i^* - \boldsymbol{\gamma}_i\|_1 = O_P(s_i\lambda_i)$, and by following similar arguments as (D.24), we have $\|\boldsymbol{\gamma}_i\|_1 = O(\sqrt{s_i})$.

Using a similar line of arguments with C_1, we have that $C_2 = O_P\left(s_i\lambda_i\right)$, completing the proof. □

Proof of Lemma C.4

We use the following decomposition of the left hand side of (C.10):

$$\begin{aligned}\frac{1}{\sqrt{T}}\left\|\boldsymbol{a}'\widehat{\mathbf{\Theta}}\widetilde{\boldsymbol{X}}'\widehat{\boldsymbol{L}}\boldsymbol{u} - \boldsymbol{a}'\mathbf{\Theta}\boldsymbol{X}'\boldsymbol{L}\boldsymbol{u}\right\|_1 \leq \frac{1}{\sqrt{T}}\Big[&\left\|\boldsymbol{a}'\widehat{\mathbf{\Theta}}\widetilde{\boldsymbol{X}}'\widehat{\boldsymbol{L}}\boldsymbol{u} - \boldsymbol{a}'\mathbf{\Theta}^*\widetilde{\boldsymbol{X}}'\widehat{\boldsymbol{L}}\boldsymbol{u}\right\|_1 \\ &+ \left\|\boldsymbol{a}'\mathbf{\Theta}^*\widetilde{\boldsymbol{X}}'\widehat{\boldsymbol{L}}\boldsymbol{u} - \boldsymbol{a}'\mathbf{\Theta}^*\boldsymbol{X}^{*\prime}\widehat{\boldsymbol{L}}\boldsymbol{u}\right\|_1 \\ &+ \left\|\boldsymbol{a}'\mathbf{\Theta}^*\boldsymbol{X}^{*\prime}\widehat{\boldsymbol{L}}\boldsymbol{u} - \boldsymbol{a}'\mathbf{\Theta}^*\boldsymbol{X}^{*\prime}\boldsymbol{L}\boldsymbol{u}\right\|_1 \\ &+ \left\|\boldsymbol{a}'\mathbf{\Theta}^*\boldsymbol{X}^{*\prime}\boldsymbol{L}\boldsymbol{u} - \boldsymbol{a}'\mathbf{\Theta}\boldsymbol{X}^{*\prime}\boldsymbol{L}\boldsymbol{u}\right\|_1 \\ &+ \left\|\boldsymbol{a}'\mathbf{\Theta}\boldsymbol{X}^{*\prime}\boldsymbol{L}\boldsymbol{u} - \boldsymbol{a}'\mathbf{\Theta}\boldsymbol{X}'\boldsymbol{L}\boldsymbol{u}\right\|_1\Big] = \sum_{k=1}^{5} D_k.\end{aligned} \tag{D.30}$$

It suffices to show the following:

$$\begin{aligned}&D_1 = O_P\left(s_iT^{-1/2}\lambda_i\right), \ D_2 = O_P\left(\sqrt{s_i}T^{-1/2}\right), \ D_3 = O_P\left(s_i{}^{3/2}T^{-1/2}\lambda_i\right) \\ &D_4 = O_P\left(\sqrt{s_i}\lambda_i\right), \ D_5 = O\left(\sqrt{s_iT^{-1}}\right).\end{aligned} \tag{D.31}$$

We consider each term, starting with D_1:

$$\begin{aligned}D_1 &\leq \frac{1}{\sqrt{T}}\|\boldsymbol{a}'\|_1\left\|\widehat{\mathbf{\Theta}} - \mathbf{\Theta}^*\right\|_1\left[\left\|\widetilde{\boldsymbol{X}}'\widehat{\boldsymbol{L}}\boldsymbol{u} - \boldsymbol{X}^{*\prime}\widehat{\boldsymbol{L}}\boldsymbol{u}\right\|_1 + \left\|\boldsymbol{X}^{*\prime}\widehat{\boldsymbol{L}}\boldsymbol{u} - \boldsymbol{X}^{*\prime}\boldsymbol{L}\boldsymbol{u}\right\|_1 + \left\|\boldsymbol{X}^{*\prime}\boldsymbol{L}\boldsymbol{u}\right\|_1\right] \\ &= \|\boldsymbol{a}'\|_1\left\|\widehat{\mathbf{\Theta}} - \mathbf{\Theta}^*\right\|_1\left[D_{11} + D_{12} + D_{13}\right].\end{aligned}$$

By (D.29), $\left\|\widehat{\mathbf{\Theta}}_i - \mathbf{\Theta}_i^*\right\|_1 = O_P\left(s_i\lambda_i\right)$, as for term D_{11}, we have that

$$\begin{aligned}\frac{1}{\sqrt{T}}\left\|\widetilde{\boldsymbol{X}}'\widehat{\boldsymbol{L}}\boldsymbol{u} - \boldsymbol{X}^{*\prime}\widehat{\boldsymbol{L}}\boldsymbol{u}\right\|_1 &\leq \frac{1}{\sqrt{T}}\left|\sum_{t=1}^{T}\sum_{j=1}^{q}(\widehat{\phi}_j - \phi_j)\sum_{i=1}^{p}x_{t-j,i}\right| \\ &\quad\times\left|\sum_{t=1}^{T}\left[u_t + \sum_{j=1}^{q}\left[(\widehat{\phi}_j - \phi_j)\sum_{i=1}^{p}x_{t-j,i}(\widehat{\beta}_i - \beta_i)\right] + \sum_{j=1}^{q}\phi_j x_{t-j,i}\right]\right| \\ &\leq \frac{1}{\sqrt{T}}\left|\sum_{t=1}^{T}\sum_{j=1}^{q}(\widehat{\phi}_j - \phi_j)\sum_{i=1}^{p}x_{t-j,i}u_t\right| \\ &\quad + \frac{1}{\sqrt{T}}\left|\sum_{t=1}^{T}\sum_{j=1}^{q}(\widehat{\phi}_j - \phi_j)^2\sum_{i=1}^{p}x_{t-j,i}^2\right|\left|\sum_{i=1}^{p}(\widehat{\beta}_i - \beta_i)\right| \\ &\quad + \frac{1}{\sqrt{T}}\left|\sum_{t=1}^{T}\sum_{j=1}^{q}(\widehat{\phi}_j - \phi_j)\phi_j\sum_{i=1}^{p}x_{t-j,i}^2\right| = d_1 + (d_2\times d_3) + d_4.\end{aligned} \tag{D.32}$$

By Corollary 1 and (A.16) $d_1 = O_P(T^{-1/2})$, $d_2 = O_P(T^{-1/2})$, and by Lemma 1, $d_3 = O_P(s_0\lambda)$, and by (A.16), $d_4 = O_P(1)$, hence $D_{11} = O_P(T^{-1/2})$. We proceed to analyse D_{12}:

$$\frac{1}{\sqrt{T}}\left\|\boldsymbol{X}^{*\prime}\widehat{\boldsymbol{L}}\boldsymbol{u} - \boldsymbol{X}^{*\prime}\boldsymbol{L}\boldsymbol{u}\right\|_1 \leq \frac{1}{\sqrt{T}}\left\|\boldsymbol{X}'\boldsymbol{L}'\boldsymbol{X}\right\|_\infty\left\|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right\|_1 = O_P(s_0\lambda), \tag{D.33}$$

hence $D_{12} = O_P(s_0\lambda)$. We continue with D_{13}, by (A.16), $D_{13} = O_P(s_0\lambda)$. By consequence, $D_1 = O_P(s_0\lambda)$, showing the first part of (D.31). For the second part of (D.31), by (D.29) and (D.32) $D_2 \leq \|\boldsymbol{a}'\|_1\|\mathbf{\Theta}^*\|_1\|\boldsymbol{X}^{*\prime}\widehat{\boldsymbol{u}} - \boldsymbol{X}^{*\prime}\boldsymbol{u}\|_1 = O_P(\sqrt{\frac{s_i}{T}})$. For D_3, by (D.29) and by (D.33) $D_3 = O_P\left({s_i}^{3/2}T^{-1/2}\lambda_i\right)$. For D_4

$$\left\|\boldsymbol{a}'\mathbf{\Theta}^*\boldsymbol{X}^{*\prime}\boldsymbol{L}\boldsymbol{u} - \boldsymbol{a}'\mathbf{\Theta}\boldsymbol{X}^{*\prime}\boldsymbol{L}\boldsymbol{u}\right\|_1 \leq \|\boldsymbol{a}'\|_1\left\|\mathbf{\Theta}^* - \mathbf{\Theta}\right\|_1\left\|\boldsymbol{X}^{*\prime}\boldsymbol{L}\boldsymbol{u}\right\|_1.$$

By the arguments of (D.27), $\|\mathbf{\Theta}_i^* - \mathbf{\Theta}_i\|_1 = O_P(s_i\lambda_i)$ and by Lemma 11 $\left\|\boldsymbol{X}^{*\prime}\boldsymbol{u}\right\|_1 \leq pT^{-1/2}\left\|\boldsymbol{X}^{*\prime}\boldsymbol{u}\right\|_\infty \leq \frac{\sqrt{T}\lambda_0}{2}$. Lastly, for D_4

$$T^{-1/2}\left\|\boldsymbol{a}'\mathbf{\Theta}\boldsymbol{X}^{*\prime}\boldsymbol{L}\boldsymbol{u} - \boldsymbol{a}'\mathbf{\Theta}\boldsymbol{X}'\boldsymbol{L}\boldsymbol{u}\right\|_1 \leq T^{-1/2}\|\boldsymbol{a}'\|_1\left\|\mathbf{\Theta}\right\|_1\left\|(\boldsymbol{X}^* - \boldsymbol{X})'\boldsymbol{L}\boldsymbol{u}\right\|_1$$

Recall, $\sum_{i\in\{1,\dots\tilde{p}\}}|a_i|\|\mathbf{\Theta}_i\|_1 = O_P(\sqrt{s_i})$, where $\tilde{p}\subseteq\{1,\dots,p\}$. By Lemma C.3, $\|\mathbf{\Theta}_i\|_1 = O(\sqrt{s}_i)$, and

$$T^{-1/2}\left\|(\boldsymbol{X}^* - \boldsymbol{X})'\boldsymbol{L}\boldsymbol{u}\right\|_1 \leq T^{-1/2}\left\|\boldsymbol{L}\boldsymbol{u}\right\|_1.$$

Note that $\boldsymbol{X}$ corresponds to the design matrix when ϕ_j, for all $j = 1,\dots,q$, is known, which coincides with the case of $\boldsymbol{X}^*$, where the design contains the infeasible estimates of ϕ_j. Therefore $D_5 = O(\sqrt{s_iT^{-1}})$, and the result follows. □

Proof of Lemma C.5.

To show the result in (C.11), we use the following decomposition:

$$\begin{aligned}\left|\boldsymbol{a}'\widehat{\boldsymbol{\Theta}}\widehat{\boldsymbol{\Sigma}}_{xu}\widehat{\boldsymbol{\Theta}}'\boldsymbol{a}-\boldsymbol{a}'\boldsymbol{\Theta}\boldsymbol{\Sigma}_{xu}\boldsymbol{\Theta}'\boldsymbol{a}\right| &\leq \left|\boldsymbol{a}'\widehat{\boldsymbol{\Theta}}\widehat{\boldsymbol{\Sigma}}_{xu}\widehat{\boldsymbol{\Theta}}'\boldsymbol{a}-\boldsymbol{a}'\widehat{\boldsymbol{\Theta}}\boldsymbol{\Sigma}_{xu}\widehat{\boldsymbol{\Theta}}'\boldsymbol{a}\right|+\left|\boldsymbol{a}'\widehat{\boldsymbol{\Theta}}\boldsymbol{\Sigma}_{xu}\widehat{\boldsymbol{\Theta}}'\boldsymbol{a}-\boldsymbol{a}'\boldsymbol{\Theta}\boldsymbol{\Sigma}_{xu}\boldsymbol{\Theta}'\boldsymbol{a}\right| \\ &\leq \left|\boldsymbol{a}'\widehat{\boldsymbol{\Theta}}\widehat{\boldsymbol{\Sigma}}_{xu}\widehat{\boldsymbol{\Theta}}'\boldsymbol{a}-\boldsymbol{a}'\widehat{\boldsymbol{\Theta}}\boldsymbol{\Sigma}^*_{xu}\widehat{\boldsymbol{\Theta}}'\boldsymbol{a}\right|+\left|\boldsymbol{a}'\widehat{\boldsymbol{\Theta}}\boldsymbol{\Sigma}^*_{xu}\widehat{\boldsymbol{\Theta}}'\boldsymbol{a}-\boldsymbol{a}'\widehat{\boldsymbol{\Theta}}\boldsymbol{\Sigma}_{xu}\widehat{\boldsymbol{\Theta}}'\boldsymbol{a}\right| \\ &\quad+\left|\boldsymbol{a}'\widehat{\boldsymbol{\Theta}}\boldsymbol{\Sigma}_{xu}\widehat{\boldsymbol{\Theta}}'\boldsymbol{a}-\boldsymbol{a}'\boldsymbol{\Theta}^*\boldsymbol{\Sigma}_{xu}\boldsymbol{\Theta}^{*\prime}\boldsymbol{a}\right| = F_1+F_2+F_3.\end{aligned}$$

It is sufficient to show the following:

$$F_1 = o_P(1), \quad F_2 = o_P(1), \quad F_3 = o_P(1). \tag{D.34}$$

We consider each term, starting with F_1:

$$F_1 = \left|\widehat{\boldsymbol{\Theta}}_i\widehat{\boldsymbol{\Sigma}}_{xu}\widehat{\boldsymbol{\Theta}}_i' - \widehat{\boldsymbol{\Theta}}_i\boldsymbol{\Sigma}^*_{xu}\widehat{\boldsymbol{\Theta}}_i'\right| \leq \left\|\widehat{\boldsymbol{\Sigma}}_{xu}-\boldsymbol{\Sigma}^*_{xu}\right\|_\infty \left\|\widehat{\boldsymbol{\Theta}}'\boldsymbol{a}\right\|_1^2. \tag{D.35}$$

Using the definition $\widehat{u}_t = u_t + \widetilde{\boldsymbol{x}}_t\left(\boldsymbol{\beta}-\widehat{\boldsymbol{\beta}}\right)$, we analyse the term $\left\|\widehat{\boldsymbol{\Sigma}}_{xu}-\boldsymbol{\Sigma}^*_{xu}\right\|_\infty$ as:

$$\begin{aligned}\left\|\widehat{\boldsymbol{\Sigma}}_{xu}-\boldsymbol{\Sigma}^*_{xu}\right\|_\infty &\leq \left\|\frac{1}{T}\sum_{t=1}^T\left(\widetilde{\boldsymbol{x}}_t-\boldsymbol{x}^*_t\right)u_t\right\|_2^2 + \left\|\frac{1}{T}\sum_{t=1}^T\widetilde{\boldsymbol{x}}_t\left(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}\right)\right\|_2^2 \\ &\quad+\left\|\frac{2}{T}\sum_{t=1}^T\widetilde{\boldsymbol{x}}_tu_t\right\|_1\left\|\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}\right\|_1 \\ &\leq \left\|\frac{1}{T}\sum_{t=1}^T\left[\sum_{j=1}^q\left(\widehat{\phi}_j-\phi_j\right)\boldsymbol{x}_{t-j}\right]u_t\right\|_2^2+\left\|\frac{1}{T}\sum_{t=1}^T\widetilde{\boldsymbol{x}}_t\left(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}\right)\right\|_2^2 \\ &\quad+\left\|\frac{2}{T}\sum_{t=1}^T\widetilde{\boldsymbol{x}}_tu_t\right\|_1\left\|\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}\right\|_1 = G_1+G_2+G_3.\end{aligned}$$

Notice that $T^{-1}\sum_{t=1}^T\left[\sum_{j=1}^q\left(\widehat{\phi}_j-\phi_j\right)\boldsymbol{x}_{t-j}\right]u_t$ is bounded in probability by the same arguments used in (I) o and as a consequence $G_1 = o_P(1)$. G_2 attains a non-asymptotic bound by direct application of Theorem 1, hence $G_2 = o_P(1)$. The term G_3 is bounded in probability in Theorem 1, hence $G_3 = o_P(1)$, therefore $\left\|\widehat{\boldsymbol{\Sigma}}_{xu}-\boldsymbol{\Sigma}^*_{xu}\right\|_\infty = o_p(1)$. Then, $\sum_{i\in\{1,\ldots\tilde{p}\}}|a_i|\|\boldsymbol{\Theta}^*_i\|_1 = O_P(\sqrt{s_i})$, by direct application of Lemma C.3, which completes the analysis for F_1. We continue with F_2:

$$F_2 = \left|\boldsymbol{a}'\widehat{\boldsymbol{\Theta}}\boldsymbol{\Sigma}^*_{xu}\widehat{\boldsymbol{\Theta}}'\boldsymbol{a}-\boldsymbol{a}'\widehat{\boldsymbol{\Theta}}\boldsymbol{\Sigma}_{xu}\widehat{\boldsymbol{\Theta}}'\boldsymbol{a}\right| \leq \|\boldsymbol{\Sigma}^*_{xu}-\boldsymbol{\Sigma}_{xu}\|_\infty\|\widehat{\boldsymbol{\Theta}}'\boldsymbol{a}\|_1^2,$$

where $\|\boldsymbol{\Sigma}^*_{xu}-\boldsymbol{\Sigma}_{xu}\|_\infty = \boldsymbol{\Delta}^*_{\boldsymbol{\Sigma}}$,

$$\boldsymbol{\Delta}^*_{\boldsymbol{\Sigma}} \leq \left|\frac{1}{T}\sum_{t=1}^T\left(\boldsymbol{x}^*_tu_t\right)'\left(\boldsymbol{x}^*_tu_t\right)-E\left[\frac{1}{T}\sum_{t=1}^T\left(\widetilde{\boldsymbol{x}}_tu_t\right)'\left(\widetilde{\boldsymbol{x}}_tu_t\right)\right]\right|$$

$$
\begin{aligned}
&= \left| \frac{1}{T} \sum_{t=1}^{T} \left(\left(\boldsymbol{x}_t - \sum_{j=1}^{q} \phi_j \boldsymbol{x}_{t-j} \right) \left(u_t - \sum_{j=1}^{q} \phi_j u_{t-j} \right) \right)' \right. \\
&\qquad \left(\left(\boldsymbol{x}_t - \sum_{j=1}^{q} \phi_j \boldsymbol{x}_{t-j} \right) \left(u_t - \sum_{j=1}^{q} \phi_j u_{t-j} \right) \right) \\
&\quad - E \left[\sum_{t=1}^{T} \left(\left(\boldsymbol{x}_t - \sum_{j=1}^{q} \phi_j \boldsymbol{x}_{t-j} \right) \left(u_t - \sum_{j=1}^{q} \phi_j u_{t-j} \right) \right)' \right. \\
&\qquad \left. \left. \left(\left(\boldsymbol{x}_t - \sum_{j=1}^{q} \phi_j \boldsymbol{x}_{t-j} \right) \left(u_t - \sum_{j=1}^{q} \phi_j u_{t-j} \right) \right) \right] \right| .
\end{aligned}
$$

Set $\boldsymbol{\xi}_t = \left(\boldsymbol{x}_t - \sum_{j=1}^{q} \phi_j \boldsymbol{x}_{t-j} \right) \left(u_t - \sum_{j=1}^{q} \phi_j u_{t-j} \right)$, and its expectation

$$
E\left(\boldsymbol{\xi}_t\right) = E\left[\left(\boldsymbol{x}_t - \sum_{j=1}^{q} \phi_j \boldsymbol{x}_{t-j} \right) \left(u_t - \sum_{j=1}^{q} \phi_j u_{t-j} \right) \right].
$$

By Theorem 14.1 of Davidson (1994), and as a product of α-mixing series, $\{\boldsymbol{\xi}_t\}$, is also a p-dimensional α-mixing series with similar properties as the ones outlined in Assumptions 1 and 3. We then write

$$
\begin{aligned}
\boldsymbol{\Delta}_{\boldsymbol{\Sigma}}^* &\leq \left| \frac{1}{T} \sum_{t=1}^{T} \boldsymbol{\xi}_t' \boldsymbol{\xi}_t - \sum_{t=1}^{T} E\left[\frac{1}{T} \sum_{t=1}^{T} \boldsymbol{\xi}_t' \right] E\left[\frac{1}{T} \sum_{t=1}^{T} \boldsymbol{\xi}_t \right] \right| \\
&\leq \frac{1}{T} \left\| \sum_{t=1}^{T} \boldsymbol{\xi}_t - E\left(\boldsymbol{\xi}_t\right) \right\|_2^2 + \frac{2}{T} \left\| E\left(\boldsymbol{\xi}_t\right) \right\|_2^2 \leq \frac{1}{T} \left\| \sum_{t=1}^{T} \boldsymbol{\xi}_t - E\left(\boldsymbol{\xi}_t\right) \right\|_2^2 + \frac{2}{T} E \left\| \boldsymbol{\xi}_t \right\|_2^2 = J_1 + J_2.
\end{aligned} \tag{D.36}
$$

Note that the term J_1 includes an α-mixing series and can be bounded following similar arguments as in Lemma A1 of Dendramis et al. (2021), J_2 is bounded by Assumptions 1 and 3, which implies that second moments exist and are bounded. Therefore, $\left\| \boldsymbol{\Sigma}_{xu}^* - \boldsymbol{\Sigma}_{xu} \right\|_\infty = o_P(1)$, while $\sum_{i \in \{1,\ldots \tilde{p}\}} |a_i| \|\boldsymbol{\Theta}_i^*\|_1 = O_P(\sqrt{s_i})$ by direct application of Lemma C.3, which completes the analysis for F_2. Analyse F_3:

$$
\left| \boldsymbol{a}' \widehat{\boldsymbol{\Theta}} \boldsymbol{\Sigma}_{xu} \widehat{\boldsymbol{\Theta}}' \boldsymbol{a} - \boldsymbol{a}' \boldsymbol{\Theta}^* \boldsymbol{\Sigma}_{xu} \boldsymbol{\Theta}^* \boldsymbol{a} \right| \leq \left\| \boldsymbol{\Sigma}_{xu} \right\|_\infty \left\| (\widehat{\boldsymbol{\Theta}} - \boldsymbol{\Theta}^*)' \boldsymbol{a} \right\|_1^2 + 2 \left\| \boldsymbol{a}' \boldsymbol{\Sigma}_{xu} \right\|_2 \left\| (\widehat{\boldsymbol{\Theta}} - \boldsymbol{\Theta}^*)' \boldsymbol{a} \right\|_2,
$$

where, by (C.9)

$$
\left\| \boldsymbol{\Theta}^{*\prime} \boldsymbol{a} \right\|_2 \leq \left\| \boldsymbol{\Theta}^{*\prime} \boldsymbol{a} \right\|_1 \leq \sum_{i \in \{1,\ldots \tilde{p}\}} |a_i| \|\boldsymbol{\Theta}_i^{*\prime}\|_1 = O_P\left(\sqrt{s_i}\right),
$$

and by Lemma C.3 we have that,

$$
\left\| \widehat{\boldsymbol{\Theta}}_i - \boldsymbol{\Theta}_i^* \right\|_1 = O_P\left(s_i \lambda_i\right), \text{ and } \left\| \widehat{\boldsymbol{\Theta}}_i - \boldsymbol{\Theta}_i^* \right\|_2 = O_P\left(\sqrt{s_i} \lambda_i\right).
$$

Further $\|\mathbf{\Theta}^{*\prime}\boldsymbol{a} - E\mathbf{\Theta}'\boldsymbol{a}\|_2 = o_P(1)$, by implication of Lemma C.3 Note that $\mathbf{\Sigma}_{xu}$ and $\mathbf{\Theta}$ satisfy the following properties:

$$0 < \Lambda_{\min}(\mathbf{\Sigma}) \leq \Lambda_{\max}(\mathbf{\Sigma}), \quad 0 < \Lambda_{\min}(\mathbf{\Theta}) \leq \Lambda_{\max}(\mathbf{\Theta}),$$

where $\Lambda_{\min}(\mathbf{\Sigma}), \Lambda_{\max}(\mathbf{\Sigma})$ are the smallest and largest eigenvalues of $\mathbf{\Sigma}$ respectively and $\Lambda_{\min}(\mathbf{\Theta}), \Lambda_{\max}(\mathbf{\Theta})$ are the smallest and largest eigenvalues of $\mathbf{\Theta}$ respectively. Denote $E[u_t^2] := \mathrm{Var}(u_t) = \sigma_u^2$ which is a scalar and $0 < \sigma_u^2 < \infty$. Then $\mathbf{\Sigma}_{xu} := \mathbf{\Sigma}\sigma_u^2$. Recall that by (B.8)

$$\boldsymbol{a}'\mathbf{\Theta}^*\mathbf{\Sigma}_{xu}^*\mathbf{\Theta}^*\boldsymbol{a} \geq \Lambda_{\min}(\mathbf{\Sigma}_{xu}^*)\|\mathbf{\Theta}^{*\prime}\boldsymbol{a}\|_2^2 \geq \Lambda_{\min}(\mathbf{\Sigma}_{xu}^*)\lambda_{\min}^2(\mathbf{\Theta}^*)\|\boldsymbol{a}\|_2^2 \geq \Lambda_{\min}(\mathbf{\Sigma}_{xu}^*)\lambda_{\min}^{-2}(\mathbf{\Theta}^*),$$

therefore the following statements hold:

$$\|\mathbf{\Sigma}_{xu}\|_\infty \leq \|\mathbf{\Sigma}_{xu}\|_2 = \Lambda_{\max}(\mathbf{\Sigma}_{xu}) = \sigma_u^2\Lambda_{\max}(\mathbf{\Sigma}) = O_P(1), \tag{D.37}$$

$$\begin{aligned} \left\|\mathbf{\Sigma}_{xu}\mathbf{\Theta}^{*\prime}\boldsymbol{a}\right\|_\infty &\leq \|\mathbf{\Sigma}_{xu}\|_2 \left\|\mathbf{\Theta}^{*\prime}\boldsymbol{a}\right\|_\infty \leq O_P(1)\|\mathbf{\Theta}^{*\prime}\|_2\|\boldsymbol{a}\|_2 \\ &= O_P(1)\Lambda_{\max}(\mathbf{\Theta}^*) = \frac{O_P(1)}{\Lambda_{\min}(\mathbf{\Sigma})} = O_P(1), \end{aligned} \tag{D.38}$$

completing the proof. □

Proof of Lemma C.6

Notice that, for some $i, k \in [1, p]$

$$\begin{aligned} \left|(\widetilde{\boldsymbol{X}}'\widetilde{\boldsymbol{X}})_{i,k} - E\left(\widetilde{\boldsymbol{X}}'\widetilde{\boldsymbol{X}}\right)_{i,k}\right| \leq & \left|(\boldsymbol{X}'\widehat{\boldsymbol{L}}'\widehat{\boldsymbol{L}}\boldsymbol{X})_{i,k} - (\boldsymbol{X}'\boldsymbol{L}'\boldsymbol{L}\boldsymbol{X})_{i,k}\right| \\ & + \left|(\boldsymbol{X}'\boldsymbol{L}'\boldsymbol{L}\boldsymbol{X})_{i,k} - E\left(\boldsymbol{X}'L'L\boldsymbol{X}\right)_{i,k}\right| \\ & + \left|E\left(\boldsymbol{X}'\boldsymbol{L}'\boldsymbol{L}\boldsymbol{X}\right)_{i,k} - E\left(\boldsymbol{X}'\widehat{\boldsymbol{L}}'\widehat{\boldsymbol{L}}\boldsymbol{X}\right)_{i,k}\right| = A + B + C. \end{aligned}$$

By (D.16), $A = O_P(T^{-1/2})$, we analyse B:

$$\begin{aligned} B \leq & \left|\boldsymbol{x}_i'\widehat{\boldsymbol{L}} - E(\boldsymbol{x_i})'\widehat{\boldsymbol{L}}\right| \left|\boldsymbol{x}_k'\widehat{\boldsymbol{L}} - E(\boldsymbol{x}_k)'\widehat{\boldsymbol{L}}\right| + |E(\boldsymbol{x}_i'\boldsymbol{L})| \left|\boldsymbol{x}_k'\widehat{\boldsymbol{L}} - E(\boldsymbol{x_k})'\boldsymbol{L}\right| \\ & + |E(\boldsymbol{x}_k'\boldsymbol{L})| \left|\boldsymbol{x}_i'\widehat{\boldsymbol{L}} - E(\boldsymbol{x_i})'\boldsymbol{L}\right| \\ = & \left|\left(\boldsymbol{x}_i'(\widehat{\boldsymbol{L}} - \boldsymbol{L}) - E\boldsymbol{x}_i\right) + \boldsymbol{x}_i'\boldsymbol{L} - E\boldsymbol{x}_i'\widehat{\boldsymbol{L}}\right| \times \left|\left(\boldsymbol{x}_k'(\widehat{\boldsymbol{L}} - \boldsymbol{L}) - E\boldsymbol{x}_k\right) + \boldsymbol{x}_k'\boldsymbol{L} - E\boldsymbol{x}_k'\widehat{\boldsymbol{L}}\right| \\ & + |E(\boldsymbol{x}_i'\boldsymbol{L})| \left|(\boldsymbol{x_k} - E\boldsymbol{x}_k)'(\widehat{\boldsymbol{L}} - \boldsymbol{L}) + \boldsymbol{x}_k'\boldsymbol{L} - E\boldsymbol{x}_k'\widehat{\boldsymbol{L}}\right| \\ & + |E(\boldsymbol{x}_k'\boldsymbol{L})| \left|(\boldsymbol{x_i} - E\boldsymbol{x}_i)'(\widehat{\boldsymbol{L}} - \boldsymbol{L}) + \boldsymbol{x}_i'\boldsymbol{L} - E\boldsymbol{x}_i'\widehat{\boldsymbol{L}}\right|, \end{aligned}$$

where $\boldsymbol{L}$ and $\widehat{\boldsymbol{L}}$ are defined similarly to (N.1). Under Assumption 3, $\boldsymbol{x}_{i,k}$ is a thin tailed α-mixing sequence with properties in Assumption 3 of the main paper, with $\max_{i,k} E|\boldsymbol{x}_{i,k}| \leq d < \infty$, for some $d > 0$, we have

$$|\widehat{\sigma}_{i,k} - \sigma_{ik}| \leq \left|(\boldsymbol{x}_i - E\boldsymbol{x}_i)'(\widehat{\boldsymbol{L}} - \boldsymbol{L}) + \boldsymbol{x}_i'\boldsymbol{L} - E\boldsymbol{x}_i'\widehat{\boldsymbol{L}}\right|$$

$$\times \left|(\boldsymbol{x}_k - E\boldsymbol{x}_k)'(\widehat{\boldsymbol{L}} - \boldsymbol{L}) + \boldsymbol{x}_k\boldsymbol{L} - \widehat{\boldsymbol{L}}E\boldsymbol{x}_k\right|$$
$$+ d\left|(\boldsymbol{x_k} - E\boldsymbol{x}_k)'(\widehat{\boldsymbol{L}} - \boldsymbol{L}) + \boldsymbol{x}_k'\boldsymbol{L} - E\boldsymbol{x}_k'\widehat{\boldsymbol{L}}\right|$$
$$+ d\left|(\boldsymbol{x_i} - E\boldsymbol{x}_i)'(\widehat{\boldsymbol{L}} - \boldsymbol{L}) + \boldsymbol{x}_i'\boldsymbol{L} - E\boldsymbol{x}_i'\widehat{\boldsymbol{L}}\right|.$$

By Assumption 3, $\{\boldsymbol{x}_t\}$ is a p-dimensional ergodic α-mixing sequences, and consequently by Theorem 14.1 of Davidson (1994), $\boldsymbol{x}_i'\boldsymbol{L}, \boldsymbol{x}_k'\boldsymbol{L}$ are α-mixing series satisfying Assumption 3. Therefore,

$$P\left(\max_{0<i,k\leq p} |\widehat{\sigma}_{i,k} - \sigma_{ik}| > tT\right) \leq \sum_{i,k} P\left(\left|(\boldsymbol{x}_i - E\boldsymbol{x}_i)'(\widehat{\boldsymbol{L}} - \boldsymbol{L}) + \boldsymbol{x}_i'\boldsymbol{L} - E\boldsymbol{x}_i'\widehat{\boldsymbol{L}}\right|\right.$$
$$\left.\times \left|(\boldsymbol{x}_k - E\boldsymbol{x}_k)'(\widehat{\boldsymbol{L}} - \boldsymbol{L}) + \boldsymbol{x}_k\boldsymbol{L} - \widehat{\boldsymbol{L}}E\boldsymbol{x}_k\right| > tT/3\right)$$
$$+ P\left(d\left|(\boldsymbol{x_k} - E\boldsymbol{x}_k)'(\widehat{\boldsymbol{L}} - \boldsymbol{L}) + \boldsymbol{x}_k'\boldsymbol{L} - E\boldsymbol{x}_k'\widehat{\boldsymbol{L}}\right| > tT/3\right)$$
$$+ P\left(d\left|(\boldsymbol{x_i} - E\boldsymbol{x}_i)'(\widehat{\boldsymbol{L}} - \boldsymbol{L}) + \boldsymbol{x}_i'\boldsymbol{L} - E\boldsymbol{x}_i'\widehat{\boldsymbol{L}}\right| > tT/3\right).$$

Notice that

$$\left|(\boldsymbol{x}_i - E\boldsymbol{x}_i)'\left(\widehat{\boldsymbol{L}} - \boldsymbol{L}\right) + \boldsymbol{x}_i'\boldsymbol{L} - E\boldsymbol{x}_i'\widehat{\boldsymbol{L}}\right| \times \left|(\boldsymbol{x}_k - E\boldsymbol{x}_k)'\left(\widehat{\boldsymbol{L}} - \boldsymbol{L}\right) + \boldsymbol{x}_k\boldsymbol{L} - \widehat{\boldsymbol{L}}E\boldsymbol{x}_k\right|$$
$$\leq \left|(\boldsymbol{x}_i - E\boldsymbol{x}_i)'\left(\widehat{\boldsymbol{L}} - \boldsymbol{L}\right)\right| \times \left|(\boldsymbol{x}_k - E\boldsymbol{x}_k)'\left(\widehat{\boldsymbol{L}} - \boldsymbol{L}\right)\right|$$
$$= |(\boldsymbol{x}_i - E\boldsymbol{x}_i)| \times \left|(\boldsymbol{x}_k - E\boldsymbol{x}_k)'\right| \|\widehat{\boldsymbol{L}} - \boldsymbol{L}\|_2^2,$$
$$d\left|(\boldsymbol{x_k} - E\boldsymbol{x}_k)'\left(\widehat{\boldsymbol{L}} - \boldsymbol{L}\right) + \boldsymbol{x}_k'\boldsymbol{L} - E\boldsymbol{x}_k'\widehat{\boldsymbol{L}}\right| \leq d\left|(\boldsymbol{x_k} - E\boldsymbol{x}_k)'\left(\widehat{\boldsymbol{L}} - \boldsymbol{L}\right)\right|$$

and

$$d\left|(\boldsymbol{x_i} - E\boldsymbol{x}_i)'\left(\widehat{\boldsymbol{L}} - \boldsymbol{L}\right) + \boldsymbol{x}_i'\boldsymbol{L} - E\boldsymbol{x}_i'\widehat{\boldsymbol{L}}\right| \leq d\left|(\boldsymbol{x_i} - E\boldsymbol{x}_i)'\left(\widehat{\boldsymbol{L}} - \boldsymbol{L}\right)\right|.$$

Then we can write,

$$P\left(\max_{0<i,k\leq p} T^{-1}|\widehat{\sigma}_{i,k} - \sigma_{i,k}| > \nu\right) \leq \sum_{i,k} P\left(T^{-1}|(\boldsymbol{x}_i - E\boldsymbol{x}_i)| \times \left|(\boldsymbol{x}_k - E\boldsymbol{x}_k)'\right| \left\|\widehat{\boldsymbol{L}} - \boldsymbol{L}\right\|^2 > \frac{\nu}{3}\right)$$
$$+ \sum_{i,k} P\left(dT^{-1}\left|(\boldsymbol{x_k} - E\boldsymbol{x}_k)'\left(\widehat{\boldsymbol{L}} - \boldsymbol{L}\right)\right| > \frac{\nu}{3}\right)$$
$$+ \sum_{i,k} P\left(dT^{-1}\left|(\boldsymbol{x_i} - E\boldsymbol{x}_i)'\left(\widehat{\boldsymbol{L}} - \boldsymbol{L}\right)\right| > \frac{\nu}{3}\right). \tag{D.39}$$

Since $(\boldsymbol{x}_i - E\boldsymbol{x}_i)$ is α-mixing with properties outlined in Assumption 3, and $\|\widehat{\boldsymbol{L}} - \boldsymbol{L}\|_1 = O_P(s_0^2\lambda)$ by direct application of Corollary 1 and the order of λ defined in (18). The three terms are bounded using Lemma A1 of the Online Supplement of Dendramis et al. (2021). To show that (D.39) holds, it suffices to show that

$$\sum_{i,k} P\left(T^{-1}|(\boldsymbol{x}_i - E\boldsymbol{x}_i)| \times |(\boldsymbol{x}_k - E\boldsymbol{x}_k)| \left\|\widehat{\boldsymbol{L}} - \boldsymbol{L}\right\|^2 > \frac{\nu}{3}\right) \to 0 \tag{D.40}$$

$$\sum_k P\left(dT^{-1}\left|(\boldsymbol{x_k} - E\boldsymbol{x}_k)'(\widehat{\boldsymbol{L}} - \boldsymbol{L})\right| > \frac{\nu}{3}\right) \to 0 \tag{D.41}$$

$$\sum_i P\left(dT^{-1}\left|(\boldsymbol{x_i} - E\boldsymbol{x}_i)'(\widehat{\boldsymbol{L}} - \boldsymbol{L})\right| > \frac{\nu}{3}\right) \to 0, \tag{D.42}$$

for some finite constants $c_1, \delta_1, \delta_2, d_1, d_2, e_1, e_2 > 0$ independent of T, p, where $c_1, \delta_1, \delta_2, d_1, d_2$ are large constants. By Lemma A11 , Equation B.61 of Chudik et al. (2018) we have that

$$\begin{aligned}
&\sum_{i,k} P\left(T^{-1}\left|(\boldsymbol{x}_i - E\boldsymbol{x}_i)\right| \times \left|(\boldsymbol{x}_k - E\boldsymbol{x}_k)\right| \times \left\|\widehat{\boldsymbol{L}} - \boldsymbol{L}\right\|^2 > \frac{\nu}{3}\right) \\
&\leq \sum_i P\left(T^{-1}\left|(\boldsymbol{x}_i - E\boldsymbol{x}_i)\right| > \nu/(3\delta_1\delta_2)\right) + \sum_k P\left(\left|(\boldsymbol{x}_k - E\boldsymbol{x}_k)\right| > \delta_1\right) \\
&\quad + P\left(\left\|\widehat{\boldsymbol{L}} - \boldsymbol{L}\right\|^2 > \delta_2\right) = S_1 + S_2 + S_3.
\end{aligned} \tag{D.43}$$

We analyse S_1:

Case 1: $\boldsymbol{x}_i \in \mathcal{T}(r), r > 0$

$$\begin{aligned}
\sum_i P\left(T^{-1/2}\left|(\boldsymbol{x}_i - E\boldsymbol{x}_i)\right| > \frac{\nu T^{1/2}}{3\delta_1\delta_2}\right) &\leq pc_0\left\{\exp\left(-c_1\left(\frac{\nu T^{1/2}}{3\delta_1\delta_2}\right)^2\right) + \exp\left(-c_2\left(\frac{\nu T^{1/2}}{3\delta_1\delta_2\log^2 T}\right)^{\gamma_1}\right)\right\} \\
&\leq pc_0\exp\left(-c_1\left(\frac{\nu T^{1/2}}{3\delta_1\delta_2}\right)^2\right) \to 0,\ \text{ for } \nu = \sqrt{\log p/T}
\end{aligned} \tag{D.44}$$

Case 2: $\boldsymbol{x}_i \in \mathcal{H}(\theta), \theta > 4$

$$\begin{aligned}
\sum_i P\left(T^{-1/2}\left|(\boldsymbol{x}_i - E\boldsymbol{x}_i)\right| > \frac{\nu T^{1/2}}{3\delta_1\delta_2}\right) &\leq pc_0\left\{\exp\left(-c_1\left(\frac{\nu T^{1/2}}{3\delta_1\delta_2}\right)^2\right) + \left(\frac{\nu T^{1/2}}{3\delta_1\delta_2}\right)^{-\theta} T^{-\frac{\theta}{2}-1}\right\} \\
&\leq c_0 p\left(\frac{\nu T^{1/2}}{3\delta_1\delta_2}\right)^{-\theta} T^{-\frac{\theta}{2}-1} \to 0, \text{for } \nu = p^{\frac{1}{\theta}}T^{-(\frac{\theta}{2}+1)}
\end{aligned} \tag{D.45}$$

For S_2 in case 1 ($\boldsymbol{x}_i \in \mathcal{T}(r), r > 0$), let $\delta_1 \geq \nu/3$, then $S_2 \leq pc_* \exp\left(-c_2 9\log pT^{-2}\right)$, for some $c_* > 0$. For S_2 in case 2 ($\boldsymbol{x}_i \in \mathcal{H}(\theta), \theta > 4$) we have that $S_2 \leq cp\left(\frac{T^{1/2}}{\delta_1}\right)^{-\theta} T^{-\frac{\theta}{2}-1}$ for some $c, \delta_1 > 0$. For S_3, notice that $\widehat{\boldsymbol{L}}$, and $\boldsymbol{L}$ are non-singular matrices defined similarly to (N.1), then by implication of Corollary 1 we have that

$$S_3 \leq P\left(\max_{s=1,\dots,T}\left|\widehat{\boldsymbol{l}}_s - \boldsymbol{l}_s\right|^2 > \delta_2\right) = o\left(1\right),\ \delta > 0.$$

(D.41)–(D.42) can be bounded using similar arguments to (D.43). Finally, by Assumption 3 and following a similar analysis with term B, by implication of Corollary 1 term $C = o(1)$, for some $\delta > 0$. The result follows. □

E Simulation Study Supplement

E.1 Main designs, $|S_0| = 7$

In the main paper we consider the experimental design in (29), studying the small sample properties of our methodology compared to methods previously seen in the literature, e.g. LASSO, DEBIASED LASSO where the cardinality of the non-sparse set is $s_0 = |S_0| = 3$. In this section we examine the case where the cardinality of the active set is increased to 7 non-sparse elements, i.e. $s_0 = 7$. Notice that the selection of the regularisation parameter corresponding to this simulation study, follows Section 5 of the main paper, with $k = 10$.

The first panel of Table E.1 reports the ratio of the average root mean squared error (RMSE) of the LASSO estimator over the RMSE of the GLS LASSO. The second panel of Table E.1, reports the ratio of the average RMSE of the DEBIASED LASSO estimator over the RMSE of DEBIASED GLS LASSO. Entries larger than 1 indicate superiority of the competing model (GLS LASSO). In the parenthesis reported are the entries corresponding to the RMSE of the GLS LASSOin Panel I and the DEBIASED GLS LASSO in Panel II. Table E.2 below reports and size-adjusted power and standard errors of the DEBIASED estimators. The results in Table E.1 suggest similar patterns with Table 1 of the main paper.

As the results in Tables 2 and 1 of the main paper suggest, when autocorrelation is present, i.e. $\phi > 0$ our method outperforms the DEBIASED LASSO, while in the cases of $\phi = 0$, i.e. $u_t \sim$ *i.i.d.* both the DEBIASED LASSO and DEBIASED GLS LASSO report similar results.

In Table E.2 we report size-adjusted power (refer to this term in Section 6 of the main paper) for both the DEBIASED models under the same sparsity level, $s_0 = 7$. The results here, suggest that our method reports power closer to 95% as p, T and ϕ increase, while $|\phi| < 1$. These results are in line with the results of Table 2 of the main paper. It is evident that the size-adjustment uncovers the underlying behaviour of the methods, given that DEBIASED LASSO exhibits significant size distortions as p and ϕ increase as a pair and/or individually.

These findings, are in line with our theoretical results, while the behaviour of the DEBIASED estimators in higher degrees of sparsity are in line with the empirical results of Van de Geer et al. (2014), suggesting that inference with the DEBIASED LASSO has its limit when the problem is not sufficiently sparse.

In Table E.3 we evaluate the performance measures reported in (31) across 1000 replications of model (29) of the main paper. Our findings are in line with the finding os Table 3 of the main paper.

Table E.1: See description of Table 1, for $|S_0| = 7$.

	p/T			200					500					1000		
ϕ		0	0.5	0.8	0.9	0.95	0	0.5	0.8	0.9	0.95	0	0.5	0.8	0.9	0.95
LASSO/GLS LASSO	100	1.00	1.23	1.76	2.38	2.98	1.00	1.30	2.18	2.73	3.66	1.00	1.25	2.21	3.30	4.08
GLS LASSO		(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.01)	(0.01)	(0.01)
LASSO/GLS LASSO	200	1.00	1.20	1.76	2.36	2.96	1.00	1.31	1.99	2.45	3.48	1.00	1.31	2.43	3.20	3.97
GLS LASSO		(0.04)	(0.03)	(0.03)	(0.03)	(0.03)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
LASSO/GLS LASSO	500	1.00	1.19	1.83	2.42	2.97	1.00	1.30	1.70	2.38	3.31	1.00	1.34	2.38	2.51	3.83
GLS LASSO		(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Panel II																
Debiased LASSO/DEBIASED GLS	100	1.00	1.26	1.99	2.67	3.43	1.00	1.29	2.10	2.93	3.92	1.00	1.29	2.11	3.00	4.17
DEBIASED GLS		(0.07)	(0.06)	(0.06)	(0.06)	(0.06)	(0.04)	(0.04)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.02)	(0.02)	(0.02)
Debiased LASSO/DEBIASED GLS	200	1.00	1.27	1.99	2.68	3.46	1.00	1.29	2.13	2.95	3.91	1.00	1.29	2.12	3.02	4.16
DEBIASED GLS		(0.06)	(0.06)	(0.05)	(0.05)	(0.05)	(0.04)	(0.04)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.02)	(0.02)	(0.02)
Debiased LASSO/DEBIASED GLS	500	1.01	1.26	1.94	2.62	3.38	1.00	1.30	2.22	2.95	3.86	1.00	1.30	2.14	3.11	4.13
DEBIASED GLS		(0.06)	(0.05)	(0.05)	(0.05)	(0.05)	(0.04)	(0.04)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.02)	(0.02)	(0.02)

Table E.2: See description of Table 2, for $|S_0| = 7$

Size adjusted power and standard errors																
	p/T			200					500					1000		
ϕ		0	0.5	0.8	0.9	0.95	0	0.5	0.8	0.9	0.95	0	0.5	0.8	0.9	0.95
GLS LASSO	100	0.75	0.76	0.76	0.76	0.75	0.74	0.74	0.75	0.75	0.75	0.75	0.75	0.75	0.75	0.74
(s.e.)		(0.07)	(0.07)	(0.06)	(0.06)	(0.07)	(0.05)	(0.04)	(0.04)	(0.04)	(0.04)	(0.03)	(0.03)	(0.03)	(0.02)	(0.03)
LASSO		0.75	0.73	0.65	0.58	0.50	0.74	0.73	0.70	0.68	0.64	0.75	0.74	0.72	0.69	0.65
(s.e.)		(0.07)	(0.08)	(0.12)	(0.15)	(0.20)	(0.05)	(0.05)	(0.07)	(0.10)	(0.14)	(0.03)	(0.04)	(0.05)	(0.07)	(0.10)
GLS LASSO	200	0.83	0.84	0.85	0.85	0.84	0.89	0.90	0.92	0.92	0.92	0.93	0.94	0.95	0.95	0.95
(s.e.)		(0.07)	(0.07)	(0.06)	(0.06)	(0.07)	(0.05)	(0.04)	(0.04)	(0.04)	(0.04)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
LASSO		0.83	0.81	0.74	0.65	0.55	0.89	0.88	0.83	0.77	0.69	0.93	0.92	0.88	0.84	0.78
(s.e.)		(0.07)	(0.08)	(0.12)	(0.16)	(0.20)	(0.05)	(0.05)	(0.07)	(0.10)	(0.14)	(0.03)	(0.04)	(0.05)	(0.07)	(0.10)
GLS LASSO	500	0.84	0.85	0.86	0.86	0.86	0.90	0.91	0.92	0.92	0.92	0.93	0.94	0.95	0.95	0.95
(s.e.)		(0.08)	(0.07)	(0.07)	(0.07)	(0.07)	(0.05)	(0.04)	(0.04)	(0.04)	(0.04)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
LASSO		0.84	0.82	0.74	0.66	0.59	0.90	0.88	0.84	0.79	0.71	0.93	0.92	0.88	0.85	0.79
(s.e.)		(0.08)	(0.09)	(0.12)	(0.16)	(0.20)	(0.05)	(0.05)	(0.08)	(0.10)	(0.14)	(0.03)	(0.04)	(0.05)	(0.07)	(0.10)

Table E.3: See description of Table 3, for $|S_0| = 7$.

		p/T			200					500					1000		
ϕ			0	0.5	0.8	0.9	0.95	0	0.5	0.8	0.9	0.95	0	0.5	0.8	0.9	0.95
DEBIASED LASSO	AvgCov S_0	100	0.916	0.919	0.930	0.935	0.939	0.927	0.928	0.931	0.938	0.944	0.931	0.935	0.946	0.947	0.949
	AvgCov S_0^c		0.957	0.956	0.959	0.956	0.955	0.954	0.954	0.952	0.954	0.954	0.950	0.951	0.952	0.950	0.951
	AvgLength		0.285	0.326	0.458	0.606	0.786	0.177	0.204	0.291	0.395	0.536	0.125	0.144	0.206	0.281	0.386
	AvgLength^c		0.285	0.326	0.458	0.606	0.785	0.177	0.204	0.291	0.395	0.536	0.125	0.144	0.206	0.281	0.386
DEBIASED GLS	AvgCov S_0		0.914	0.901	0.896	0.911	0.920	0.926	0.918	0.916	0.923	0.932	0.931	0.925	0.917	0.923	0.929
	AvgCov S_0^c		0.956	0.957	0.961	0.967	0.972	0.953	0.954	0.958	0.962	0.967	0.950	0.952	0.954	0.960	0.964
	AvgLength		0.284	0.261	0.238	0.244	0.256	0.177	0.159	0.143	0.141	0.147	0.125	0.112	0.098	0.097	0.098
	AvgLength^c		0.283	0.261	0.238	0.244	0.255	0.177	0.159	0.143	0.141	0.147	0.125	0.112	0.098	0.097	0.098
DEBIASED LASSO	AvgCov S_0	200	0.903	0.914	0.926	0.928	0.933	0.927	0.929	0.936	0.942	0.945	0.929	0.931	0.939	0.946	0.949
	AvgCov S_0^c		0.968	0.967	0.967	0.966	0.962	0.959	0.958	0.956	0.958	0.957	0.955	0.956	0.954	0.953	0.955
	AvgLength		0.291	0.332	0.466	0.614	0.793	0.179	0.205	0.293	0.397	0.538	0.125	0.144	0.207	0.282	0.387
	AvgLength^c		0.291	0.332	0.465	0.614	0.793	0.178	0.205	0.293	0.397	0.537	0.125	0.144	0.207	0.282	0.387
DEBIASED GLS	AvgCov S_0		0.898	0.886	0.883	0.899	0.912	0.926	0.912	0.908	0.914	0.927	0.928	0.926	0.924	0.926	0.932
	AvgCov S_0^c		0.967	0.968	0.973	0.977	0.980	0.958	0.961	0.964	0.967	0.974	0.955	0.956	0.960	0.964	0.968
	AvgLength		0.289	0.268	0.247	0.255	0.268	0.178	0.161	0.146	0.144	0.153	0.125	0.112	0.100	0.099	0.100
	AvgLength^c		0.289	0.268	0.247	0.255	0.268	0.178	0.161	0.146	0.144	0.153	0.125	0.112	0.100	0.099	0.100
DEBIASED LASSO	AvgCov S_0	500	0.889	0.902	0.913	0.921	0.930	0.917	0.916	0.937	0.939	0.945	0.927	0.934	0.944	0.949	0.947
	AvgCov S_0^c		0.982	0.981	0.981	0.977	0.973	0.969	0.968	0.966	0.967	0.965	0.960	0.961	0.958	0.958	0.959
	AvgLength		0.299	0.341	0.476	0.624	0.802	0.180	0.207	0.295	0.398	0.536	0.126	0.145	0.208	0.284	0.389
	AvgLength^c		0.299	0.341	0.476	0.624	0.802	0.180	0.207	0.295	0.398	0.536	0.126	0.145	0.208	0.284	0.389
DEBIASED GLS	AvgCov S_0		0.882	0.872	0.878	0.892	0.901	0.915	0.906	0.891	0.909	0.929	0.927	0.920	0.919	0.917	0.935
	AvgCov S_0^c		0.982	0.982	0.985	0.987	0.989	0.969	0.972	0.973	0.978	0.981	0.960	0.963	0.968	0.969	0.975
	AvgLength		0.297	0.276	0.264	0.271	0.284	0.180	0.165	0.147	0.151	0.161	0.126	0.113	0.103	0.099	0.104
	AvgLength^c		0.297	0.276	0.263	0.271	0.285	0.180	0.165	0.147	0.151	0.161	0.126	0.113	0.103	0.099	0.105

E.2 Simulations with $MA(1)$ errors

In this section we examine the small sample properties of the estimator considering the case of misspecifying the error. We consider the following data generating process:

$$y_t = \boldsymbol{x}_t'\boldsymbol{\beta} + u_t, \quad u_t = \varepsilon_t + \theta\varepsilon_{t-1}, \tag{E.1}$$

where ε_t, $\boldsymbol{x}_t \sim$ i.i.d. By using the lag operator, $\theta L\varepsilon_t = \varepsilon_{t-1}$, we write

$$u_t = (1+\theta L)\varepsilon_t \Leftrightarrow \frac{u_t}{(1+\theta L)} = \varepsilon_t, \text{ and } |\theta| < 1$$

then by applying infinite geometric series we have that $MA(1) \equiv AR(\infty)$

$$u_t = \sum_{j=0}^{\infty} \theta^j L\varepsilon_t = \sum_{j=0}^{\infty} \theta^j u_{t-j}, \tag{E.2}$$

which is indicative of the invertibility between AR and MA models. The latter covers both the cases of mispecifying an AR(q) model, and the case where q can diverge to infinity. While our theoretical results imply both cases, we showcase the performance of GLS LASSO and DEBIASED GLS through simulation experiments. The data generating process followed is (E.1), with $s_0 \in \{3, 7\}$, similarly to the experimental design outlined in Section E of the main paper. Notice that the selection of the regularisation parameter corresponding to this simulation study, follows Section 5 of the main paper, with $k = 10$.

We report our findings in Tables E.4 – E.9. The first panel of Table E.4 – E.5 reports the ratio of the average root mean squared error (RMSE) of the LASSO estimator over the RMSE of the GLS LASSO. The second panel of Tables E.4 – E.5, report the ratio of the average RMSE of the DEBIASED LASSO estimator over the RMSE of DEBIASED GLS LASSO. Entries larger than 1 indicate superiority of the competing model (GLS LASSO). In the parenthesis reported are the entries corresponding to the RMSE of the GLS LASSO in Panel I and the DEBIASED GLS LASSO in Panel II. Tables E.6 – E.7 below report and size-adjusted power and standard errors of the DEBIASED estimators. In Tables E.8 – E.9 we evaluate the performance measures reported in (31) across 1000 replications of model (E.1).

Table E.4: See description of Table 1, for $S_0 = 3$

Panel I										
	p/T		200			500			1000	
ϕ		0.5	0.8	0.9	0.5	0.8	0.9	0.5	0.8	0.9
LASSO/GLS LASSO	100	1.208	1.548	1.634	1.259	1.733	1.905	1.284	1.894	2.221
GLS LASSO		(0.03)	(0.03)	(0.03)	(0.02)	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)
LASSO/GLS LASSO	200	1.201	1.516	1.603	1.240	1.674	1.838	1.289	1.882	2.167
GLS LASSO		(0.02)	(0.02)	(0.02)	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
LASSO/GLS LASSO	500	1.196	1.471	1.529	1.231	1.665	1.838	1.283	1.853	2.078
GLS LASSO		(0.02)	(0.02)	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Panel II										
DEBIASED LASSO/DEBIASED GLS	100	1.237	1.585	1.680	1.268	1.795	1.997	1.287	1.951	2.298
DEBIASED GLS		(0.06)	(0.06)	(0.06)	(0.04)	(0.03)	(0.03)	(0.03)	(0.02)	(0.02)
DEBIASED LASSO/DEBIASED GLS	200	1.236	1.559	1.629	1.275	1.797	2.000	1.287	1.926	2.215
DEBIASED GLS		(0.06)	(0.05)	(0.05)	(0.04)	(0.03)	(0.03)	(0.03)	(0.02)	(0.02)
DEBIASED LASSO/DEBIASED GLS	500	1.230	1.499	1.549	1.282	1.778	1.964	1.292	1.894	2.163
DEBIASED GLS		(0.05)	(0.05)	(0.05)	(0.04)	(0.03)	(0.03)	(0.03)	(0.02)	(0.02)

Table E.5: See description of Table 1, for $S_0 = 7$

Panel I										
	p/T		200			500			1000	
ϕ		0.5	0.8	0.9	0.5	0.8	0.9	0.5	0.8	0.9
LASSO/GLS LASSO	100	1.161	1.339	1.370	1.255	1.689	1.823	1.252	1.804	2.063
GLS LASSO		(0.04)	(0.04)	(0.04)	(0.02)	(0.02)	(0.02)	(0.02)	(0.01)	(0.01)
LASSO/GLS LASSO	200	1.138	1.293	1.344	1.251	1.618	1.702	1.261	1.803	2.027
GLS LASSO		(0.03)	(0.03)	(0.03)	(0.02)	(0.02)	(0.02)	(0.01)	(0.01)	(0.01)
LASSO/GLS LASSO	500	1.124	1.294	1.337	1.236	1.496	1.534	1.291	1.825	1.994
GLS LASSO		(0.02)	(0.02)	(0.02)	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Panel II										
DEBIASED LASSO/DEBIASED GLS	100	1.194	1.434	1.487	1.263	1.709	1.844	1.276	1.892	2.183
DEBIASED GLS		(0.06)	(0.06)	(0.06)	(0.04)	(0.03)	(0.03)	(0.03)	(0.02)	(0.02)
DEBIASED LASSO/DEBIASED GLS	200	1.193	1.409	1.457	1.255	1.636	1.734	1.281	1.865	2.089
DEBIASED GLS		(0.06)	(0.06)	(0.06)	(0.04)	(0.03)	(0.03)	(0.03)	(0.02)	(0.02)
DEBIASED LASSO/DEBIASED GLS	500	1.191	1.350	1.377	1.242	1.579	1.691	1.283	1.784	1.933
DEBIASED GLS		(0.05)	(0.06)	(0.06)	(0.04)	(0.03)	(0.03)	(0.03)	(0.02)	(0.02)

Table E.6: See description of Table 2, $S_0 = 3$

Size adjusted power and standard errors										
	p/T	200			500			1000		
ϕ		0.5	0.8	0.9	0.5	0.8	0.9	0.5	0.8	0.9
GLS LASSO	100	0.848	0.870	0.861	0.893	0.917	0.922	0.932	0.946	0.956
(s.e.)		(0.07)	(0.06)	(0.06)	(0.04)	(0.03)	(0.03)	(0.03)	(0.02)	(0.02)
LASSO		0.791	0.772	0.760	0.878	0.868	0.863	0.913	0.899	0.891
(s.e.)		(0.08)	(0.09)	(0.09)	(0.05)	(0.06)	(0.06)	(0.04)	(0.04)	(0.04)
GLS LASSO	200	0.860	0.874	0.878	0.917	0.935	0.938	0.952	0.962	0.964
(s.e.)		(0.07)	(0.06)	(0.06)	(0.04)	(0.04)	(0.03)	(0.03)	(0.02)	(0.02)
LASSO		0.834	0.810	0.802	0.905	0.885	0.880	0.934	0.922	0.917
(s.e.)		(0.08)	(0.09)	(0.10)	(0.05)	(0.06)	(0.06)	(0.04)	(0.04)	(0.04)
GLS LASSO	500	0.848	0.860	0.857	0.920	0.931	0.937	0.944	0.955	0.959
(s.e.)		(0.07)	(0.07)	(0.07)	(0.04)	(0.04)	(0.04)	(0.03)	(0.02)	(0.02)
LASSO		0.828	0.815	0.802	0.897	0.878	0.872	0.929	0.914	0.908
(s.e.)		(0.08)	(0.09)	(0.10)	(0.05)	(0.06)	(0.06)	(0.04)	(0.04)	(0.04)

Table E.7: See description of Table 2, $S_0 = 7$

Size adjusted power and standard errors										
	p/T	200			500			1000		
ϕ		0.5	0.8	0.9	0.5	0.8	0.9	0.5	0.8	0.9
GLS LASSO	100	0.708	0.714	0.705	0.730	0.718	0.718	0.725	0.730	0.738
(s.e.)		(0.07)	(0.07)	(0.07)	(0.04)	(0.04)	(0.04)	(0.03)	(0.02)	(0.02)
LASSO		0.701	0.688	0.679	0.717	0.706	0.705	0.724	0.720	0.719
(s.e.)		(0.08)	(0.09)	(0.10)	(0.05)	(0.06)	(0.06)	(0.04)	(0.04)	(0.04)
GLS LASSO	200	0.835	0.830	0.826	0.905	0.915	0.916	0.934	0.947	0.950
(s.e.)		(0.07)	(0.07)	(0.07)	(0.04)	(0.04)	(0.04)	(0.03)	(0.02)	(0.02)
LASSO		0.801	0.772	0.765	0.884	0.869	0.855	0.923	0.909	0.904
(s.e.)		(0.08)	(0.09)	(0.10)	(0.05)	(0.06)	(0.06)	(0.04)	(0.04)	(0.04)
GLS LASSO	500	0.841	0.827	0.823	0.908	0.910	0.911	0.940	0.948	0.951
(s.e.)		(0.07)	(0.07)	(0.08)	(0.04)	(0.04)	(0.04)	(0.03)	(0.03)	(0.03)
LASSO		0.823	0.784	0.774	0.891	0.871	0.864	0.925	0.914	0.909
(s.e.)		(0.08)	(0.10)	(0.10)	(0.05)	(0.06)	(0.06)	(0.04)	(0.04)	(0.04)

Table E.8: See description of Table 3, for $S_0 = 3$

		p/T	200			500			1000		
ϕ			0.5	0.8	0.9	0.5	0.8	0.9	0.5	0.8	0.9
DEBIASED LASSO	AvgCov S_0	100	0.924	0.929	0.930	0.924	0.933	0.938	0.925	0.927	0.927
	AvgCov S_0^c		0.954	0.954	0.955	0.953	0.952	0.953	0.949	0.950	0.950
	AvgLength		0.309	0.353	0.371	0.196	0.224	0.236	0.138	0.159	0.167
	AvgLength^c		0.308	0.353	0.371	0.196	0.224	0.235	0.138	0.159	0.167
DEBIASED GLS	AvgCov S_0		0.915	0.915	0.919	0.916	0.917	0.922	0.920	0.918	0.920
	AvgCov S_0^c		0.958	0.965	0.967	0.955	0.966	0.970	0.953	0.965	0.973
	AvgLength		0.256	0.241	0.243	0.157	0.136	0.133	0.109	0.088	0.083
	AvgLength^c		0.256	0.240	0.243	0.157	0.136	0.133	0.109	0.088	0.083
DEBIASED LASSO	AvgCov S_0	200	0.913	0.917	0.917	0.932	0.932	0.937	0.935	0.937	0.938
	AvgCov S_0^c		0.962	0.962	0.962	0.955	0.955	0.955	0.953	0.953	0.953
	AvgLength		0.311	0.356	0.373	0.196	0.225	0.236	0.139	0.159	0.167
	AvgLength^c		0.311	0.355	0.373	0.196	0.225	0.236	0.139	0.159	0.167
DEBIASED GLS	AvgCov S_0		0.897	0.892	0.897	0.921	0.911	0.914	0.926	0.917	0.919
	AvgCov S_0^c		0.965	0.971	0.973	0.959	0.969	0.973	0.956	0.968	0.974
	AvgLength		0.258	0.247	0.252	0.158	0.138	0.134	0.110	0.091	0.087
	AvgLength^c		0.258	0.247	0.251	0.158	0.138	0.134	0.110	0.091	0.087
DEBIASED LASSO	AvgCov S_0	500	0.900	0.900	0.902	0.922	0.927	0.928	0.929	0.938	0.940
	AvgCov S_0^c		0.974	0.973	0.973	0.961	0.961	0.961	0.958	0.957	0.956
	AvgLength		0.314	0.359	0.377	0.197	0.225	0.237	0.139	0.159	0.167
	AvgLength^c		0.314	0.359	0.376	0.197	0.225	0.237	0.139	0.159	0.167
DEBIASED GLS	AvgCov S_0		0.885	0.872	0.874	0.912	0.908	0.912	0.917	0.908	0.913
	AvgCov S_0^c		0.976	0.980	0.980	0.966	0.974	0.977	0.962	0.973	0.978
	AvgLength		0.263	0.258	0.264	0.159	0.141	0.138	0.111	0.094	0.091
	AvgLength^c		0.263	0.258	0.264	0.159	0.141	0.138	0.111	0.094	0.091

Table E.9: See description of Table 3, for $S_0 = 7$

		p/T	200			500			1000		
ϕ			0.5	0.8	0.9	0.5	0.8	0.9	0.5	0.8	0.9
DEBIASED LASSO	AvgCov S_0	100	0.917	0.923	0.923	0.932	0.932	0.932	0.932	0.938	0.940
	AvgCov S_0^c		0.958	0.958	0.958	0.953	0.953	0.952	0.950	0.951	0.951
	AvgLength		0.318	0.363	0.382	0.198	0.227	0.238	0.139	0.160	0.168
	AvgLength^c		0.318	0.363	0.382	0.198	0.227	0.238	0.139	0.160	0.168
DEBIASED GLS	AvgCov S_0		0.902	0.897	0.897	0.920	0.919	0.926	0.922	0.920	0.922
	AvgCov S_0^c		0.960	0.963	0.964	0.956	0.966	0.970	0.953	0.965	0.972
	AvgLength		0.270	0.265	0.270	0.160	0.144	0.144	0.110	0.091	0.087
	AvgLength^c		0.270	0.265	0.270	0.160	0.144	0.144	0.110	0.091	0.087
DEBIASED LASSO	AvgCov S_0	200	0.908	0.914	0.915	0.926	0.929	0.930	0.927	0.932	0.933
	AvgCov S_0^c		0.967	0.967	0.967	0.958	0.957	0.957	0.954	0.955	0.955
	AvgLength		0.323	0.369	0.387	0.199	0.228	0.240	0.140	0.160	0.168
	AvgLength^c		0.323	0.369	0.387	0.199	0.228	0.240	0.140	0.160	0.168
DEBIASED GLS	AvgCov S_0		0.887	0.883	0.888	0.906	0.912	0.916	0.920	0.914	0.918
	AvgCov S_0^c		0.968	0.970	0.972	0.962	0.970	0.972	0.957	0.969	0.974
	AvgLength		0.276	0.275	0.281	0.163	0.152	0.154	0.111	0.094	0.092
	AvgLength^c		0.276	0.275	0.281	0.163	0.152	0.154	0.111	0.094	0.092
DEBIASED LASSO	AvgCov S_0	500	0.899	0.906	0.909	0.915	0.917	0.921	0.930	0.931	0.933
	AvgCov S_0^c		0.981	0.981	0.981	0.968	0.967	0.966	0.961	0.961	0.960
	AvgLength		0.332	0.379	0.397	0.201	0.230	0.242	0.141	0.161	0.169
	AvgLength^c		0.332	0.379	0.397	0.201	0.230	0.242	0.141	0.161	0.169
DEBIASED GLS	AvgCov S_0		0.873	0.870	0.876	0.902	0.894	0.896	0.918	0.917	0.923
	AvgCov S_0^c		0.982	0.982	0.983	0.971	0.976	0.977	0.965	0.976	0.979
	AvgLength		0.285	0.288	0.297	0.167	0.160	0.160	0.113	0.101	0.101
	AvgLength^c		0.285	0.288	0.297	0.167	0.160	0.160	0.113	0.101	0.101

E.3 Simulations with t-distributed errors and/or covariates

In this section we use simulations to examine the performance of the proposed and competing debiased methods under the Assumption of heavy tails in both (either) $\{\boldsymbol{x}_t\}$, $\{u_t\}$, in terms of size-adjusted power, standard errors, average coverage rates and average lengths of the confidence interval, as seen in Section 4 of the main paper. We further report the relative to GLS LASSO RMSE's, and examine the performance of the proposed estimators using the following data generating processes.

We consider the following three data generating processes that involve t-distributed errors and/or covariates:

DGP 1: $y_t = \boldsymbol{x}_t'\boldsymbol{\beta} + u_t,\ u_t = \phi u_{t-1} + \varepsilon_t,\ \boldsymbol{x}_t \sim$ i.i.d. $\mathcal{N}(0, 1)$ and $\varepsilon_t \sim t_d$,

DGP 2: $y_t = \boldsymbol{x}_t'\boldsymbol{\beta} + u_t,\ u_t = \phi u_{t-1} + \varepsilon_t,\ \boldsymbol{x}_t \sim t_d$ and $\varepsilon_t \sim t_d$,

DGP 3: $y_t = \boldsymbol{x}_t'\boldsymbol{\beta} + u_t,\ u_t = \phi u_{t-1} + \varepsilon_t,\ \boldsymbol{x}_t \sim t_d$ and $\varepsilon_t \sim$ i.i.d. $\mathcal{N}(0, 1)$,

where $t = 1, \ldots, T$ and $d \in \{4, 8, 16\}$. We report these results in Tables E.10 – E.63, where the reported quantities have been described in detail in Section 6 of the main paper.

In DGP 1 we consider the case where only $\varepsilon_t \sim t_d$ and $\boldsymbol{x}_t \sim$ i.i.d. $\mathcal{N}(0, 1)$. The first panel of Tables E.10 – E.15 report the ratio of the average root mean squared error (RMSE) of the LASSO estimator over the RMSE of the GLS LASSO. The second panel of Table E.1, reports the ratio of the average RMSE of the DEBIASED LASSO estimator over the RMSE of DEBIASED GLS LASSO. Entries larger than 1 indicate superiority of the competing model (GLS LASSO). In the parenthesis, reported are the entries corresponding to the RMSE of the GLS LASSO in Panel I and the DEBIASED GLS LASSO in Panel II. The results in Tables E.10 – E.15 suggest similar patterns with Table 1, of the main paper and E.1 of Section E, in their respective cases. Note that as d increase the proposed (debiased) estimator report smaller RMSE, see for example Table E.10 compared to Table E.12 in the case of $T = 1000$, $|S_0| = 3$, any p, ϕ, where the reported quantities are slightly larger in the case of $d = 4$ compared to $d = 16$. Similar, patterns appear in the cases of $|S_0| = 7$, in Tables E.13 – E.15.

Further, in terms of average size-adjusted power, and considering the case where $\varepsilon_t \sim t_4$, $\phi = 0$, $|S_0| = 3$ in Table E.16, we find that both debiased estimators slightly under-reject the null, $H_0 : \beta_i = 0,\ i \in S_0$ as the sample size increases, with the highest reported size-adjusted power for the DEBIASED GLS LASSO to be 89.5%. The latter is expected as the sampling distribution of $\{\varepsilon_t\}$ has less moments, indicative of a heavy tailed distribution, which in turn affects the distribution of $\{u_t\}$ and consequently the frequency of rejecting the null, while the null is false. This phenomena is corrected as the degrees of freedom increase, i.e. as the t-distribution moves closer to the standard normal $N(0, 1)$. Indeed, looking at Table E.18, for $T = 1000$, $p = 500$ DEBIASED GLS LASSO reports 92% power, incrementally higher than the cases where $d = 4$. Interestingly, looking at Tables E.16 – E.18, and as ϕ and (T, p) increase, GLS LASSO reports power approximately equal to the nominal rate 95%, while in cases of $\phi > 0$, the DEBIASED LASSO severely under-performs compared to the proposed methodology, showing signs of inefficiency, while it reports power as much as 83.6% in the case where $T = 1000$, $p = 500$, $d = 16$, $\phi = 0.9$ and $|S_0| = 3$, as opposed to the DEBIASED GLS LASSO which reports power as high as 94.9% in the same design. Similar behaviour is reported in cases of less sparsity, where $|S_0| = 7$, see for example Tables E.19 – E.21.

In DGP 2 we consider the case of where both the covariates and the error term are sampled from a t-student distribution, $\boldsymbol{x}_t \sim t_d$, $\varepsilon_t \sim t_d$, indicative of heavier tails compared to the normal. While all DGP's in this section have an empirical motivation, DGP 2 presents the most general case among the three, and one usually found in financial returns. Looking at Tables E.28 – E.33, similar patterns as in Tables E.10 – E.15 appear to hold, with a slight increase in the debiased estimators' RMSE's is reported. The latter is more profound in Table E.28 and smaller sample sizes, and disappears as T increase.

Further, in terms of average size-adjusted power, similar patterns with DGP 1 can be observed, with a slight increase now in the s.e. followed by a decrease in size-adjusted power, for both $|S_0| \in \{3, 7\}$. This behaviour vanishes as T, p increase. The proposed estimator, DEBIASED GLS LASSO reports size-adjusted power close to 95% as T, p and importantly increase, while the competing debiased estimator under-performs severly reporting the highest power close to $\sim 76 - 77\%$, see for example Table E.20 $T = 1000$, $p = 500$, $\phi = 0.9, 0.95$.

Finally, in DGP 3, similar patterns as in DGP 1, and DGP 2 arise.

In summary, in this set of simulations, we outline cases that are of empirical interest, while showcasing that the proposed estimators perform as expected under both the Assumption of heavier or thin tailed distributions in $\{\boldsymbol{x}_t\}$ and $\{u_t\}$ are in place. Further, in the cases of autocorrelation in the error term, it appears that the properties of standard methods, such as the LASSO and DEBIASED LASSO are violated, hence these methods under-perform severely, making them unreliable, at least in time-series settings.

In Tables E.22 – E.27, E.40 – E.45, and E.58 – E.63 we evaluate the performance measures reported in (31) across 1000 replications of DGP 1, DGP 2, and DGP 3 respectively. Our findings are in line with the finding os Table 3 of the main paper and E.3 of the supplement.

Table E.10: See description of Table 1, for $|S_0| = 3$, $d = 4$, DGP 1 ($\boldsymbol{x}_t \sim$ i.i.d. $\mathcal{N}(0, 1)$ and $\varepsilon_t \sim t_d$).

								$d = 4$								
Panel I																
	p/T			200					500					1000		
ϕ		0	0.5	0.8	0.9	0.95	0	0.5	0.8	0.9	0.95	0	0.5	0.8	0.9	0.95
LASSO/ GLS LASSO	100	0.993	1.223	1.749	2.171	2.573	0.994	1.183	1.896	2.598	3.129	0.999	1.238	1.731	2.431	3.458
GLS LASSO		(0.04)	(0.04)	(0.03)	(0.03)	(0.03)	(0.03)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
LASSO/ GLS LASSO	200	0.997	1.212	1.672	1.991	2.389	0.991	1.182	2.013	2.598	3.040	0.999	1.218	1.672	2.601	3.466
GLS LASSO		(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.01)	(0.01)	(0.01)
LASSO/ GLS LASSO	500	0.997	1.185	1.605	1.967	2.331	0.992	1.265	1.989	2.324	2.829	0.997	1.142	1.893	2.693	3.226
GLS LASSO		(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Panel II																
DEBIASED LASSO/ DEBIASED GLS	100	1.004	1.285	2.069	2.819	3.610	1.001	1.293	2.105	2.967	4.048	1.000	1.292	2.114	3.004	4.178
DEBIASED GLS		(0.10)	(0.08)	(0.07)	(0.07)	(0.07)	(0.06)	(0.06)	(0.05)	(0.05)	(0.05)	(0.04)	(0.04)	(0.03)	(0.03)	(0.03)
DEBIASED LASSO/ DEBIASED GLS	200	1.005	1.290	2.102	2.906	3.657	1.002	1.298	2.109	2.979	4.073	1.001	1.295	2.124	3.016	4.212
DEBIASED GLS		(0.09)	(0.08)	(0.07)	(0.07)	(0.07)	(0.06)	(0.05)	(0.05)	(0.05)	(0.05)	(0.04)	(0.04)	(0.03)	(0.03)	(0.03)
DEBIASED LASSO/ DEBIASED GLS	500	1.007	1.287	2.146	2.954	3.657	1.003	1.300	2.122	3.067	4.190	1.001	1.303	2.117	3.004	4.253
DEBIASED GLS		(0.08)	(0.07)	(0.07)	(0.06)	(0.06)	(0.06)	(0.05)	(0.05)	(0.04)	(0.04)	(0.04)	(0.04)	(0.03)	(0.03)	(0.03)

Table E.11: See description of Table 1, for $|S_0| = 3,\ d = 8$, DGP 1 ($\boldsymbol{x}_t \sim$ i.i.d. $\mathcal{N}(0,1)$ and $\varepsilon_t \sim t_d$).

				$d = 8$												
Panel I																
	p/T			200					500					1000		
ϕ		0	0.5	0.8	0.9	0.95	0	0.5	0.8	0.9	0.95	0	0.5	0.8	0.9	0.95
LASSO/ GLS LASSO	100	0.987	1.212	1.865	2.302	2.699	0.996	1.209	1.755	2.522	3.281	0.998	1.270	1.859	2.387	3.549
GLS LASSO		(0.04)	(0.03)	(0.03)	(0.03)	(0.03)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)
LASSO/ GLS LASSO	200	0.995	1.214	1.810	2.225	2.643	0.996	1.163	1.806	2.577	3.237	0.999	1.251	1.649	2.287	3.513
GLS LASSO		(0.03)	(0.03)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
LASSO/ GLS LASSO	500	0.992	1.234	1.731	2.133	2.553	0.991	1.174	2.012	2.636	3.022	0.998	1.205	1.532	2.389	3.354
GLS LASSO		(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Panel II																
DEBIASED LASSO/ DEBIASED GLS	100	1.004	1.275	2.033	2.757	3.529	1.001	1.290	2.099	2.955	4.021	1.001	1.287	2.112	3.012	4.185
DEBIASED GLS		(0.08)	(0.07)	(0.06)	(0.06)	(0.06)	(0.05)	(0.05)	(0.04)	(0.04)	(0.04)	(0.04)	(0.03)	(0.03)	(0.03)	(0.03)
DEBIASED LASSO/ DEBIASED GLS	200	1.006	1.284	2.057	2.811	3.572	1.002	1.296	2.101	2.937	4.005	1.001	1.295	2.135	3.029	4.184
DEBIASED GLS		(0.07)	(0.07)	(0.06)	(0.06)	(0.06)	(0.05)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.03)	(0.03)	(0.03)	(0.03)
DEBIASED LASSO/ DEBIASED GLS	500	1.009	1.266	2.068	2.850	3.510	1.003	1.305	2.095	2.935	4.069	1.001	1.303	2.159	3.031	4.208
DEBIASED GLS		(0.07)	(0.06)	(0.05)	(0.05)	(0.05)	(0.05)	(0.04)	(0.04)	(0.04)	(0.04)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)

Table E.12: See description of Table 1, for $|S_0| = 3,\ d = 16$, DGP 1 ($\boldsymbol{x}_t \sim$ i.i.d. $\mathcal{N}(0,1)$ and $\varepsilon_t \sim t_d$).

				$d = 16$												
Panel I																
	p/T			200					500					1000		
ϕ		0	0.5	0.8	0.9	0.95	0	0.5	0.8	0.9	0.95	0	0.5	0.8	0.9	0.95
LASSO/ GLS LASSO	100	0.994	1.177	1.779	2.316	2.737	0.999	1.239	1.762	2.482	3.350	0.999	1.272	1.969	2.425	3.541
GLS LASSO		(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
LASSO/ GLS LASSO	200	0.989	1.210	1.862	2.291	2.709	0.994	1.185	1.677	2.463	3.214	0.999	1.271	1.768	2.280	3.530
GLS LASSO		(0.03)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
LASSO/ GLS LASSO	500	0.993	1.236	1.801	2.192	2.633	0.994	1.149	1.865	2.586	3.063	1.000	1.220	1.514	2.299	3.430
GLS LASSO		(0.02)	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Panel II																
DEBIASED LASSO/ DEBIASED GLS	100	1.004	1.274	2.022	2.736	3.507	1.001	1.294	2.109	2.957	3.995	1.000	1.289	2.114	3.016	4.190
DEBIASED GLS		(0.07)	(0.06)	(0.06)	(0.06)	(0.06)	(0.05)	(0.04)	(0.04)	(0.04)	(0.03)	(0.03)	(0.03)	(0.03)	(0.02)	(0.02)
DEBIASED LASSO/ DEBIASED GLS	200	1.007	1.281	2.028	2.759	3.515	1.002	1.299	2.114	2.940	3.991	1.001	1.294	2.135	3.047	4.201
DEBIASED GLS		(0.07)	(0.06)	(0.05)	(0.05)	(0.05)	(0.05)	(0.04)	(0.04)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.02)	(0.02)
DEBIASED LASSO/ DEBIASED GLS	500	1.008	1.261	2.020	2.794	3.471	1.003	1.309	2.103	2.913	4.026	1.001	1.302	2.172	3.055	4.194
DEBIASED GLS		(0.06)	(0.06)	(0.05)	(0.05)	(0.05)	(0.04)	(0.04)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.02)	(0.02)	(0.02)

Table E.13: See description of Table 1, for $|S_0| = 7,\ d = 4$, DGP 1 ($\boldsymbol{x}_t \sim$ i.i.d. $\mathcal{N}(0,1)$ and $\varepsilon_t \sim t_d$).

				$d = 4$												
Panel I																
	p/T			200					500					1000		
ϕ		0	0.5	0.8	0.9	0.95	0	0.5	0.8	0.9	0.95	0	0.5	0.8	0.9	0.95
LASSO/ GLS LASSO	100	0.993	1.223	1.749	2.171	2.573	0.994	1.183	1.896	2.598	3.129	0.999	1.238	1.731	2.431	3.458
GLS LASSO		(0.04)	(0.04)	(0.03)	(0.03)	(0.03)	(0.03)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
LASSO/ GLS LASSO	200	0.997	1.212	1.672	1.991	2.389	0.991	1.182	2.013	2.598	3.040	0.999	1.218	1.672	2.601	3.466
GLS LASSO		(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.01)	(0.01)	(0.01)
LASSO/ GLS LASSO	500	0.997	1.185	1.605	1.967	2.331	0.992	1.265	1.989	2.324	2.829	0.997	1.142	1.893	2.693	3.226
GLS LASSO		(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Panel II																
DEBIASED LASSO/ DEBIASED GLS	100	1.002	1.262	1.977	2.658	3.368	1.000	1.283	2.068	2.878	3.892	1.000	1.288	2.105	2.979	4.103
DEBIASED GLS		(0.10)	(0.09)	(0.08)	(0.08)	(0.08)	(0.06)	(0.06)	(0.05)	(0.05)	(0.05)	(0.04)	(0.04)	(0.03)	(0.03)	(0.03)
DEBIASED LASSO/ DEBIASED GLS	200	1.006	1.255	1.980	2.696	3.419	1.001	1.294	2.087	2.881	3.894	1.000	1.293	2.130	3.009	4.127
DEBIASED GLS		(0.09)	(0.08)	(0.07)	(0.07)	(0.07)	(0.06)	(0.05)	(0.05)	(0.05)	(0.05)	(0.04)	(0.04)	(0.03)	(0.03)	(0.03)
DEBIASED LASSO/ DEBIASED GLS	500	1.007	1.247	2.014	2.798	3.461	1.003	1.307	2.055	2.865	3.960	1.001	1.307	2.178	3.012	4.124
DEBIASED GLS		(0.08)	(0.07)	(0.07)	(0.07)	(0.07)	(0.06)	(0.05)	(0.05)	(0.05)	(0.05)	(0.04)	(0.04)	(0.03)	(0.03)	(0.03)

Table E.14: See description of Table 1, for $|S_0| = 7,\ d =$, DGP 1 ($\boldsymbol{x}_t \sim$ i.i.d. $\mathcal{N}(0,1)$ and $\varepsilon_t \sim t_d$).

								$d = 8$								
Panel I																
	p/T			200					500					1000		
ϕ		0	0.5	0.8	0.9	0.95	0	0.5	0.8	0.9	0.95	0	0.5	0.8	0.9	0.95
LASSO/ GLS LASSO	100	0.987	1.212	1.865	2.302	2.699	0.996	1.209	1.755	2.522	3.281	0.998	1.270	1.859	2.387	3.549
GLS LASSO		(0.04)	(0.03)	(0.03)	(0.03)	(0.03)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)
LASSO/ GLS LASSO	200	0.995	1.214	1.810	2.225	2.643	0.996	1.163	1.806	2.577	3.237	0.999	1.251	1.649	2.287	3.513
GLS LASSO		(0.03)	(0.03)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
LASSO/ GLS LASSO	500	0.992	1.234	1.731	2.133	2.553	0.991	1.174	2.012	2.636	3.022	0.998	1.205	1.532	2.389	3.354
GLS LASSO		(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Panel II																
DEBIASED LASSO/ DEBIASED GLS	100	1.002	1.257	1.959	2.617	3.302	1.001	1.285	2.084	2.899	3.882	1.000	1.285	2.100	2.981	4.117
DEBIASED GLS		(0.08)	(0.07)	(0.06)	(0.06)	(0.06)	(0.05)	(0.05)	(0.04)	(0.04)	(0.04)	(0.04)	(0.03)	(0.03)	(0.03)	(0.03)
DEBIASED LASSO/ DEBIASED GLS	200	1.005	1.259	1.953	2.609	3.284	1.001	1.292	2.106	2.890	3.840	1.001	1.289	2.117	3.013	4.117
DEBIASED GLS		(0.07)	(0.07)	(0.06)	(0.06)	(0.06)	(0.05)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.03)	(0.03)	(0.03)	(0.03)
DEBIASED LASSO/ DEBIASED GLS	500	1.007	1.239	1.885	2.598	3.268	1.002	1.312	2.151	2.867	3.863	1.001	1.298	2.170	3.067	4.083
DEBIASED GLS		(0.07)	(0.06)	(0.06)	(0.06)	(0.06)	(0.05)	(0.04)	(0.04)	(0.04)	(0.04)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)

Table E.15: See description of Table 1, for $|S_0| = 7,\ d =$, DGP 1 ($\boldsymbol{x}_t \sim$ i.i.d. $\mathcal{N}(0,1)$ and $\varepsilon_t \sim t_d$).

								$d = 16$								
Panel I																
	p/T			200					500					1000		
ϕ		0	0.5	0.8	0.9	0.95	0	0.5	0.8	0.9	0.95	0	0.5	0.8	0.9	0.95
LASSO/ GLS LASSO	100	0.994	1.177	1.779	2.316	2.737	0.999	1.239	1.762	2.482	3.350	0.999	1.272	1.969	2.425	3.541
GLS LASSO		(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
LASSO/ GLS LASSO	200	0.989	1.210	1.862	2.291	2.709	0.994	1.185	1.677	2.463	3.214	0.999	1.271	1.768	2.280	3.530
GLS LASSO		(0.03)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
LASSO/ GLS LASSO	500	0.993	1.236	1.801	2.192	2.633	0.994	1.149	1.865	2.586	3.063	1.000	1.220	1.514	2.299	3.430
GLS LASSO		(0.02)	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Panel II																
DEBIASED LASSO/ DEBIASED GLS	100	1.003	1.257	1.959	2.605	3.278	1.001	1.287	2.090	2.903	3.875	1.000	1.284	2.102	2.980	4.121
DEBIASED GLS		(0.07)	(0.07)	(0.06)	(0.06)	(0.06)	(0.05)	(0.04)	(0.04)	(0.04)	(0.04)	(0.03)	(0.03)	(0.03)	(0.03)	(0.02)
DEBIASED LASSO/ DEBIASED GLS	200	1.005	1.254	1.949	2.579	3.262	1.001	1.290	2.110	2.898	3.828	1.001	1.289	2.112	3.006	4.124
DEBIASED GLS		(0.07)	(0.06)	(0.06)	(0.06)	(0.06)	(0.05)	(0.04)	(0.04)	(0.04)	(0.04)	(0.03)	(0.03)	(0.03)	(0.02)	(0.02)
DEBIASED LASSO/ DEBIASED GLS	500	1.009	1.241	1.857	2.506	3.175	1.002	1.304	2.196	2.897	3.826	1.001	1.298	2.148	3.085	4.093
DEBIASED GLS		(0.06)	(0.06)	(0.05)	(0.05)	(0.05)	(0.04)	(0.04)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.02)	(0.02)	(0.02)

Table E.16: See description of Table 2, for $|S_0| = 3,\ d = 4$, DGP 1 ($\boldsymbol{x}_t \sim$ i.i.d. $\mathcal{N}(0,1)$ and $\varepsilon_t \sim t_d$).

								Size adjusted power and standard errors								
								$d = 4$								
	p/T			200					500					1000		
ϕ		0	0.5	0.8	0.9	0.95	0	0.5	0.8	0.9	0.95	0	0.5	0.8	0.9	0.95
GLS LASSO	100	0.764	0.786	0.823	0.825	0.825	0.855	0.872	0.883	0.884	0.891	0.895	0.916	0.927	0.930	0.930
(s.e.)		(0.10)	(0.09)	(0.08)	(0.08)	(0.08)	(0.06)	(0.06)	(0.05)	(0.05)	(0.05)	(0.04)	(0.04)	(0.04)	(0.03)	(0.03)
LASSO		0.764	0.724	0.630	0.532	0.410	0.857	0.835	0.773	0.710	0.582	0.894	0.884	0.839	0.786	0.717
(s.e.)		(0.10)	(0.11)	(0.16)	(0.21)	(0.27)	(0.06)	(0.07)	(0.10)	(0.14)	(0.19)	(0.04)	(0.05)	(0.07)	(0.10)	(0.14)
GLS LASSO	200	0.785	0.823	0.839	0.852	0.854	0.875	0.899	0.903	0.911	0.908	0.915	0.921	0.933	0.937	0.941
(s.e.)		(0.10)	(0.09)	(0.08)	(0.08)	(0.08)	(0.06)	(0.06)	(0.05)	(0.05)	(0.05)	(0.04)	(0.04)	(0.03)	(0.03)	(0.03)
LASSO		0.791	0.753	0.677	0.568	0.441	0.878	0.856	0.780	0.719	0.639	0.916	0.898	0.849	0.813	0.730
(s.e.)		(0.10)	(0.11)	(0.16)	(0.21)	(0.27)	(0.06)	(0.07)	(0.10)	(0.14)	(0.19)	(0.04)	(0.05)	(0.07)	(0.10)	(0.14)
GLS LASSO	500	0.805	0.811	0.832	0.837	0.836	0.867	0.885	0.892	0.900	0.903	0.912	0.919	0.932	0.937	0.935
(s.e.)		(0.10)	(0.09)	(0.08)	(0.08)	(0.08)	(0.06)	(0.06)	(0.05)	(0.05)	(0.05)	(0.04)	(0.04)	(0.04)	(0.03)	(0.03)
LASSO		0.803	0.780	0.693	0.621	0.527	0.868	0.854	0.787	0.699	0.616	0.911	0.898	0.855	0.801	0.723
(s.e.)		(0.10)	(0.11)	(0.16)	(0.21)	(0.27)	(0.06)	(0.07)	(0.10)	(0.14)	(0.19)	(0.04)	(0.05)	(0.07)	(0.10)	(0.14)

Table E.17: See description of Table 2, for $|S_0| = 3,\ d = 8$, DGP 1 ($\boldsymbol{x}_t \sim$ i.i.d. $\mathcal{N}(0,1)$ and $\varepsilon_t \sim t_d$).

Size adjusted power and standard errors																
$d=8$																
	p/T			200					500					1000		
ϕ		0	0.5	0.8	0.9	0.95	0	0.5	0.8	0.9	0.95	0	0.5	0.8	0.9	0.95
GLS LASSO	100	0.823	0.826	0.857	0.863	0.853	0.891	0.892	0.902	0.906	0.908	0.920	0.928	0.933	0.937	0.941
(s.e.)		(0.08)	(0.07)	(0.07)	(0.07)	(0.07)	(0.05)	(0.05)	(0.04)	(0.04)	(0.04)	(0.04)	(0.03)	(0.03)	(0.03)	(0.03)
LASSO		0.824	0.772	0.714	0.627	0.505	0.888	0.888	0.824	0.759	0.662	0.919	0.903	0.876	0.837	0.773
(s.e.)		(0.08)	(0.09)	(0.13)	(0.17)	(0.22)	(0.05)	(0.06)	(0.09)	(0.12)	(0.16)	(0.04)	(0.04)	(0.06)	(0.08)	(0.11)
GLS LASSO	200	0.834	0.856	0.877	0.874	0.884	0.900	0.907	0.924	0.919	0.928	0.923	0.932	0.944	0.945	0.949
(s.e.)		(0.08)	(0.07)	(0.07)	(0.07)	(0.07)	(0.05)	(0.05)	(0.04)	(0.04)	(0.04)	(0.04)	(0.03)	(0.03)	(0.03)	(0.03)
LASSO		0.836	0.790	0.717	0.628	0.533	0.900	0.886	0.825	0.767	0.684	0.922	0.913	0.883	0.832	0.776
(s.e.)		(0.08)	(0.09)	(0.13)	(0.17)	(0.22)	(0.05)	(0.06)	(0.09)	(0.12)	(0.16)	(0.04)	(0.04)	(0.06)	(0.08)	(0.11)
GLS LASSO	500	0.828	0.835	0.847	0.848	0.845	0.890	0.902	0.915	0.918	0.914	0.925	0.930	0.943	0.947	0.948
(s.e.)		(0.08)	(0.08)	(0.07)	(0.07)	(0.07)	(0.05)	(0.05)	(0.04)	(0.04)	(0.04)	(0.04)	(0.03)	(0.03)	(0.03)	(0.03)
LASSO		0.822	0.806	0.748	0.664	0.568	0.891	0.873	0.827	0.766	0.686	0.925	0.911	0.870	0.826	0.763
(s.e.)		(0.08)	(0.09)	(0.13)	(0.17)	(0.23)	(0.05)	(0.06)	(0.09)	(0.12)	(0.16)	(0.04)	(0.04)	(0.06)	(0.08)	(0.11)

Table E.18: See description of Table 2, for $|S_0| = 3,\ d = 16$, DGP 1 ($\boldsymbol{x}_t \sim$ i.i.d. $\mathcal{N}(0,1)$ and $\varepsilon_t \sim t_d$).

Size adjusted power and standard errors																
$d=16$																
	p/T			200					500					1000		
ϕ		0	0.5	0.8	0.9	0.95	0	0.5	0.8	0.9	0.95	0	0.5	0.8	0.9	0.95
GLS LASSO	100	0.829	0.851	0.860	0.856	0.855	0.881	0.895	0.912	0.917	0.916	0.922	0.923	0.932	0.936	0.934
(s.e.)		(0.07)	(0.07)	(0.06)	(0.06)	(0.06)	(0.05)	(0.04)	(0.04)	(0.04)	(0.04)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
LASSO		0.833	0.799	0.703	0.627	0.566	0.879	0.873	0.809	0.748	0.641	0.923	0.916	0.882	0.835	0.796
(s.e.)		(0.08)	(0.09)	(0.12)	(0.16)	(0.21)	(0.05)	(0.06)	(0.08)	(0.11)	(0.14)	(0.03)	(0.04)	(0.06)	(0.08)	(0.10)
GLS LASSO	200	0.847	0.859	0.877	0.882	0.886	0.918	0.921	0.930	0.936	0.936	0.936	0.943	0.949	0.951	0.955
(s.e.)		(0.08)	(0.07)	(0.06)	(0.06)	(0.06)	(0.05)	(0.04)	(0.04)	(0.04)	(0.04)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
LASSO		0.847	0.816	0.741	0.649	0.548	0.918	0.899	0.849	0.796	0.746	0.934	0.925	0.891	0.852	0.801
(s.e.)		(0.08)	(0.09)	(0.12)	(0.16)	(0.21)	(0.05)	(0.06)	(0.08)	(0.11)	(0.14)	(0.03)	(0.04)	(0.06)	(0.08)	(0.11)
GLS LASSO	500	0.834	0.836	0.857	0.858	0.859	0.905	0.915	0.923	0.929	0.929	0.922	0.932	0.943	0.949	0.951
(s.e.)		(0.08)	(0.07)	(0.06)	(0.06)	(0.07)	(0.05)	(0.04)	(0.04)	(0.04)	(0.04)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
LASSO		0.836	0.817	0.761	0.682	0.583	0.906	0.889	0.839	0.783	0.720	0.923	0.916	0.879	0.836	0.775
(s.e.)		(0.08)	(0.09)	(0.12)	(0.16)	(0.21)	(0.05)	(0.06)	(0.08)	(0.11)	(0.14)	(0.03)	(0.04)	(0.06)	(0.08)	(0.11)

Table E.19: See description of Table 2, for $|S_0| = 7,\ d = 4$, DGP 1 ($\boldsymbol{x}_t \sim$ i.i.d. $\mathcal{N}(0,1)$ and $\varepsilon_t \sim t_d$).

Size adjusted power and standard errors																
$d=4$																
	p/T			200					500					1000		
ϕ		0	0.5	0.8	0.9	0.95	0	0.5	0.8	0.9	0.95	0	0.5	0.8	0.9	0.95
GLS LASSO	100	0.721	0.725	0.734	0.733	0.719	0.761	0.762	0.761	0.753	0.748	0.731	0.737	0.740	0.749	0.749
(s.e.)		(0.10)	(0.09)	(0.09)	(0.09)	(0.09)	(0.06)	(0.06)	(0.05)	(0.05)	(0.05)	(0.04)	(0.04)	(0.04)	(0.03)	(0.04)
LASSO		0.723	0.677	0.580	0.464	0.428	0.762	0.741	0.676	0.602	0.529	0.730	0.737	0.726	0.727	0.625
(s.e.)		(0.10)	(0.12)	(0.16)	(0.21)	(0.28)	(0.06)	(0.07)	(0.11)	(0.14)	(0.19)	(0.04)	(0.05)	(0.07)	(0.10)	(0.14)
GLS LASSO	200	0.755	0.773	0.795	0.799	0.802	0.851	0.861	0.882	0.886	0.884	0.894	0.909	0.919	0.921	0.922
(s.e.)		(0.10)	(0.10)	(0.09)	(0.09)	(0.09)	(0.06)	(0.06)	(0.05)	(0.05)	(0.05)	(0.05)	(0.04)	(0.04)	(0.03)	(0.04)
LASSO		0.754	0.723	0.634	0.524	0.432	0.851	0.839	0.772	0.699	0.609	0.894	0.879	0.832	0.778	0.704
(s.e.)		(0.10)	(0.12)	(0.16)	(0.22)	(0.28)	(0.06)	(0.07)	(0.11)	(0.14)	(0.19)	(0.05)	(0.05)	(0.07)	(0.10)	(0.14)
GLS LASSO	500	0.776	0.781	0.799	0.792	0.796	0.861	0.877	0.888	0.890	0.891	0.903	0.911	0.925	0.926	0.925
(s.e.)		(0.11)	(0.10)	(0.09)	(0.09)	(0.10)	(0.06)	(0.06)	(0.05)	(0.05)	(0.06)	(0.05)	(0.04)	(0.04)	(0.04)	(0.04)
LASSO		0.774	0.742	0.660	0.529	0.421	0.861	0.844	0.779	0.702	0.605	0.903	0.888	0.840	0.784	0.708
(s.e.)		(0.11)	(0.12)	(0.17)	(0.22)	(0.28)	(0.06)	(0.07)	(0.11)	(0.14)	(0.19)	(0.05)	(0.05)	(0.07)	(0.10)	(0.14)

Table E.20: See description of Table 2, for $|S_0| = 7,\ d = 8$, DGP 1 ($\boldsymbol{x}_t \sim$ i.i.d. $\mathcal{N}(0,1)$ and $\varepsilon_t \sim t_d$).

Size adjusted power and standard errors																
$d = 8$																
	p/T	200					500					1000				
ϕ		0	0.5	0.8	0.9	0.95	0	0.5	0.8	0.9	0.95	0	0.5	0.8	0.9	0.95
GLS LASSO	100	0.697	0.684	0.700	0.717	0.699	0.762	0.784	0.784	0.786	0.792	0.735	0.737	0.737	0.747	0.735
(s.e.)		(0.08)	(0.08)	(0.07)	(0.07)	(0.07)	(0.05)	(0.05)	(0.04)	(0.04)	(0.04)	(0.04)	(0.03)	(0.03)	(0.03)	(0.03)
LASSO		0.692	0.658	0.600	0.551	0.481	0.760	0.763	0.718	0.674	0.562	0.734	0.726	0.712	0.695	0.660
(s.e.)		(0.08)	(0.10)	(0.13)	(0.18)	(0.23)	(0.05)	(0.06)	(0.09)	(0.12)	(0.16)	(0.04)	(0.04)	(0.06)	(0.08)	(0.11)
GLS LASSO	200	0.808	0.817	0.836	0.843	0.838	0.884	0.896	0.907	0.910	0.911	0.914	0.924	0.934	0.934	0.935
(s.e.)		(0.08)	(0.08)	(0.07)	(0.07)	(0.08)	(0.05)	(0.05)	(0.04)	(0.04)	(0.05)	(0.04)	(0.03)	(0.03)	(0.03)	(0.03)
LASSO		0.809	0.782	0.687	0.598	0.514	0.883	0.870	0.810	0.744	0.652	0.914	0.903	0.867	0.817	0.755
(s.e.)		(0.09)	(0.10)	(0.14)	(0.18)	(0.23)	(0.05)	(0.06)	(0.09)	(0.12)	(0.16)	(0.04)	(0.04)	(0.06)	(0.08)	(0.11)
GLS LASSO	500	0.811	0.831	0.838	0.838	0.839	0.890	0.898	0.909	0.910	0.908	0.919	0.927	0.937	0.938	0.938
(s.e.)		(0.09)	(0.08)	(0.08)	(0.08)	(0.08)	(0.05)	(0.05)	(0.04)	(0.04)	(0.05)	(0.04)	(0.03)	(0.03)	(0.03)	(0.03)
LASSO		0.812	0.786	0.715	0.632	0.546	0.888	0.868	0.812	0.753	0.663	0.920	0.908	0.871	0.828	0.755
(s.e.)		(0.09)	(0.10)	(0.14)	(0.18)	(0.23)	(0.05)	(0.06)	(0.09)	(0.12)	(0.16)	(0.04)	(0.04)	(0.06)	(0.08)	(0.11)

Table E.21: See description of Table 2, for $|S_0| = 7,\ d = 16$, DGP 1 ($\boldsymbol{x}_t \sim$ i.i.d. $\mathcal{N}(0,1)$ and $\varepsilon_t \sim t_d$).

Size adjusted power and standard errors																
$d = 16$																
	p/T	200					500					1000				
ϕ		0	0.5	0.8	0.9	0.95	0	0.5	0.8	0.9	0.95	0	0.5	0.8	0.9	0.95
GLS LASSO	100	0.677	0.700	0.709	0.715	0.705	0.777	0.779	0.777	0.778	0.772	0.747	0.737	0.735	0.738	0.739
(s.e.)		(0.08)	(0.07)	(0.06)	(0.07)	(0.07)	(0.05)	(0.04)	(0.04)	(0.04)	(0.04)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
LASSO		0.676	0.675	0.630	0.559	0.488	0.776	0.765	0.719	0.691	0.638	0.746	0.747	0.753	0.752	0.719
(s.e.)		(0.08)	(0.09)	(0.12)	(0.16)	(0.21)	(0.05)	(0.06)	(0.08)	(0.11)	(0.15)	(0.03)	(0.04)	(0.06)	(0.08)	(0.11)
GLS LASSO	200	0.806	0.818	0.837	0.839	0.841	0.891	0.897	0.911	0.914	0.912	0.928	0.931	0.942	0.945	0.945
(s.e.)		(0.08)	(0.07)	(0.07)	(0.07)	(0.07)	(0.05)	(0.04)	(0.04)	(0.04)	(0.04)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
LASSO		0.807	0.788	0.700	0.623	0.534	0.891	0.869	0.817	0.754	0.683	0.928	0.915	0.874	0.829	0.774
(s.e.)		(0.08)	(0.09)	(0.13)	(0.17)	(0.21)	(0.05)	(0.06)	(0.08)	(0.11)	(0.15)	(0.03)	(0.04)	(0.06)	(0.08)	(0.11)
GLS LASSO	500	0.818	0.827	0.841	0.842	0.840	0.894	0.901	0.912	0.912	0.914	0.925	0.931	0.941	0.944	0.943
(s.e.)		(0.08)	(0.08)	(0.07)	(0.07)	(0.08)	(0.05)	(0.04)	(0.04)	(0.04)	(0.04)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
LASSO		0.818	0.796	0.722	0.662	0.555	0.893	0.878	0.830	0.770	0.695	0.924	0.913	0.876	0.833	0.772
(s.e.)		(0.08)	(0.09)	(0.13)	(0.17)	(0.22)	(0.05)	(0.06)	(0.08)	(0.11)	(0.15)	(0.03)	(0.04)	(0.06)	(0.08)	(0.11)

Table E.22: See description of Table 3, for $|S_0| = 3$. $d = 4$, DGP 1 ($\boldsymbol{x}_t \sim$ i.i.d. $\mathcal{N}(0,1)$ and $\varepsilon_t \sim t_d$).

$d = 4$																	
		p/T	200					500					1000				
ϕ			0	0.5	0.8	0.9	0.95	0	0.5	0.8	0.9	0.95	0	0.5	0.8	0.9	0.95
DEBIASED LASSO	AvgCov S_0	100	0.928	0.927	0.937	0.939	0.939	0.940	0.937	0.943	0.945	0.950	0.942	0.936	0.935	0.937	0.942
	AvgCov S_0^c		0.955	0.954	0.951	0.950	0.948	0.951	0.952	0.951	0.950	0.950	0.951	0.951	0.952	0.952	0.953
	AvgLength		0.389	0.444	0.624	0.824	1.065	0.247	0.284	0.406	0.551	0.747	0.175	0.202	0.290	0.396	0.545
	AvgLength^c		0.389	0.444	0.624	0.824	1.065	0.247	0.284	0.406	0.551	0.747	0.175	0.202	0.290	0.396	0.545
DEBIASED GLS	AvgCov S_0		0.926	0.925	0.925	0.926	0.932	0.939	0.935	0.930	0.934	0.941	0.940	0.940	0.938	0.937	0.943
	AvgCov S_0^c		0.954	0.956	0.960	0.963	0.966	0.951	0.953	0.954	0.958	0.962	0.951	0.952	0.952	0.954	0.959
	AvgLength		0.386	0.351	0.314	0.309	0.316	0.246	0.221	0.195	0.190	0.193	0.175	0.157	0.137	0.132	0.134
	AvgLength^c		0.386	0.351	0.314	0.309	0.315	0.246	0.221	0.194	0.190	0.192	0.175	0.157	0.137	0.132	0.134
DEBIASED LASSO	AvgCov S_0	200	0.916	0.932	0.936	0.937	0.938	0.934	0.939	0.944	0.945	0.944	0.933	0.932	0.940	0.943	0.944
	AvgCov S_0^c		0.962	0.961	0.957	0.955	0.954	0.955	0.954	0.954	0.953	0.951	0.953	0.952	0.953	0.952	0.951
	AvgLength		0.391	0.447	0.627	0.824	1.057	0.248	0.285	0.407	0.551	0.744	0.175	0.202	0.290	0.395	0.544
	AvgLength^c		0.391	0.447	0.627	0.823	1.057	0.248	0.285	0.407	0.551	0.744	0.175	0.202	0.290	0.395	0.544
DEBIASED GLS	AvgCov S_0		0.911	0.908	0.903	0.907	0.915	0.932	0.932	0.924	0.928	0.933	0.933	0.933	0.929	0.932	0.938
	AvgCov S_0^c		0.961	0.964	0.968	0.969	0.972	0.954	0.956	0.958	0.962	0.965	0.952	0.953	0.954	0.957	0.963
	AvgLength		0.389	0.354	0.319	0.311	0.320	0.247	0.222	0.196	0.194	0.194	0.175	0.157	0.137	0.134	0.136
	AvgLength^c		0.389	0.353	0.318	0.311	0.319	0.247	0.222	0.196	0.194	0.194	0.175	0.157	0.137	0.134	0.136
DEBIASED LASSO	AvgCov S_0	500	0.919	0.922	0.937	0.950	0.951	0.934	0.929	0.943	0.948	0.950	0.944	0.941	0.939	0.943	0.948
	AvgCov S_0^c		0.972	0.971	0.964	0.960	0.959	0.961	0.962	0.961	0.957	0.955	0.956	0.955	0.957	0.956	0.954
	AvgLength		0.392	0.448	0.628	0.827	1.066	0.250	0.288	0.411	0.557	0.752	0.175	0.202	0.290	0.396	0.544
	AvgLength^c		0.392	0.447	0.628	0.826	1.066	0.250	0.288	0.411	0.557	0.752	0.175	0.202	0.290	0.396	0.544
DEBIASED GLS	AvgCov S_0		0.913	0.903	0.895	0.898	0.904	0.934	0.923	0.916	0.915	0.923	0.944	0.937	0.929	0.936	0.940
	AvgCov S_0^c		0.972	0.974	0.976	0.978	0.980	0.961	0.964	0.968	0.970	0.973	0.956	0.957	0.958	0.963	0.965
	AvgLength		0.389	0.356	0.320	0.317	0.329	0.249	0.224	0.200	0.195	0.198	0.175	0.157	0.138	0.136	0.136
	AvgLength^c		0.389	0.356	0.319	0.316	0.328	0.249	0.224	0.200	0.195	0.198	0.175	0.157	0.138	0.136	0.136

Table E.23: See description of Table 3, for $|S_0| = 3$. $d = 8$, DGP 1 ($\boldsymbol{x}_t \sim$ i.i.d. $\mathcal{N}(0,1)$ and $\varepsilon_t \sim t_d$).

			$d=8$														
		p/T	200					500					1000				
ϕ			0	0.5	0.8	0.9	0.95	0	0.5	0.8	0.9	0.95	0	0.5	0.8	0.9	0.95
DEBIASED LASSO	AvgCov S_0	100	0.918	0.922	0.934	0.933	0.934	0.927	0.926	0.938	0.943	0.948	0.925	0.937	0.942	0.942	0.948
	AvgCov S_0^c		0.954	0.955	0.953	0.952	0.951	0.952	0.952	0.951	0.951	0.950	0.951	0.951	0.952	0.951	0.951
	AvgLength		0.319	0.364	0.511	0.675	0.874	0.202	0.233	0.333	0.452	0.612	0.143	0.165	0.237	0.323	0.444
	AvgLength^c		0.319	0.364	0.511	0.675	0.874	0.202	0.233	0.333	0.452	0.612	0.143	0.165	0.237	0.323	0.443
DEBIASED GLS	AvgCov S_0		0.915	0.906	0.904	0.914	0.918	0.928	0.915	0.916	0.925	0.934	0.924	0.920	0.923	0.923	0.927
	AvgCov S_0^c		0.954	0.956	0.961	0.964	0.967	0.952	0.953	0.953	0.957	0.963	0.951	0.952	0.953	0.955	0.960
	AvgLength		0.317	0.288	0.260	0.257	0.263	0.202	0.182	0.160	0.157	0.160	0.143	0.128	0.112	0.108	0.109
	AvgLength^c		0.317	0.288	0.259	0.257	0.263	0.202	0.182	0.160	0.157	0.160	0.143	0.128	0.112	0.108	0.109
DEBIASED LASSO	AvgCov S_0	200	0.918	0.925	0.939	0.939	0.941	0.930	0.936	0.947	0.951	0.951	0.932	0.938	0.946	0.950	0.950
	AvgCov S_0^c		0.962	0.963	0.959	0.957	0.955	0.955	0.954	0.955	0.954	0.952	0.953	0.953	0.952	0.953	0.952
	AvgLength		0.321	0.367	0.514	0.676	0.870	0.203	0.233	0.334	0.451	0.611	0.143	0.165	0.237	0.323	0.444
	AvgLength^c		0.321	0.366	0.514	0.676	0.870	0.203	0.233	0.333	0.451	0.611	0.143	0.165	0.237	0.323	0.444
DEBIASED GLS	AvgCov S_0		0.915	0.904	0.907	0.908	0.918	0.931	0.927	0.917	0.924	0.927	0.930	0.927	0.915	0.917	0.926
	AvgCov S_0^c		0.962	0.965	0.969	0.972	0.974	0.955	0.957	0.958	0.963	0.967	0.953	0.954	0.955	0.957	0.964
	AvgLength		0.319	0.290	0.263	0.261	0.269	0.202	0.182	0.160	0.160	0.163	0.143	0.128	0.112	0.109	0.112
	AvgLength^c		0.319	0.290	0.263	0.261	0.269	0.202	0.182	0.160	0.160	0.163	0.143	0.128	0.112	0.109	0.112
DEBIASED LASSO	AvgCov S_0	500	0.901	0.911	0.926	0.942	0.941	0.925	0.932	0.939	0.944	0.947	0.929	0.936	0.944	0.944	0.946
	AvgCov S_0^c		0.973	0.974	0.969	0.964	0.962	0.961	0.961	0.962	0.960	0.956	0.958	0.957	0.956	0.957	0.955
	AvgLength		0.324	0.370	0.519	0.685	0.885	0.204	0.235	0.335	0.454	0.612	0.143	0.165	0.237	0.323	0.445
	AvgLength^c		0.324	0.370	0.519	0.685	0.884	0.204	0.235	0.335	0.454	0.611	0.143	0.165	0.237	0.323	0.445
DEBIASED GLS	AvgCov S_0		0.893	0.885	0.872	0.876	0.890	0.924	0.921	0.918	0.925	0.927	0.928	0.925	0.923	0.928	0.933
	AvgCov S_0^c		0.973	0.976	0.979	0.981	0.983	0.961	0.963	0.967	0.971	0.973	0.958	0.959	0.959	0.964	0.968
	AvgLength		0.322	0.295	0.268	0.268	0.281	0.203	0.183	0.163	0.163	0.164	0.143	0.129	0.112	0.111	0.114
	AvgLength^c		0.322	0.294	0.267	0.268	0.281	0.203	0.183	0.163	0.163	0.164	0.143	0.129	0.112	0.111	0.114

Table E.24: See description of Table 3, for $|S_0| = 3$. $d = 16$, DGP 1 ($\boldsymbol{x}_t \sim$ i.i.d. $\mathcal{N}(0,1)$ and $\varepsilon_t \sim t_d$).

			$d=16$														
		p/T	200					500					1000				
ϕ			0	0.5	0.8	0.9	0.95	0	0.5	0.8	0.9	0.95	0	0.5	0.8	0.9	0.95
DEBIASED LASSO	AvgCov S_0	100	0.912	0.921	0.932	0.936	0.939	0.929	0.931	0.936	0.945	0.946	0.930	0.936	0.943	0.948	0.947
	AvgCov S_0^c		0.954	0.954	0.954	0.953	0.951	0.951	0.951	0.951	0.950	0.950	0.951	0.951	0.950	0.951	0.951
	AvgLength		0.296	0.337	0.474	0.626	0.809	0.187	0.216	0.309	0.418	0.566	0.133	0.153	0.219	0.300	0.412
	AvgLength^c		0.295	0.337	0.474	0.625	0.809	0.187	0.216	0.308	0.418	0.566	0.132	0.153	0.219	0.299	0.412
DEBIASED GLS	AvgCov S_0		0.909	0.902	0.901	0.908	0.920	0.928	0.928	0.923	0.923	0.932	0.931	0.928	0.923	0.924	0.931
	AvgCov S_0^c		0.953	0.956	0.960	0.963	0.967	0.951	0.953	0.954	0.958	0.964	0.950	0.951	0.952	0.954	0.959
	AvgLength		0.294	0.267	0.240	0.239	0.246	0.187	0.168	0.148	0.145	0.149	0.132	0.119	0.104	0.100	0.101
	AvgLength^c		0.294	0.267	0.239	0.239	0.245	0.187	0.168	0.148	0.145	0.149	0.132	0.119	0.104	0.100	0.101
DEBIASED LASSO	AvgCov S_0	200	0.911	0.921	0.933	0.934	0.941	0.926	0.928	0.929	0.939	0.942	0.932	0.927	0.939	0.935	0.939
	AvgCov S_0^c		0.963	0.963	0.961	0.958	0.956	0.955	0.954	0.955	0.955	0.953	0.953	0.953	0.952	0.952	0.952
	AvgLength		0.298	0.340	0.476	0.627	0.807	0.188	0.216	0.309	0.418	0.566	0.133	0.153	0.219	0.299	0.412
	AvgLength^c		0.298	0.340	0.476	0.627	0.807	0.188	0.216	0.309	0.418	0.566	0.133	0.153	0.219	0.299	0.412
DEBIASED GLS	AvgCov S_0		0.905	0.896	0.894	0.900	0.911	0.923	0.925	0.919	0.924	0.935	0.932	0.925	0.914	0.915	0.931
	AvgCov S_0^c		0.963	0.965	0.969	0.971	0.974	0.955	0.957	0.958	0.962	0.967	0.953	0.955	0.956	0.957	0.964
	AvgLength		0.296	0.269	0.244	0.244	0.251	0.187	0.169	0.148	0.148	0.151	0.132	0.119	0.104	0.101	0.103
	AvgLength^c		0.296	0.269	0.244	0.243	0.251	0.187	0.169	0.148	0.148	0.151	0.132	0.119	0.104	0.101	0.103
DEBIASED LASSO	AvgCov S_0	500	0.900	0.905	0.919	0.934	0.943	0.929	0.930	0.938	0.940	0.946	0.926	0.931	0.941	0.946	0.954
	AvgCov S_0^c		0.974	0.974	0.970	0.965	0.963	0.962	0.961	0.963	0.961	0.958	0.959	0.958	0.956	0.958	0.956
	AvgLength		0.301	0.343	0.482	0.636	0.823	0.189	0.217	0.311	0.420	0.567	0.133	0.153	0.220	0.300	0.413
	AvgLength^c		0.301	0.343	0.482	0.636	0.822	0.189	0.217	0.311	0.420	0.567	0.133	0.153	0.220	0.300	0.413
DEBIASED GLS	AvgCov S_0		0.892	0.881	0.877	0.878	0.889	0.927	0.922	0.909	0.920	0.927	0.926	0.914	0.905	0.909	0.925
	AvgCov S_0^c		0.974	0.976	0.979	0.981	0.983	0.962	0.963	0.966	0.971	0.973	0.958	0.960	0.960	0.964	0.970
	AvgLength		0.299	0.273	0.249	0.250	0.262	0.188	0.170	0.150	0.152	0.153	0.133	0.119	0.104	0.102	0.106
	AvgLength^c		0.298	0.273	0.249	0.250	0.262	0.188	0.170	0.150	0.152	0.153	0.133	0.119	0.104	0.102	0.106

Table E.25: See description of Table 3, for $|S_0| = 7$. $d = 4$, DGP 1 ($\boldsymbol{x}_t \sim$ i.i.d. $\mathcal{N}(0, 1)$ and $\varepsilon_t \sim t_d$).

		$d=4$															
		p/T	200					500					1000				
ϕ			0	0.5	0.8	0.9	0.95	0	0.5	0.8	0.9	0.95	0	0.5	0.8	0.9	0.95
DEBIASED LASSO	AvgCov S_0	100	0.927	0.929	0.934	0.941	0.943	0.933	0.939	0.944	0.948	0.946	0.937	0.942	0.943	0.946	0.946
	AvgCov S_0^c		0.958	0.957	0.956	0.953	0.952	0.955	0.954	0.955	0.954	0.953	0.952	0.951	0.951	0.952	0.953
	AvgLength		0.400	0.457	0.640	0.841	1.083	0.250	0.288	0.412	0.558	0.755	0.176	0.203	0.291	0.397	0.546
	AvgLength^c		0.400	0.457	0.639	0.841	1.082	0.250	0.288	0.412	0.558	0.754	0.176	0.203	0.291	0.397	0.546
DEBIASED GLS	AvgCov S_0		0.923	0.922	0.922	0.930	0.935	0.932	0.930	0.930	0.936	0.946	0.937	0.936	0.936	0.938	0.947
	AvgCov S_0^c		0.957	0.958	0.963	0.967	0.970	0.955	0.954	0.955	0.962	0.968	0.952	0.952	0.953	0.957	0.963
	AvgLength		0.398	0.365	0.334	0.337	0.348	0.250	0.225	0.199	0.199	0.206	0.176	0.158	0.139	0.135	0.139
	AvgLength^c		0.398	0.364	0.333	0.337	0.348	0.250	0.225	0.199	0.199	0.206	0.176	0.158	0.139	0.135	0.139
DEBIASED LASSO	AvgCov S_0	200	0.922	0.929	0.937	0.941	0.948	0.933	0.938	0.945	0.941	0.942	0.941	0.941	0.947	0.948	0.949
	AvgCov S_0^c		0.967	0.968	0.965	0.960	0.957	0.957	0.956	0.958	0.956	0.954	0.955	0.955	0.954	0.954	0.953
	AvgLength		0.407	0.464	0.645	0.843	1.082	0.251	0.289	0.412	0.558	0.754	0.177	0.204	0.292	0.399	0.549
	AvgLength^c		0.407	0.464	0.645	0.843	1.081	0.251	0.289	0.412	0.558	0.754	0.177	0.204	0.292	0.399	0.549
DEBIASED GLS	AvgCov S_0		0.920	0.916	0.917	0.924	0.931	0.932	0.932	0.927	0.938	0.944	0.939	0.939	0.936	0.941	0.952
	AvgCov S_0^c		0.967	0.969	0.974	0.977	0.979	0.957	0.959	0.961	0.968	0.972	0.954	0.956	0.957	0.961	0.968
	AvgLength		0.404	0.373	0.346	0.349	0.359	0.251	0.227	0.202	0.205	0.212	0.176	0.159	0.140	0.137	0.143
	AvgLength^c		0.404	0.373	0.346	0.348	0.359	0.251	0.227	0.202	0.205	0.212	0.176	0.159	0.140	0.137	0.143
DEBIASED LASSO	AvgCov S_0	500	0.907	0.911	0.930	0.940	0.945	0.930	0.936	0.938	0.941	0.946	0.928	0.934	0.944	0.949	0.951
	AvgCov S_0^c		0.981	0.980	0.973	0.965	0.963	0.965	0.965	0.967	0.963	0.960	0.961	0.959	0.958	0.960	0.958
	AvgLength		0.416	0.474	0.659	0.862	1.107	0.254	0.293	0.418	0.564	0.759	0.178	0.205	0.294	0.401	0.552
	AvgLength^c		0.416	0.474	0.659	0.862	1.106	0.254	0.293	0.418	0.564	0.759	0.178	0.205	0.294	0.401	0.552
DEBIASED GLS	AvgCov S_0		0.903	0.891	0.888	0.892	0.908	0.928	0.917	0.922	0.933	0.936	0.928	0.927	0.921	0.930	0.941
	AvgCov S_0^c		0.981	0.981	0.984	0.986	0.987	0.965	0.967	0.972	0.976	0.978	0.961	0.963	0.963	0.969	0.974
	AvgLength		0.413	0.383	0.358	0.363	0.378	0.253	0.229	0.208	0.213	0.217	0.177	0.160	0.141	0.142	0.148
	AvgLength^c		0.413	0.383	0.358	0.363	0.378	0.253	0.229	0.208	0.212	0.217	0.177	0.160	0.141	0.142	0.148

Table E.26: See description of Table 3, for $|S_0| = 7$. $d = 8$, DGP 1 ($\boldsymbol{x}_t \sim$ i.i.d. $\mathcal{N}(0, 1)$ and $\varepsilon_t \sim t_d$).

		$d=8$															
		p/T	200					500					1000				
ϕ			0	0.5	0.8	0.9	0.95	0	0.5	0.8	0.9	0.95	0	0.5	0.8	0.9	0.95
DEBIASED LASSO	AvgCov S_0	100	0.918	0.922	0.930	0.934	0.938	0.927	0.932	0.940	0.945	0.946	0.928	0.933	0.946	0.948	0.947
	AvgCov S_0^c		0.957	0.957	0.957	0.955	0.952	0.955	0.953	0.953	0.952	0.952	0.951	0.951	0.950	0.951	0.952
	AvgLength		0.329	0.375	0.526	0.692	0.892	0.205	0.236	0.337	0.456	0.618	0.144	0.166	0.238	0.325	0.447
	AvgLength^c		0.329	0.375	0.526	0.692	0.891	0.205	0.236	0.337	0.456	0.618	0.144	0.166	0.238	0.325	0.447
DEBIASED GLS	AvgCov S_0		0.916	0.909	0.907	0.919	0.930	0.928	0.925	0.923	0.927	0.937	0.927	0.925	0.919	0.922	0.930
	AvgCov S_0^c		0.956	0.957	0.961	0.967	0.971	0.954	0.955	0.957	0.962	0.968	0.951	0.952	0.955	0.959	0.964
	AvgLength		0.327	0.300	0.273	0.279	0.290	0.204	0.184	0.164	0.163	0.170	0.144	0.129	0.114	0.112	0.113
	AvgLength^c		0.327	0.300	0.273	0.279	0.290	0.204	0.184	0.164	0.163	0.170	0.144	0.129	0.114	0.112	0.113
DEBIASED LASSO	AvgCov S_0	200	0.913	0.921	0.934	0.936	0.939	0.930	0.931	0.941	0.944	0.944	0.934	0.936	0.941	0.942	0.944
	AvgCov S_0^c		0.967	0.967	0.967	0.964	0.961	0.958	0.956	0.956	0.957	0.955	0.954	0.954	0.953	0.954	0.955
	AvgLength		0.335	0.382	0.533	0.697	0.893	0.206	0.237	0.338	0.457	0.618	0.144	0.166	0.239	0.326	0.448
	AvgLength^c		0.335	0.382	0.532	0.697	0.893	0.206	0.237	0.338	0.457	0.618	0.144	0.166	0.239	0.326	0.448
DEBIASED GLS	AvgCov S_0		0.911	0.904	0.902	0.913	0.923	0.928	0.921	0.914	0.924	0.936	0.933	0.929	0.926	0.927	0.939
	AvgCov S_0^c		0.966	0.968	0.972	0.976	0.980	0.958	0.960	0.962	0.967	0.973	0.954	0.955	0.959	0.961	0.968
	AvgLength		0.333	0.307	0.284	0.290	0.303	0.205	0.186	0.166	0.168	0.177	0.144	0.130	0.116	0.112	0.116
	AvgLength^c		0.333	0.307	0.284	0.289	0.303	0.205	0.186	0.166	0.167	0.177	0.144	0.130	0.116	0.112	0.116
DEBIASED LASSO	AvgCov S_0	500	0.902	0.911	0.922	0.936	0.946	0.919	0.920	0.939	0.939	0.943	0.929	0.929	0.935	0.941	0.947
	AvgCov S_0^c		0.981	0.981	0.979	0.973	0.968	0.968	0.967	0.967	0.967	0.963	0.960	0.960	0.958	0.959	0.959
	AvgLength		0.344	0.392	0.547	0.716	0.917	0.208	0.240	0.342	0.462	0.622	0.145	0.167	0.240	0.327	0.451
	AvgLength^c		0.344	0.392	0.547	0.716	0.916	0.208	0.240	0.342	0.462	0.622	0.145	0.167	0.240	0.327	0.451
DEBIASED GLS	AvgCov S_0		0.896	0.884	0.884	0.897	0.911	0.917	0.907	0.900	0.912	0.925	0.928	0.924	0.922	0.925	0.937
	AvgCov S_0^c		0.981	0.981	0.985	0.987	0.988	0.968	0.970	0.972	0.978	0.980	0.960	0.963	0.965	0.969	0.975
	AvgLength		0.342	0.317	0.299	0.305	0.320	0.208	0.189	0.169	0.175	0.182	0.145	0.131	0.117	0.115	0.122
	AvgLength^c		0.342	0.317	0.299	0.305	0.319	0.208	0.189	0.169	0.175	0.182	0.145	0.131	0.117	0.115	0.122

Table E.27: See description of Table 3, for $|S_0| = 7$. $d = 16$, DGP 1 ($\boldsymbol{x}_t \sim$ i.i.d. $\mathcal{N}(0, 1)$ and $\varepsilon_t \sim t_d$).

			$d=16$														
		p/T	200					500					1000				
ϕ			0	0.5	0.8	0.9	0.95	0	0.5	0.8	0.9	0.95	0	0.5	0.8	0.9	0.95
DEBIASED LASSO	AvgCov S_0	100	0.916	0.921	0.931	0.935	0.941	0.922	0.924	0.935	0.938	0.944	0.934	0.939	0.947	0.949	0.952
	AvgCov S_0^c		0.956	0.957	0.957	0.956	0.953	0.954	0.954	0.951	0.951	0.952	0.952	0.953	0.953	0.951	0.952
	AvgLength		0.305	0.349	0.489	0.643	0.830	0.190	0.218	0.312	0.423	0.572	0.133	0.154	0.221	0.301	0.414
	AvgLength^c		0.305	0.348	0.489	0.643	0.830	0.190	0.218	0.312	0.423	0.572	0.133	0.154	0.221	0.301	0.414
DEBIASED GLS	AvgCov S_0		0.914	0.906	0.903	0.914	0.927	0.921	0.918	0.918	0.923	0.934	0.932	0.928	0.925	0.930	0.935
	AvgCov S_0^c		0.955	0.958	0.962	0.967	0.971	0.954	0.954	0.957	0.962	0.968	0.952	0.953	0.955	0.960	0.964
	AvgLength		0.303	0.279	0.255	0.259	0.272	0.189	0.171	0.153	0.151	0.157	0.133	0.119	0.105	0.104	0.105
	AvgLength^c		0.303	0.279	0.255	0.259	0.272	0.189	0.171	0.153	0.151	0.157	0.133	0.119	0.105	0.104	0.105
DEBIASED LASSO	AvgCov S_0	200	0.905	0.914	0.925	0.930	0.937	0.924	0.926	0.942	0.944	0.945	0.932	0.932	0.938	0.939	0.948
	AvgCov S_0^c		0.967	0.967	0.967	0.965	0.962	0.958	0.957	0.957	0.957	0.956	0.955	0.955	0.954	0.953	0.954
	AvgLength		0.310	0.354	0.494	0.647	0.829	0.191	0.220	0.314	0.425	0.573	0.134	0.154	0.221	0.302	0.415
	AvgLength^c		0.310	0.354	0.494	0.647	0.829	0.191	0.220	0.314	0.424	0.573	0.134	0.154	0.221	0.302	0.415
DEBIASED GLS	AvgCov S_0		0.900	0.893	0.893	0.907	0.915	0.924	0.918	0.911	0.921	0.932	0.932	0.926	0.923	0.926	0.930
	AvgCov S_0^c		0.966	0.967	0.972	0.977	0.979	0.958	0.960	0.963	0.968	0.973	0.954	0.955	0.959	0.962	0.968
	AvgLength		0.309	0.285	0.263	0.270	0.282	0.191	0.173	0.155	0.155	0.164	0.134	0.120	0.107	0.105	0.107
	AvgLength^c		0.308	0.285	0.263	0.270	0.282	0.191	0.173	0.155	0.155	0.164	0.134	0.120	0.107	0.105	0.107
DEBIASED LASSO	AvgCov S_0	500	0.898	0.910	0.918	0.930	0.938	0.918	0.923	0.940	0.937	0.943	0.928	0.933	0.938	0.944	0.943
	AvgCov S_0^c		0.981	0.981	0.981	0.975	0.971	0.969	0.967	0.966	0.967	0.965	0.961	0.961	0.959	0.959	0.960
	AvgLength		0.320	0.364	0.507	0.665	0.852	0.193	0.222	0.317	0.428	0.576	0.134	0.155	0.222	0.303	0.417
	AvgLength^c		0.319	0.364	0.507	0.665	0.852	0.193	0.222	0.317	0.428	0.576	0.134	0.155	0.222	0.303	0.417
DEBIASED GLS	AvgCov S_0		0.892	0.879	0.877	0.890	0.901	0.917	0.902	0.894	0.912	0.924	0.926	0.920	0.921	0.922	0.938
	AvgCov S_0^c		0.981	0.982	0.985	0.987	0.988	0.969	0.971	0.973	0.977	0.980	0.960	0.964	0.968	0.969	0.975
	AvgLength		0.317	0.295	0.278	0.285	0.300	0.193	0.176	0.156	0.162	0.170	0.134	0.121	0.109	0.106	0.112
	AvgLength^c		0.317	0.294	0.278	0.285	0.299	0.193	0.176	0.157	0.162	0.170	0.134	0.121	0.109	0.106	0.112

Table E.28: See description of Table 1, for $|S_0| = 3,\ d = 4$, DGP 2 ($\boldsymbol{x}_t \sim t_d$ and $\varepsilon_t \sim t_d$).

		$d=4$														
Panel I																
	p/T	200					500					1000				
ϕ		0	0.5	0.8	0.9	0.95	0	0.5	0.8	0.9	0.95	0	0.5	0.8	0.9	0.95
LASSO/ GLS LASSO	100	0.994	1.043	1.398	1.678	1.946	0.997	1.031	1.089	1.335	1.704	0.998	1.006	1.049	1.067	1.340
GLS LASSO		(0.05)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)
LASSO/ GLS LASSO	200	0.994	1.072	1.423	1.745	2.030	0.995	1.002	1.113	1.423	1.731	1.000	1.015	1.011	1.065	1.498
GLS LASSO		(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
LASSO/ GLS LASSO	500	0.991	1.100	1.507	1.753	2.115	0.998	0.956	1.214	1.609	1.836	0.996	1.012	0.928	1.171	1.633
GLS LASSO		(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Panel II																
DEBIASED LASSO/ DEBIASED GLS	100	1.004	1.231	1.860	2.484	3.151	1.001	1.203	1.749	2.322	3.089	1.000	1.156	1.584	2.079	2.759
DEBIASED GLS		(0.10)	(0.09)	(0.08)	(0.08)	(0.08)	(0.07)	(0.07)	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)	(0.05)	(0.05)	(0.05)
DEBIASED LASSO/ DEBIASED GLS	200	1.004	1.242	1.902	2.536	2.334	1.002	1.238	1.871	2.540	3.412	1.001	1.197	1.748	2.368	3.179
DEBIASED GLS		(0.10)	(0.09)	(0.08)	(0.08)	(0.20)	(0.07)	(0.06)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.04)	(0.04)	(0.04)
DEBIASED LASSO/ DEBIASED GLS	500	0.984	0.889	1.299	0.259	2.859	1.001	1.279	2.005	2.740	3.760	1.001	1.246	1.957	2.706	3.648
DEBIASED GLS		(0.23)	(0.21)	(1.20)	(1.02)	(0.15)	(0.06)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.04)	(0.04)	(0.04)	(0.04)

Table E.29: See description of Table 1, for $|S_0| = 3,\ d = 8$, DGP 2 ($\boldsymbol{x}_t \sim t_d$ and $\varepsilon_t \sim t_d$).

		$d=8$														
Panel I																
	p/T	200					500					1000				
ϕ		0	0.5	0.8	0.9	0.95	0	0.5	0.8	0.9	0.95	0	0.5	0.8	0.9	0.95
LASSO/ GLS LASSO	100	0.991	1.119	1.771	2.311	2.777	0.994	1.132	1.413	2.059	2.924	0.995	1.109	1.495	1.620	2.474
GLS LASSO		(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
LASSO/ GLS LASSO	200	0.985	1.155	1.843	2.344	2.806	0.988	1.096	1.406	2.155	2.951	0.997	1.157	1.314	1.606	2.659
GLS LASSO		(0.03)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
LASSO/ GLS LASSO	500	0.996	1.223	1.904	2.339	2.797	0.990	1.048	1.649	2.548	3.027	0.998	1.050	1.092	1.787	2.957
GLS LASSO		(0.02)	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Panel II																
DEBIASED LASSO/ DEBIASED GLS	100	1.004	1.264	1.974	2.662	3.404	1.001	1.269	2.008	2.776	3.747	1.001	1.243	1.922	2.677	3.671
DEBIASED GLS		(0.08)	(0.07)	(0.06)	(0.06)	(0.06)	(0.05)	(0.05)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.03)	(0.03)	(0.03)
DEBIASED LASSO/ DEBIASED GLS	200	1.006	1.271	1.996	2.702	3.433	1.002	1.283	2.064	2.846	3.849	1.001	1.267	2.018	2.841	3.886
DEBIASED GLS		(0.07)	(0.07)	(0.06)	(0.06)	(0.06)	(0.05)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.03)	(0.03)	(0.03)	(0.03)
DEBIASED LASSO/ DEBIASED GLS	500	1.004	1.260	1.982	1.164	3.418	1.003	1.304	2.093	2.862	3.940	1.001	1.289	2.124	2.984	4.070
DEBIASED GLS		(0.07)	(0.06)	(0.05)	(0.12)	(0.05)	(0.05)	(0.04)	(0.04)	(0.04)	(0.04)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)

Table E.30: See description of Table 1, for $|S_0| = 3,\ d = 16$, DGP 2 ($\boldsymbol{x}_t \sim t_d$ and $\varepsilon_t \sim t_d$).

								$d = 16$								
Panel I																
	p/T	200					500					1000				
ϕ		0	0.5	0.8	0.9	0.95	0	0.5	0.8	0.9	0.95	0	0.5	0.8	0.9	0.95
LASSO/ GLS LASSO	100	0.980	1.153	1.811	2.448	2.911	0.996	1.216	1.639	2.334	3.334	0.998	1.252	1.936	2.325	3.435
GLS LASSO		(0.03)	(0.03)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
LASSO/ GLS LASSO	200	0.981	1.180	1.914	2.444	2.878	0.995	1.167	1.568	2.434	3.342	0.998	1.247	1.680	2.078	3.474
GLS LASSO		(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
LASSO/ GLS LASSO	500	0.995	1.243	1.927	2.361	2.806	0.991	1.102	1.853	2.656	3.270	0.997	1.220	1.363	2.109	3.464
GLS LASSO		(0.02)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Panel II																
DEBIASED LASSO/ DEBIASED GLS	100	1.005	1.272	2.005	2.710	3.461	1.001	1.283	2.078	2.902	3.927	1.001	1.278	2.065	2.929	4.058
DEBIASED GLS		(0.07)	(0.06)	(0.06)	(0.06)	(0.06)	(0.05)	(0.04)	(0.04)	(0.04)	(0.04)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
DEBIASED LASSO/ DEBIASED GLS	200	1.005	1.273	2.000	2.705	3.444	1.002	1.296	2.112	2.928	3.954	1.001	1.287	2.106	2.997	4.123
DEBIASED GLS		(0.07)	(0.06)	(0.05)	(0.05)	(0.05)	(0.05)	(0.04)	(0.04)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.02)	(0.02)
DEBIASED LASSO/ DEBIASED GLS	500	1.006	1.207	1.220	2.174	3.421	1.002	1.312	2.109	2.908	3.988	1.001	1.299	2.170	3.066	4.171
DEBIASED GLS		(0.06)	(0.06)	(0.09)	(0.07)	(0.05)	(0.04)	(0.04)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.02)	(0.02)	(0.02)

Table E.31: See description of Table 1, for $|S_0| = 7,\ d = 4$, DGP 2 ($\boldsymbol{x}_t \sim t_d$ and $\varepsilon_t \sim t_d$).

								$d = 4$								
Panel I																
	p/T	200					500					1000				
ϕ		0	0.5	0.8	0.9	0.95	0	0.5	0.8	0.9	0.95	0	0.5	0.8	0.9	0.95
LASSO/ GLS LASSO	100	0.994	1.043	1.398	1.678	1.946	0.997	1.031	1.089	1.335	1.704	0.998	1.006	1.049	1.067	1.340
GLS LASSO		(0.05)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)
LASSO/ GLS LASSO	200	0.994	1.072	1.423	1.745	2.030	0.995	1.002	1.113	1.423	1.731	1.000	1.015	1.011	1.065	1.498
GLS LASSO		(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
LASSO/ GLS LASSO	500	0.991	1.100	1.507	1.753	2.115	0.998	0.956	1.214	1.609	1.836	0.996	1.012	0.928	1.171	1.633
GLS LASSO		(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Panel II																
DEBIASED LASSO/ DEBIASED GLS	100	1.002	1.190	1.695	2.172	2.713	1.001	1.140	1.529	1.953	2.512	1.000	1.093	1.349	1.665	2.120
DEBIASED GLS		(0.11)	(0.11)	(0.10)	(0.10)	(0.10)	(0.09)	(0.08)	(0.08)	(0.08)	(0.08)	(0.08)	(0.07)	(0.07)	(0.07)	(0.07)
DEBIASED LASSO/ DEBIASED GLS	200	1.001	1.200	1.775	2.337	2.967	1.001	1.183	1.690	2.217	2.894	1.000	1.139	1.516	1.964	2.574
DEBIASED GLS		(0.10)	(0.09)	(0.09)	(0.08)	(0.08)	(0.07)	(0.07)	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)	(0.05)	(0.05)
DEBIASED LASSO/ DEBIASED GLS	500	1.002	0.068	1.459	1.573	2.586	1.002	1.234	1.924	2.554	3.369	1.001	1.193	1.727	2.387	3.139
DEBIASED GLS		(0.15)	(3.29)	(0.16)	(0.22)	(0.13)	(0.06)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.04)	(0.04)	(0.04)

Table E.32: See description of Table 1, for $|S_0| = 7,\ d = 8$, DGP 2 ($\boldsymbol{x}_t \sim t_d$ and $\varepsilon_t \sim t_d$).

								$d = 8$								
Panel I																
	p/T	200					500					1000				
ϕ		0	0.5	0.8	0.9	0.95	0	0.5	0.8	0.9	0.95	0	0.5	0.8	0.9	0.95
LASSO/ GLS LASSO	100	0.991	1.119	1.771	2.311	2.777	0.994	1.132	1.413	2.059	2.924	0.995	1.109	1.495	1.620	2.474
GLS LASSO		(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
LASSO/ GLS LASSO	200	0.985	1.155	1.843	2.344	2.806	0.988	1.096	1.406	2.155	2.951	0.997	1.157	1.314	1.606	2.659
GLS LASSO		(0.03)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
LASSO/ GLS LASSO	500	0.996	1.223	1.904	2.339	2.797	0.990	1.048	1.649	2.548	3.027	0.998	1.050	1.092	1.787	2.957
GLS LASSO		(0.02)	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Panel II																
DEBIASED LASSO/ DEBIASED GLS	100	1.002	1.243	1.893	2.490	3.130	1.001	1.234	1.888	2.562	3.397	1.000	1.197	1.747	2.364	3.192
DEBIASED GLS		(0.08)	(0.07)	(0.07)	(0.07)	(0.07)	(0.06)	(0.05)	(0.05)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)
DEBIASED LASSO/ DEBIASED GLS	200	1.005	1.247	1.914	2.518	3.179	1.001	1.261	1.994	2.722	3.572	1.001	1.236	1.891	2.615	3.556
DEBIASED GLS		(0.07)	(0.07)	(0.06)	(0.06)	(0.06)	(0.05)	(0.05)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.03)	(0.03)	(0.03)
DEBIASED LASSO/ DEBIASED GLS	500	1.007	1.236	1.806	2.527	3.224	1.001	1.285	2.154	2.869	3.737	1.001	1.266	2.010	2.889	3.832
DEBIASED GLS		(0.06)	(0.06)	(0.06)	(0.06)	(0.06)	(0.05)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.03)	(0.03)	(0.03)	(0.03)

Table E.33: See description of Table 1, for $|S_0| = 7,\ d = 16$, DGP 2 ($\boldsymbol{x}_t \sim t_d$ and $\varepsilon_t \sim t_d$).

								$d = 16$								
Panel I																
	p/T			200					500					1000		
ϕ		0	0.5	0.8	0.9	0.95	0	0.5	0.8	0.9	0.95	0	0.5	0.8	0.9	0.95
LASSO/ GLS LASSO	100	0.980	1.153	1.811	2.448	2.911	0.996	1.216	1.639	2.334	3.334	0.998	1.252	1.936	2.325	3.435
GLS LASSO		(0.03)	(0.03)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
LASSO/ GLS LASSO	200	0.981	1.180	1.914	2.444	2.878	0.995	1.167	1.568	2.434	3.342	0.998	1.247	1.680	2.078	3.474
GLS LASSO		(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
LASSO/ GLS LASSO	500	0.995	1.243	1.927	2.361	2.806	0.991	1.102	1.853	2.656	3.270	0.997	1.220	1.363	2.109	3.464
GLS LASSO		(0.02)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Panel II																
DEBIASED LASSO/ DEBIASED GLS	100	1.003	1.259	1.957	2.578	3.240	1.001	1.272	2.031	2.805	3.742	1.001	1.258	1.985	2.775	3.811
DEBIASED GLS		(0.07)	(0.07)	(0.06)	(0.06)	(0.06)	(0.05)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.03)	(0.03)	(0.03)	(0.03)
DEBIASED LASSO/ DEBIASED GLS	200	1.006	1.259	1.949	2.572	3.221	1.002	1.282	2.086	2.879	3.786	1.001	1.275	2.053	2.900	3.983
DEBIASED GLS		(0.07)	(0.06)	(0.06)	(0.06)	(0.06)	(0.05)	(0.04)	(0.04)	(0.04)	(0.04)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
DEBIASED LASSO/ DEBIASED GLS	500	1.009	1.246	1.893	2.505	3.169	1.002	1.293	2.192	2.926	3.807	1.001	1.290	2.097	3.043	4.042
DEBIASED GLS		(0.06)	(0.06)	(0.05)	(0.05)	(0.05)	(0.04)	(0.04)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.02)	(0.02)

Table E.34: See description of Table 2, for $|S_0| = 3,\ d = 4$, DGP 2 ($\boldsymbol{x}_t \sim t_d$ and $\varepsilon_t \sim t_d$).

								Size adjusted power and standard errors								
								$d = 4$								
	p/T			200					500					1000		
ϕ		0	0.5	0.8	0.9	0.95	0	0.5	0.8	0.9	0.95	0	0.5	0.8	0.9	0.95
GLS LASSO	100	0.826	0.850	0.868	0.875	0.875	0.904	0.913	0.914	0.919	0.923	0.919	0.923	0.935	0.939	0.938
(s.e.)		(0.10)	(0.09)	(0.08)	(0.08)	(0.09)	(0.06)	(0.06)	(0.05)	(0.05)	(0.05)	(0.04)	(0.04)	(0.04)	(0.03)	(0.03)
LASSO		0.830	0.793	0.727	0.662	0.578	0.902	0.870	0.806	0.731	0.649	0.919	0.903	0.875	0.836	0.765
(s.e.)		(0.10)	(0.12)	(0.17)	(0.22)	(0.28)	(0.06)	(0.07)	(0.10)	(0.14)	(0.19)	(0.04)	(0.05)	(0.07)	(0.10)	(0.14)
GLS LASSO	200	0.850	0.857	0.872	0.875	0.881	0.911	0.922	0.936	0.938	0.936	0.930	0.943	0.952	0.956	0.956
(s.e.)		(0.10)	(0.09)	(0.09)	(0.09)	(0.09)	(0.06)	(0.06)	(0.05)	(0.05)	(0.05)	(0.04)	(0.04)	(0.04)	(0.03)	(0.03)
LASSO		0.848	0.823	0.754	0.688	0.579	0.911	0.893	0.835	0.787	0.737	0.930	0.927	0.895	0.855	0.791
(s.e.)		(0.10)	(0.12)	(0.17)	(0.22)	(0.28)	(0.06)	(0.07)	(0.11)	(0.14)	(0.19)	(0.04)	(0.05)	(0.07)	(0.10)	(0.14)
GLS LASSO	500	0.838	0.853	0.867	0.870	0.870	0.911	0.919	0.929	0.936	0.934	0.931	0.941	0.949	0.950	0.953
(s.e.)		(0.16)	(0.12)	(0.11)	(0.11)	(0.10)	(0.06)	(0.06)	(0.05)	(0.05)	(0.05)	(0.05)	(0.04)	(0.04)	(0.03)	(0.04)
LASSO		0.840	0.813	0.772	0.647	0.568	0.911	0.887	0.849	0.796	0.714	0.931	0.924	0.895	0.862	0.805
(s.e.)		(0.12)	(0.13)	(0.18)	(0.24)	(0.32)	(0.06)	(0.07)	(0.11)	(0.14)	(0.20)	(0.05)	(0.05)	(0.07)	(0.10)	(0.14)

Table E.35: See description of Table 2, for $|S_0| = 3,\ d = 8$, DGP 2 ($\boldsymbol{x}_t \sim t_d$ and $\varepsilon_t \sim t_d$).

								Size adjusted power and standard errors								
								$d = 8$								
	p/T			200					500					1000		
ϕ		0	0.5	0.8	0.9	0.95	0	0.5	0.8	0.9	0.95	0	0.5	0.8	0.9	0.95
GLS LASSO	100	0.829	0.844	0.859	0.856	0.869	0.892	0.906	0.910	0.917	0.917	0.927	0.934	0.943	0.943	0.944
(s.e.)		(0.08)	(0.07)	(0.07)	(0.07)	(0.07)	(0.05)	(0.05)	(0.04)	(0.04)	(0.04)	(0.04)	(0.03)	(0.03)	(0.03)	(0.03)
LASSO		0.832	0.814	0.737	0.683	0.599	0.892	0.877	0.831	0.764	0.673	0.928	0.921	0.884	0.840	0.772
(s.e.)		(0.08)	(0.09)	(0.13)	(0.17)	(0.22)	(0.05)	(0.06)	(0.09)	(0.12)	(0.16)	(0.04)	(0.04)	(0.06)	(0.08)	(0.11)
GLS LASSO	200	0.853	0.863	0.879	0.885	0.884	0.907	0.919	0.927	0.930	0.930	0.940	0.949	0.958	0.960	0.962
(s.e.)		(0.08)	(0.07)	(0.07)	(0.07)	(0.07)	(0.05)	(0.05)	(0.04)	(0.04)	(0.04)	(0.04)	(0.03)	(0.03)	(0.03)	(0.03)
LASSO		0.855	0.832	0.765	0.689	0.588	0.908	0.899	0.864	0.815	0.732	0.941	0.928	0.901	0.865	0.815
(s.e.)		(0.08)	(0.09)	(0.13)	(0.17)	(0.22)	(0.05)	(0.06)	(0.09)	(0.12)	(0.16)	(0.04)	(0.04)	(0.06)	(0.08)	(0.11)
GLS LASSO	500	0.845	0.863	0.878	0.879	0.874	0.907	0.913	0.921	0.928	0.924	0.945	0.949	0.949	0.951	0.952
(s.e.)		(0.08)	(0.08)	(0.07)	(0.07)	(0.07)	(0.05)	(0.05)	(0.04)	(0.04)	(0.04)	(0.04)	(0.03)	(0.03)	(0.03)	(0.03)
LASSO		0.846	0.820	0.775	0.718	0.612	0.907	0.891	0.845	0.791	0.719	0.944	0.936	0.897	0.846	0.793
(s.e.)		(0.08)	(0.10)	(0.13)	(0.18)	(0.23)	(0.05)	(0.06)	(0.09)	(0.12)	(0.16)	(0.04)	(0.04)	(0.06)	(0.08)	(0.11)

Table E.36: See description of Table 2, for $|S_0| = 3,\ d = 16$, DGP 2 ($\boldsymbol{x}_t \sim t_d$ and $\varepsilon_t \sim t_d$).

Size adjusted power and standard errors																
$d = 16$																
	p/T			200					500					1000		
ϕ		0	0.5	0.8	0.9	0.95	0	0.5	0.8	0.9	0.95	0	0.5	0.8	0.9	0.95
GLS LASSO	100	0.835	0.848	0.870	0.872	0.880	0.890	0.897	0.912	0.915	0.920	0.910	0.924	0.931	0.937	0.937
(s.e.)		(0.08)	(0.07)	(0.06)	(0.06)	(0.06)	(0.05)	(0.04)	(0.04)	(0.04)	(0.04)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
LASSO		0.835	0.817	0.739	0.667	0.588	0.891	0.878	0.826	0.770	0.712	0.910	0.903	0.871	0.836	0.759
(s.e.)		(0.08)	(0.09)	(0.12)	(0.16)	(0.21)	(0.05)	(0.06)	(0.08)	(0.11)	(0.14)	(0.03)	(0.04)	(0.06)	(0.08)	(0.11)
GLS LASSO	200	0.863	0.874	0.886	0.886	0.897	0.924	0.928	0.931	0.931	0.938	0.936	0.943	0.952	0.955	0.955
(s.e.)		(0.08)	(0.07)	(0.06)	(0.06)	(0.06)	(0.05)	(0.04)	(0.04)	(0.04)	(0.04)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
LASSO		0.865	0.833	0.775	0.691	0.636	0.922	0.909	0.866	0.817	0.737	0.935	0.924	0.898	0.864	0.811
(s.e.)		(0.08)	(0.09)	(0.12)	(0.16)	(0.21)	(0.05)	(0.06)	(0.08)	(0.11)	(0.14)	(0.03)	(0.04)	(0.06)	(0.08)	(0.11)
GLS LASSO	500	0.852	0.865	0.884	0.884	0.881	0.907	0.919	0.932	0.935	0.934	0.935	0.942	0.952	0.955	0.957
(s.e.)		(0.08)	(0.07)	(0.07)	(0.07)	(0.07)	(0.05)	(0.04)	(0.04)	(0.04)	(0.04)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
LASSO		0.852	0.837	0.782	0.703	0.627	0.908	0.894	0.842	0.796	0.720	0.935	0.927	0.892	0.859	0.803
(s.e.)		(0.08)	(0.09)	(0.12)	(0.16)	(0.21)	(0.05)	(0.06)	(0.08)	(0.11)	(0.14)	(0.03)	(0.04)	(0.06)	(0.08)	(0.11)

Table E.37: See description of Table 2, for $|S_0| = 7,\ d = 4$, DGP 2 ($\boldsymbol{x}_t \sim t_d$ and $\varepsilon_t \sim t_d$).

Size adjusted power and standard errors																
$d = 4$																
	p/T			200					500					1000		
ϕ		0	0.5	0.8	0.9	0.95	0	0.5	0.8	0.9	0.95	0	0.5	0.8	0.9	0.95
GLS LASSO	100	0.702	0.719	0.750	0.744	0.742	0.758	0.760	0.762	0.753	0.754	0.745	0.747	0.745	0.752	0.751
(s.e.)		(0.10)	(0.10)	(0.09)	(0.09)	(0.09)	(0.06)	(0.06)	(0.05)	(0.05)	(0.05)	(0.05)	(0.04)	(0.04)	(0.04)	(0.04)
LASSO		0.703	0.717	0.640	0.564	0.509	0.755	0.727	0.671	0.617	0.559	0.745	0.748	0.742	0.717	0.688
(s.e.)		(0.11)	(0.12)	(0.17)	(0.22)	(0.28)	(0.06)	(0.07)	(0.11)	(0.14)	(0.19)	(0.05)	(0.05)	(0.07)	(0.10)	(0.14)
GLS LASSO	200	0.796	0.821	0.838	0.843	0.836	0.893	0.901	0.910	0.917	0.913	0.924	0.932	0.939	0.940	0.941
(s.e.)		(0.11)	(0.10)	(0.09)	(0.09)	(0.10)	(0.06)	(0.06)	(0.05)	(0.05)	(0.06)	(0.05)	(0.04)	(0.04)	(0.04)	(0.04)
LASSO		0.793	0.775	0.701	0.611	0.511	0.894	0.876	0.814	0.761	0.687	0.924	0.922	0.878	0.830	0.776
(s.e.)		(0.11)	(0.12)	(0.17)	(0.22)	(0.29)	(0.06)	(0.07)	(0.11)	(0.14)	(0.20)	(0.05)	(0.05)	(0.08)	(0.10)	(0.14)
GLS LASSO	500	0.819	0.822	0.832	0.837	0.834	0.893	0.895	0.911	0.914	0.913	0.930	0.937	0.945	0.947	0.945
(s.e.)		(0.12)	(0.11)	(0.12)	(0.11)	(0.11)	(0.07)	(0.06)	(0.05)	(0.05)	(0.06)	(0.05)	(0.04)	(0.04)	(0.04)	(0.04)
LASSO		0.815	0.796	0.706	0.617	0.531	0.894	0.877	0.820	0.764	0.678	0.930	0.918	0.878	0.838	0.783
(s.e.)		(0.12)	(0.15)	(0.19)	(0.24)	(0.32)	(0.07)	(0.08)	(0.11)	(0.15)	(0.20)	(0.05)	(0.05)	(0.08)	(0.10)	(0.14)

Table E.38: See description of Table 2, for $|S_0| = 7,\ d = 8$, DGP 2 ($\boldsymbol{x}_t \sim t_d$ and $\varepsilon_t \sim t_d$).

Size adjusted power and standard errors																
$d = 8$																
	p/T			200					500					1000		
ϕ		0	0.5	0.8	0.9	0.95	0	0.5	0.8	0.9	0.95	0	0.5	0.8	0.9	0.95
GLS LASSO	100	0.711	0.716	0.733	0.741	0.726	0.741	0.753	0.759	0.762	0.763	0.744	0.746	0.744	0.750	0.752
(s.e.)		(0.08)	(0.08)	(0.07)	(0.07)	(0.07)	(0.05)	(0.05)	(0.04)	(0.04)	(0.04)	(0.04)	(0.03)	(0.03)	(0.03)	(0.03)
LASSO		0.712	0.710	0.650	0.576	0.492	0.740	0.729	0.723	0.677	0.604	0.744	0.738	0.730	0.724	0.680
(s.e.)		(0.08)	(0.10)	(0.13)	(0.18)	(0.23)	(0.05)	(0.06)	(0.09)	(0.12)	(0.16)	(0.04)	(0.04)	(0.06)	(0.08)	(0.11)
GLS LASSO	200	0.827	0.828	0.850	0.849	0.848	0.893	0.904	0.914	0.915	0.916	0.924	0.933	0.941	0.945	0.945
(s.e.)		(0.09)	(0.08)	(0.07)	(0.08)	(0.08)	(0.05)	(0.05)	(0.04)	(0.04)	(0.04)	(0.04)	(0.03)	(0.03)	(0.03)	(0.03)
LASSO		0.826	0.797	0.714	0.636	0.560	0.893	0.883	0.827	0.764	0.697	0.925	0.913	0.870	0.827	0.775
(s.e.)		(0.09)	(0.10)	(0.14)	(0.18)	(0.23)	(0.05)	(0.06)	(0.09)	(0.12)	(0.16)	(0.04)	(0.04)	(0.06)	(0.08)	(0.11)
GLS LASSO	500	0.830	0.840	0.851	0.857	0.851	0.896	0.907	0.916	0.917	0.919	0.929	0.938	0.944	0.946	0.946
(s.e.)		(0.09)	(0.08)	(0.08)	(0.08)	(0.08)	(0.05)	(0.05)	(0.04)	(0.04)	(0.05)	(0.04)	(0.03)	(0.03)	(0.03)	(0.03)
LASSO		0.833	0.812	0.721	0.658	0.576	0.897	0.881	0.833	0.776	0.701	0.930	0.918	0.880	0.837	0.781
(s.e.)		(0.09)	(0.10)	(0.14)	(0.18)	(0.24)	(0.05)	(0.06)	(0.09)	(0.12)	(0.16)	(0.04)	(0.04)	(0.06)	(0.08)	(0.12)

Table E.39: See description of Table 2, for $|S_0| = 7,\ d = 16$, DGP 2 ($\boldsymbol{x}_t \sim t_d$ and $\varepsilon_t \sim t_d$).

Size adjusted power and standard errors																
								$d = 16$								
	p/T			200					500					1000		
ϕ		0	0.5	0.8	0.9	0.95	0	0.5	0.8	0.9	0.95	0	0.5	0.8	0.9	0.95
GLS LASSO	100	0.688	0.714	0.752	0.759	0.750	0.750	0.754	0.760	0.762	0.759	0.748	0.754	0.759	0.765	0.759
(s.e.)		(0.08)	(0.07)	(0.07)	(0.07)	(0.07)	(0.05)	(0.04)	(0.04)	(0.04)	(0.04)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
LASSO		0.684	0.680	0.636	0.586	0.491	0.751	0.742	0.718	0.688	0.628	0.747	0.751	0.745	0.716	0.689
(s.e.)		(0.08)	(0.09)	(0.12)	(0.16)	(0.21)	(0.05)	(0.06)	(0.08)	(0.11)	(0.15)	(0.03)	(0.04)	(0.06)	(0.08)	(0.11)
GLS LASSO	200	0.824	0.839	0.851	0.857	0.851	0.896	0.905	0.916	0.917	0.914	0.927	0.933	0.940	0.943	0.944
(s.e.)		(0.08)	(0.07)	(0.07)	(0.07)	(0.07)	(0.05)	(0.04)	(0.04)	(0.04)	(0.04)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
LASSO		0.830	0.808	0.724	0.639	0.558	0.895	0.879	0.828	0.776	0.701	0.928	0.915	0.885	0.843	0.780
(s.e.)		(0.08)	(0.09)	(0.13)	(0.17)	(0.21)	(0.05)	(0.06)	(0.08)	(0.11)	(0.15)	(0.03)	(0.04)	(0.06)	(0.08)	(0.11)
GLS LASSO	500	0.839	0.844	0.851	0.855	0.853	0.898	0.907	0.915	0.918	0.918	0.928	0.937	0.944	0.947	0.946
(s.e.)		(0.08)	(0.08)	(0.07)	(0.07)	(0.08)	(0.05)	(0.04)	(0.04)	(0.04)	(0.04)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
LASSO		0.839	0.820	0.730	0.654	0.579	0.898	0.881	0.834	0.783	0.704	0.928	0.917	0.885	0.848	0.792
(s.e.)		(0.08)	(0.09)	(0.13)	(0.17)	(0.22)	(0.05)	(0.06)	(0.08)	(0.11)	(0.15)	(0.03)	(0.04)	(0.06)	(0.08)	(0.11)

Table E.40: See description of Table 3, for $|S_0| = 3$. $d = 4$, DGP 2 ($\boldsymbol{x}_t \sim t_d$ and $\varepsilon_t \sim$ i.i.d. $\mathcal{N}(0, 1)$).

										$d = 4$							
		p/T			200					500					1000		
ϕ			0	0.5	0.8	0.9	0.95	0	0.5	0.8	0.9	0.95	0	0.5	0.8	0.9	0.95
\textscdebiased lasso	AvgCov S_0	100	0.598	0.652	0.761	0.827	0.865	0.354	0.409	0.553	0.685	0.789	0.228	0.265	0.380	0.507	0.645
	AvgCov S_0^c		0.956	0.957	0.955	0.953	0.951	0.953	0.952	0.952	0.954	0.953	0.951	0.952	0.952	0.953	0.953
	AvgLength		0.403	0.461	0.648	0.856	1.107	0.249	0.287	0.410	0.556	0.754	0.176	0.203	0.291	0.398	0.547
	AvgLength^c		0.401	0.459	0.645	0.852	1.101	0.249	0.286	0.410	0.556	0.754	0.176	0.203	0.291	0.398	0.547
\textscdebiased gls	AvgCov S_0		0.597	0.565	0.524	0.525	0.542	0.351	0.322	0.281	0.274	0.279	0.228	0.209	0.183	0.173	0.175
	AvgCov S_0^c		0.955	0.958	0.962	0.965	0.967	0.953	0.953	0.955	0.957	0.963	0.951	0.952	0.953	0.954	0.959
	AvgLength		0.400	0.364	0.328	0.324	0.335	0.248	0.223	0.197	0.192	0.197	0.176	0.158	0.138	0.133	0.134
	AvgLength^c		0.399	0.363	0.327	0.323	0.334	0.248	0.223	0.197	0.191	0.197	0.176	0.158	0.138	0.133	0.134
\textscdebiased lasso	AvgCov S_0	200	0.620	0.676	0.788	0.853	0.891	0.357	0.413	0.562	0.702	0.795	0.224	0.263	0.383	0.520	0.664
	AvgCov S_0^c		0.963	0.963	0.960	0.955	0.954	0.957	0.957	0.958	0.957	0.955	0.956	0.956	0.954	0.955	0.955
	AvgLength		0.405	0.463	0.649	0.853	1.098	0.251	0.289	0.413	0.559	0.755	0.176	0.203	0.291	0.398	0.547
	AvgLength^c		0.404	0.463	0.648	0.853	1.097	0.251	0.289	0.413	0.558	0.754	0.176	0.203	0.291	0.398	0.547
\textscdebiased gls	AvgCov S_0		0.621	0.591	0.549	0.547	0.555	0.360	0.324	0.285	0.280	0.287	0.225	0.205	0.179	0.174	0.177
	AvgCov S_0^c		0.963	0.966	0.969	0.971	0.974	0.957	0.958	0.960	0.964	0.969	0.956	0.958	0.959	0.960	0.965
	AvgLength		0.403	0.369	0.345	0.334	0.342	0.250	0.225	0.199	0.196	0.202	0.176	0.158	0.139	0.134	0.137
	AvgLength^c		0.403	0.368	0.336	0.337	0.374	0.250	0.225	0.199	0.196	0.202	0.176	0.158	0.139	0.134	0.137
\textscdebiased lasso	AvgCov S_0	500	0.673	0.726	0.835	0.885	0.912	0.395	0.450	0.604	0.731	0.824	0.247	0.281	0.398	0.525	0.682
	AvgCov S_0^c		0.975	0.974	0.969	0.962	0.961	0.967	0.965	0.966	0.965	0.961	0.962	0.962	0.960	0.961	0.960
	AvgLength		0.459	0.511	0.714	0.938	1.237	0.254	0.292	0.417	0.565	0.765	0.177	0.204	0.292	0.399	0.548
	AvgLength^c		0.565	0.746	0.851	1.137	1.336	0.254	0.292	0.417	0.565	0.764	0.177	0.204	0.292	0.398	0.548
\textscdebiased gls	AvgCov S_0		0.671	0.641	0.613	0.620	0.635	0.399	0.362	0.322	0.319	0.322	0.247	0.230	0.203	0.198	0.203
	AvgCov S_0^c		0.975	0.977	0.979	0.981	0.982	0.966	0.968	0.970	0.974	0.976	0.962	0.964	0.965	0.967	0.972
	AvgLength		0.609	0.457	0.421	0.448	0.399	0.253	0.228	0.202	0.203	0.208	0.177	0.159	0.140	0.136	0.141
	AvgLength^c		0.503	0.490	0.682	0.681	0.455	0.253	0.228	0.202	0.203	0.208	0.177	0.159	0.140	0.136	0.141

Table E.41: See description of Table 3, for $|S_0| = 3$. $d = 8$, DGP 2 ($\boldsymbol{x}_t \sim t_d$ and $\varepsilon_t \sim$ i.i.d. $\mathcal{N}(0, 1)$).

			$d = 8$														
		p/T	200					500					1000				
ϕ			0	0.5	0.8	0.9	0.95	0	0.5	0.8	0.9	0.95	0	0.5	0.8	0.9	0.95
DEBIASED LASSO	AvgCov S_0	100	0.826	0.855	0.899	0.912	0.923	0.659	0.708	0.810	0.877	0.910	0.496	0.564	0.720	0.816	0.880
	AvgCov S_0^c		0.955	0.955	0.956	0.954	0.952	0.953	0.952	0.950	0.952	0.951	0.952	0.951	0.950	0.950	0.951
	AvgLength		0.320	0.365	0.513	0.678	0.877	0.203	0.233	0.334	0.452	0.613	0.143	0.165	0.237	0.323	0.444
	AvgLength^c		0.320	0.365	0.513	0.677	0.877	0.203	0.233	0.333	0.452	0.613	0.143	0.165	0.237	0.323	0.444
DEBIASED GLS	AvgCov S_0		0.827	0.805	0.784	0.791	0.807	0.658	0.622	0.564	0.558	0.575	0.495	0.448	0.392	0.381	0.382
	AvgCov S_0^c		0.954	0.957	0.960	0.965	0.968	0.952	0.955	0.955	0.958	0.964	0.952	0.952	0.953	0.955	0.960
	AvgLength		0.318	0.289	0.259	0.259	0.267	0.202	0.182	0.160	0.156	0.161	0.143	0.128	0.112	0.108	0.109
	AvgLength^c		0.318	0.289	0.259	0.259	0.267	0.202	0.182	0.160	0.156	0.161	0.143	0.128	0.112	0.108	0.109
DEBIASED LASSO	AvgCov S_0	200	0.841	0.864	0.907	0.924	0.936	0.663	0.720	0.824	0.880	0.914	0.480	0.545	0.696	0.795	0.874
	AvgCov S_0^c		0.964	0.964	0.963	0.960	0.958	0.958	0.957	0.956	0.957	0.955	0.955	0.954	0.953	0.953	0.954
	AvgLength		0.323	0.369	0.517	0.680	0.877	0.203	0.234	0.334	0.452	0.612	0.143	0.165	0.237	0.323	0.444
	AvgLength^c		0.323	0.368	0.517	0.680	0.876	0.203	0.234	0.334	0.452	0.612	0.143	0.165	0.237	0.323	0.444
DEBIASED GLS	AvgCov S_0		0.837	0.820	0.801	0.809	0.819	0.665	0.622	0.571	0.572	0.587	0.482	0.442	0.385	0.373	0.385
	AvgCov S_0^c		0.963	0.965	0.969	0.973	0.975	0.957	0.959	0.960	0.964	0.969	0.955	0.955	0.956	0.958	0.964
	AvgLength		0.321	0.292	0.265	0.266	0.274	0.203	0.183	0.161	0.159	0.165	0.143	0.128	0.113	0.109	0.111
	AvgLength^c		0.321	0.292	0.264	0.265	0.274	0.203	0.183	0.161	0.159	0.164	0.143	0.128	0.113	0.109	0.111
DEBIASED LASSO	AvgCov S_0	500	0.857	0.886	0.919	0.934	0.939	0.700	0.751	0.858	0.903	0.928	0.513	0.579	0.729	0.834	0.894
	AvgCov S_0^c		0.974	0.974	0.971	0.967	0.964	0.965	0.963	0.964	0.964	0.960	0.960	0.960	0.957	0.959	0.957
	AvgLength		0.328	0.374	0.524	0.692	0.895	0.204	0.235	0.336	0.454	0.613	0.143	0.165	0.237	0.324	0.446
	AvgLength^c		0.327	0.373	0.524	0.691	0.894	0.204	0.235	0.336	0.454	0.613	0.143	0.165	0.237	0.324	0.446
DEBIASED GLS	AvgCov S_0		0.855	0.844	0.837	0.838	0.854	0.701	0.664	0.621	0.636	0.642	0.511	0.472	0.423	0.418	0.425
	AvgCov S_0^c		0.974	0.976	0.979	0.982	0.984	0.965	0.966	0.968	0.973	0.975	0.960	0.962	0.962	0.965	0.971
	AvgLength		0.325	0.298	0.272	0.277	0.286	0.204	0.184	0.163	0.164	0.168	0.143	0.129	0.113	0.110	0.115
	AvgLength^c		0.325	0.297	0.272	0.307	0.286	0.204	0.184	0.162	0.164	0.168	0.143	0.129	0.113	0.110	0.115

Table E.42: See description of Table 3, for $|S_0| = 3$. $d = 16$, DGP 2 ($\boldsymbol{x}_t \sim t_d$ and $\varepsilon_t \sim$ i.i.d. $\mathcal{N}(0, 1)$).

			$d = 16$														
		p/T	200					500					1000				
ϕ			0	0.5	0.8	0.9	0.95	0	0.5	0.8	0.9	0.95	0	0.5	0.8	0.9	0.95
DEBIASED LASSO	AvgCov S_0	100	0.897	0.905	0.918	0.934	0.934	0.855	0.879	0.911	0.926	0.934	0.778	0.813	0.888	0.914	0.936
	AvgCov S_0^c		0.954	0.955	0.956	0.953	0.952	0.952	0.951	0.952	0.953	0.952	0.951	0.952	0.951	0.951	0.951
	AvgLength		0.296	0.338	0.475	0.627	0.811	0.187	0.216	0.309	0.418	0.566	0.133	0.153	0.219	0.300	0.412
	AvgLength^c		0.296	0.338	0.475	0.627	0.811	0.187	0.216	0.308	0.418	0.566	0.133	0.153	0.219	0.300	0.412
DEBIASED GLS	AvgCov S_0		0.894	0.892	0.883	0.891	0.905	0.856	0.844	0.820	0.820	0.833	0.776	0.745	0.710	0.699	0.702
	AvgCov S_0^c		0.953	0.956	0.960	0.964	0.968	0.951	0.953	0.954	0.958	0.964	0.951	0.953	0.954	0.956	0.961
	AvgLength		0.294	0.267	0.240	0.240	0.248	0.187	0.168	0.148	0.145	0.149	0.132	0.119	0.104	0.100	0.101
	AvgLength^c		0.294	0.267	0.240	0.240	0.248	0.187	0.168	0.148	0.145	0.149	0.132	0.119	0.104	0.100	0.101
DEBIASED LASSO	AvgCov S_0	200	0.907	0.916	0.931	0.937	0.943	0.863	0.892	0.931	0.940	0.942	0.781	0.821	0.893	0.918	0.928
	AvgCov S_0^c		0.963	0.963	0.963	0.960	0.958	0.957	0.956	0.955	0.955	0.954	0.954	0.954	0.953	0.953	0.953
	AvgLength		0.298	0.340	0.476	0.628	0.809	0.188	0.216	0.309	0.419	0.566	0.133	0.153	0.219	0.300	0.412
	AvgLength^c		0.298	0.340	0.476	0.627	0.808	0.188	0.216	0.309	0.419	0.566	0.133	0.153	0.219	0.300	0.412
DEBIASED GLS	AvgCov S_0		0.903	0.899	0.886	0.899	0.909	0.863	0.853	0.831	0.831	0.850	0.778	0.759	0.716	0.700	0.713
	AvgCov S_0^c		0.962	0.965	0.969	0.973	0.975	0.957	0.958	0.959	0.963	0.968	0.954	0.955	0.956	0.958	0.964
	AvgLength		0.296	0.271	0.245	0.247	0.254	0.188	0.169	0.149	0.148	0.153	0.132	0.119	0.104	0.101	0.103
	AvgLength^c		0.296	0.270	0.245	0.246	0.254	0.188	0.169	0.149	0.148	0.152	0.132	0.119	0.104	0.101	0.103
DEBIASED LASSO	AvgCov S_0	500	0.914	0.918	0.930	0.935	0.945	0.883	0.901	0.929	0.941	0.947	0.793	0.828	0.882	0.913	0.930
	AvgCov S_0^c		0.974	0.975	0.972	0.966	0.964	0.964	0.962	0.963	0.962	0.959	0.959	0.959	0.956	0.957	0.957
	AvgLength		0.302	0.344	0.483	0.638	0.825	0.189	0.217	0.311	0.420	0.567	0.133	0.153	0.220	0.300	0.413
	AvgLength^c		0.301	0.344	0.483	0.637	0.825	0.189	0.217	0.311	0.420	0.567	0.133	0.153	0.220	0.300	0.413
DEBIASED GLS	AvgCov S_0		0.913	0.898	0.904	0.903	0.916	0.884	0.870	0.857	0.864	0.876	0.794	0.772	0.733	0.730	0.745
	AvgCov S_0^c		0.974	0.976	0.980	0.982	0.984	0.964	0.965	0.966	0.972	0.975	0.959	0.961	0.961	0.965	0.971
	AvgLength		0.300	0.277	0.256	0.263	0.265	0.188	0.170	0.150	0.152	0.155	0.133	0.119	0.105	0.102	0.106
	AvgLength^c		0.299	0.279	0.282	0.264	0.265	0.188	0.170	0.150	0.152	0.155	0.133	0.119	0.105	0.102	0.106

Table E.43: See description of Table 3, for $|S_0| = 7$. $d = 4$, DGP 2 ($\boldsymbol{x}_t \sim t_d$ and $\varepsilon_t \sim$ i.i.d. $\mathcal{N}(0,1)$).

			$d=4$														
		p/T			200					500					1000		
ϕ			0	0.5	0.8	0.9	0.95	0	0.5	0.8	0.9	0.95	0	0.5	0.8	0.9	0.95
DEBIASED LASSO	AvgCov S_0	100	0.606	0.661	0.772	0.841	0.879	0.353	0.405	0.559	0.691	0.791	0.232	0.266	0.384	0.519	0.667
	AvgCov S_0^c		0.954	0.954	0.954	0.954	0.951	0.953	0.953	0.951	0.953	0.954	0.951	0.951	0.951	0.951	0.953
	AvgLength		0.412	0.470	0.658	0.866	1.116	0.253	0.291	0.417	0.564	0.764	0.177	0.204	0.293	0.399	0.549
	AvgLength^c		0.412	0.470	0.657	0.865	1.115	0.253	0.291	0.417	0.564	0.763	0.177	0.204	0.293	0.399	0.549
DEBIASED GLS	AvgCov S_0		0.604	0.575	0.548	0.551	0.573	0.353	0.322	0.291	0.288	0.295	0.232	0.206	0.180	0.177	0.180
	AvgCov S_0^c		0.953	0.954	0.959	0.964	0.968	0.952	0.954	0.957	0.960	0.967	0.951	0.952	0.954	0.958	0.963
	AvgLength		0.410	0.378	0.346	0.350	0.366	0.253	0.228	0.204	0.202	0.209	0.177	0.159	0.140	0.138	0.140
	AvgLength^c		0.409	0.377	0.346	0.349	0.366	0.252	0.227	0.204	0.202	0.209	0.177	0.159	0.140	0.138	0.140
DEBIASED LASSO	AvgCov S_0	200	0.646	0.705	0.808	0.864	0.892	0.365	0.421	0.579	0.709	0.806	0.235	0.269	0.387	0.523	0.671
	AvgCov S_0^c		0.967	0.966	0.966	0.962	0.958	0.960	0.959	0.958	0.959	0.958	0.955	0.955	0.955	0.955	0.955
	AvgLength		0.424	0.483	0.672	0.880	1.130	0.254	0.293	0.418	0.566	0.765	0.178	0.205	0.295	0.402	0.553
	AvgLength^c		0.423	0.481	0.671	0.879	1.128	0.254	0.292	0.417	0.565	0.764	0.178	0.205	0.295	0.402	0.553
DEBIASED GLS	AvgCov S_0		0.647	0.629	0.602	0.606	0.627	0.363	0.333	0.303	0.303	0.314	0.235	0.215	0.192	0.189	0.191
	AvgCov S_0^c		0.966	0.967	0.971	0.975	0.977	0.959	0.961	0.965	0.968	0.973	0.955	0.956	0.959	0.964	0.967
	AvgLength		0.422	0.394	0.367	0.370	0.384	0.254	0.230	0.208	0.207	0.216	0.178	0.159	0.142	0.141	0.142
	AvgLength^c		0.421	0.393	0.365	0.369	0.383	0.254	0.230	0.208	0.207	0.216	0.178	0.159	0.142	0.141	0.142
DEBIASED LASSO	AvgCov S_0	500	0.734	0.777	0.865	0.900	0.917	0.418	0.474	0.634	0.753	0.845	0.257	0.305	0.428	0.557	0.705
	AvgCov S_0^c		0.982	0.981	0.976	0.967	0.961	0.972	0.971	0.969	0.970	0.966	0.963	0.963	0.962	0.961	0.963
	AvgLength		0.477	0.586	0.746	0.955	1.235	0.259	0.298	0.425	0.574	0.773	0.180	0.208	0.298	0.407	0.560
	AvgLength^c		0.511	0.595	0.779	1.055	1.475	0.259	0.298	0.425	0.573	0.773	0.180	0.207	0.298	0.406	0.559
DEBIASED GLS	AvgCov S_0		0.733	0.715	0.695	0.703	0.714	0.418	0.383	0.355	0.357	0.370	0.257	0.233	0.216	0.211	0.216
	AvgCov S_0^c		0.982	0.982	0.984	0.986	0.987	0.972	0.973	0.976	0.979	0.981	0.963	0.965	0.969	0.971	0.976
	AvgLength		0.470	0.446	0.462	0.438	0.450	0.259	0.235	0.212	0.215	0.226	0.180	0.162	0.147	0.143	0.149
	AvgLength^c		0.491	1.865	0.507	0.521	0.475	0.258	0.235	0.212	0.215	0.226	0.180	0.162	0.147	0.143	0.148

Table E.44: See description of Table 3, for $|S_0| = 7$. $d = 8$, DGP 2 ($\boldsymbol{x}_t \sim t_d$ and $\varepsilon_t \sim$ i.i.d. $\mathcal{N}(0,1)$).

			$d=8$														
		p/T			200					500					1000		
ϕ			0	0.5	0.8	0.9	0.95	0	0.5	0.8	0.9	0.95	0	0.5	0.8	0.9	0.95
DEBIASED LASSO	AvgCov S_0	100	0.829	0.850	0.896	0.919	0.932	0.671	0.720	0.815	0.878	0.909	0.492	0.559	0.713	0.808	0.875
	AvgCov S_0^c		0.958	0.956	0.956	0.956	0.954	0.955	0.954	0.952	0.953	0.953	0.951	0.952	0.952	0.951	0.951
	AvgLength		0.331	0.378	0.529	0.696	0.898	0.205	0.236	0.337	0.457	0.618	0.144	0.166	0.238	0.325	0.447
	AvgLength^c		0.331	0.377	0.529	0.696	0.898	0.205	0.236	0.337	0.457	0.618	0.144	0.166	0.238	0.325	0.447
DEBIASED GLS	AvgCov S_0		0.824	0.809	0.792	0.806	0.820	0.671	0.629	0.587	0.583	0.596	0.492	0.449	0.390	0.390	0.394
	AvgCov S_0^c		0.956	0.958	0.962	0.967	0.971	0.955	0.956	0.959	0.963	0.969	0.951	0.952	0.954	0.960	0.964
	AvgLength		0.329	0.303	0.277	0.281	0.294	0.205	0.184	0.165	0.163	0.169	0.144	0.129	0.114	0.112	0.114
	AvgLength^c		0.329	0.303	0.277	0.280	0.293	0.205	0.184	0.165	0.163	0.169	0.144	0.129	0.114	0.112	0.114
DEBIASED LASSO	AvgCov S_0	200	0.851	0.874	0.911	0.926	0.938	0.691	0.750	0.845	0.895	0.918	0.494	0.556	0.714	0.815	0.882
	AvgCov S_0^c		0.968	0.967	0.968	0.966	0.962	0.959	0.959	0.957	0.958	0.958	0.955	0.955	0.955	0.953	0.955
	AvgLength		0.337	0.384	0.536	0.702	0.899	0.206	0.237	0.339	0.458	0.620	0.145	0.167	0.239	0.326	0.449
	AvgLength^c		0.337	0.384	0.535	0.702	0.899	0.206	0.237	0.339	0.458	0.620	0.145	0.167	0.239	0.326	0.449
DEBIASED GLS	AvgCov S_0		0.847	0.835	0.827	0.838	0.854	0.691	0.659	0.618	0.616	0.637	0.494	0.446	0.399	0.394	0.397
	AvgCov S_0^c		0.967	0.968	0.972	0.976	0.979	0.959	0.961	0.965	0.968	0.974	0.955	0.956	0.960	0.965	0.968
	AvgLength		0.334	0.311	0.286	0.294	0.308	0.206	0.186	0.168	0.167	0.176	0.144	0.130	0.115	0.114	0.115
	AvgLength^c		0.334	0.310	0.286	0.294	0.308	0.206	0.186	0.168	0.167	0.176	0.144	0.130	0.115	0.114	0.115
DEBIASED LASSO	AvgCov S_0	500	0.882	0.894	0.922	0.933	0.940	0.723	0.780	0.855	0.899	0.931	0.522	0.593	0.744	0.829	0.894
	AvgCov S_0^c		0.982	0.982	0.980	0.975	0.970	0.971	0.970	0.967	0.969	0.966	0.962	0.963	0.961	0.960	0.962
	AvgLength		0.348	0.396	0.551	0.723	0.926	0.209	0.240	0.343	0.464	0.624	0.145	0.167	0.240	0.328	0.451
	AvgLength^c		0.348	0.396	0.551	0.722	0.926	0.209	0.240	0.343	0.464	0.624	0.145	0.167	0.240	0.328	0.451
DEBIASED GLS	AvgCov S_0		0.882	0.872	0.866	0.876	0.888	0.724	0.705	0.669	0.675	0.699	0.522	0.482	0.449	0.438	0.455
	AvgCov S_0^c		0.982	0.982	0.985	0.987	0.988	0.970	0.973	0.975	0.979	0.982	0.962	0.964	0.969	0.970	0.976
	AvgLength		0.346	0.322	0.305	0.313	0.327	0.208	0.190	0.171	0.174	0.185	0.145	0.131	0.119	0.115	0.121
	AvgLength^c		0.346	0.322	0.310	0.313	0.326	0.208	0.190	0.171	0.174	0.185	0.145	0.131	0.119	0.115	0.121

Table E.45: See description of Table 3, for $|S_0| = 7$. $d = 16$, DGP 2 ($\boldsymbol{x}_t \sim t_d$ and $\varepsilon_t \sim$ i.i.d. $\mathcal{N}(0,1)$).

			$d = 16$														
		p/T			200					500					1000		
ϕ			0	0.5	0.8	0.9	0.95	0	0.5	0.8	0.9	0.95	0	0.5	0.8	0.9	0.95
DEBIASED LASSO	AvgCov S_0	100	0.906	0.911	0.934	0.939	0.945	0.857	0.880	0.914	0.930	0.938	0.776	0.816	0.882	0.912	0.927
	AvgCov S_0^c		0.957	0.957	0.957	0.957	0.955	0.953	0.953	0.952	0.953	0.953	0.952	0.951	0.952	0.951	0.951
	AvgLength		0.306	0.349	0.490	0.645	0.832	0.190	0.219	0.312	0.423	0.573	0.133	0.154	0.221	0.301	0.414
	AvgLength^c		0.306	0.349	0.490	0.645	0.832	0.190	0.219	0.312	0.423	0.573	0.133	0.154	0.221	0.301	0.414
DEBIASED GLS	AvgCov S_0		0.904	0.895	0.892	0.900	0.910	0.856	0.839	0.824	0.821	0.834	0.776	0.739	0.696	0.698	0.704
	AvgCov S_0^c		0.956	0.957	0.962	0.967	0.972	0.953	0.954	0.958	0.962	0.968	0.952	0.951	0.952	0.959	0.963
	AvgLength		0.304	0.280	0.256	0.260	0.272	0.189	0.171	0.153	0.152	0.157	0.133	0.120	0.105	0.104	0.105
	AvgLength^c		0.304	0.280	0.256	0.259	0.272	0.189	0.171	0.153	0.152	0.157	0.133	0.120	0.105	0.104	0.105
DEBIASED LASSO	AvgCov S_0	200	0.907	0.919	0.931	0.936	0.940	0.877	0.896	0.918	0.937	0.938	0.787	0.822	0.881	0.915	0.934
	AvgCov S_0^c		0.968	0.968	0.968	0.966	0.963	0.959	0.959	0.957	0.958	0.958	0.955	0.956	0.954	0.953	0.954
	AvgLength		0.312	0.355	0.495	0.649	0.832	0.191	0.220	0.314	0.425	0.574	0.134	0.154	0.221	0.302	0.415
	AvgLength^c		0.312	0.355	0.495	0.649	0.831	0.191	0.220	0.314	0.425	0.574	0.134	0.154	0.221	0.302	0.415
DEBIASED GLS	AvgCov S_0		0.906	0.900	0.898	0.908	0.917	0.877	0.864	0.845	0.847	0.862	0.787	0.759	0.729	0.727	0.731
	AvgCov S_0^c		0.968	0.969	0.973	0.977	0.979	0.959	0.961	0.964	0.968	0.974	0.955	0.956	0.960	0.964	0.968
	AvgLength		0.309	0.287	0.264	0.271	0.285	0.191	0.173	0.156	0.155	0.163	0.134	0.120	0.107	0.106	0.107
	AvgLength^c		0.309	0.287	0.264	0.271	0.285	0.191	0.173	0.156	0.155	0.163	0.134	0.120	0.107	0.106	0.107
DEBIASED LASSO	AvgCov S_0	500	0.908	0.914	0.929	0.937	0.938	0.888	0.900	0.927	0.939	0.944	0.809	0.848	0.902	0.921	0.937
	AvgCov S_0^c		0.982	0.981	0.980	0.976	0.972	0.970	0.969	0.966	0.968	0.966	0.960	0.961	0.959	0.959	0.960
	AvgLength		0.320	0.365	0.508	0.666	0.855	0.193	0.222	0.317	0.428	0.577	0.134	0.155	0.222	0.303	0.418
	AvgLength^c		0.320	0.365	0.508	0.666	0.855	0.193	0.222	0.317	0.428	0.577	0.134	0.155	0.222	0.303	0.418
DEBIASED GLS	AvgCov S_0		0.906	0.894	0.896	0.906	0.921	0.890	0.879	0.866	0.874	0.889	0.810	0.793	0.769	0.763	0.779
	AvgCov S_0^c		0.982	0.982	0.985	0.987	0.989	0.970	0.972	0.974	0.978	0.981	0.960	0.963	0.968	0.969	0.975
	AvgLength		0.318	0.295	0.278	0.288	0.302	0.193	0.176	0.158	0.161	0.171	0.134	0.121	0.110	0.106	0.112
	AvgLength^c		0.318	0.295	0.278	0.288	0.302	0.193	0.176	0.158	0.161	0.171	0.134	0.121	0.110	0.106	0.112

Table E.46: See description of Table 1, for $|S_0| = 3$, $d = 4$, DGP 3 ($\boldsymbol{x}_t \sim t_d$ and $\varepsilon_t \sim$ i.i.d. $\mathcal{N}(0,1)$).

		$d = 4$														
Panel I																
	p/T			200					500					1000		
ϕ		0	0.5	0.8	0.9	0.95	0	0.5	0.8	0.9	0.95	0	0.5	0.8	0.9	0.95
LASSO/ GLS LASSO	100	0.994	1.031	1.161	1.437	1.780	0.999	1.013	1.091	1.117	1.380	0.999	0.979	1.001	1.059	1.105
GLS LASSO		(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)
LASSO/ GLS LASSO	200	0.994	1.017	1.219	1.533	1.859	1.001	1.028	1.041	1.112	1.467	1.001	0.981	1.036	1.037	1.104
GLS LASSO		(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
LASSO/ GLS LASSO	500	0.986	1.063	1.322	1.678	1.991	0.996	1.034	1.010	1.204	1.615	0.999	1.013	1.046	1.012	1.264
GLS LASSO		(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Panel II																
DEBIASED LASSO/ DEBIASED GLS	100	1.004	1.285	2.069	2.819	3.610	1.001	1.293	2.105	2.967	4.048	1.000	1.292	2.114	3.004	4.178
DEBIASED GLS		(0.10)	(0.08)	(0.07)	(0.07)	(0.07)	(0.06)	(0.06)	(0.05)	(0.05)	(0.05)	(0.04)	(0.04)	(0.03)	(0.03)	(0.03)
DEBIASED LASSO/ DEBIASED GLS	200	1.005	1.290	2.102	2.906	3.657	1.002	1.298	2.109	2.979	4.073	1.001	1.295	2.124	3.016	4.212
DEBIASED GLS		(0.09)	(0.08)	(0.07)	(0.07)	(0.07)	(0.06)	(0.05)	(0.05)	(0.05)	(0.05)	(0.04)	(0.04)	(0.03)	(0.03)	(0.03)
DEBIASED LASSO/ DEBIASED GLS	500	1.007	1.287	2.146	2.954	3.657	1.003	1.300	2.122	3.067	4.190	1.001	1.303	2.117	3.004	4.253
DEBIASED GLS		(0.08)	(0.07)	(0.07)	(0.06)	(0.06)	(0.06)	(0.05)	(0.05)	(0.04)	(0.04)	(0.04)	(0.04)	(0.03)	(0.03)	(0.03)

Table E.47: See description of Table 1, for $|S_0| = 3$, $d = 8$, DGP 3 ($\boldsymbol{x}_t \sim t_d$ and $\varepsilon_t \sim$ i.i.d. $\mathcal{N}(0,1)$).

		$d = 8$														
Panel I																
	p/T			200					500					1000		
ϕ		0	0.5	0.8	0.9	0.95	0	0.5	0.8	0.9	0.95	0	0.5	0.8	0.9	0.95
LASSO/ GLS LASSO	100	0.984	1.103	1.593	2.211	2.830	0.997	1.161	1.405	1.796	2.759	1.000	1.089	1.448	1.641	2.173
GLS LASSO		(0.03)	(0.03)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)
LASSO/ GLS LASSO	200	0.992	1.113	1.726	2.374	2.897	1.000	1.142	1.303	1.916	2.906	0.998	1.096	1.453	1.519	2.368
GLS LASSO		(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
LASSO/ GLS LASSO	500	0.987	1.160	1.909	2.436	2.854	0.997	1.123	1.340	2.202	2.987	1.000	1.152	1.251	1.546	2.786
GLS LASSO		(0.02)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Panel II																
DEBIASED LASSO/ DEBIASED GLS	100	1.004	1.275	2.033	2.757	3.529	1.001	1.290	2.099	2.955	4.021	1.001	1.287	2.112	3.012	4.185
DEBIASED GLS		(0.08)	(0.07)	(0.06)	(0.06)	(0.06)	(0.05)	(0.05)	(0.04)	(0.04)	(0.04)	(0.04)	(0.03)	(0.03)	(0.03)	(0.03)
DEBIASED LASSO/ DEBIASED GLS	200	1.006	1.284	2.057	2.811	3.572	1.002	1.296	2.101	2.937	4.005	1.001	1.295	2.135	3.029	4.184
DEBIASED GLS		(0.07)	(0.07)	(0.06)	(0.06)	(0.06)	(0.05)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.03)	(0.03)	(0.03)	(0.03)
DEBIASED LASSO/ DEBIASED GLS	500	1.009	1.266	2.068	2.850	3.510	1.003	1.305	2.095	2.935	4.069	1.001	1.303	2.159	3.031	4.208
DEBIASED GLS		(0.07)	(0.06)	(0.05)	(0.05)	(0.05)	(0.05)	(0.04)	(0.04)	(0.04)	(0.04)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)

Table E.48: See description of Table 1, for $|S_0| = 3,\ d = 16$, DGP 3 ($\boldsymbol{x}_t \sim t_d$ and $\varepsilon_t \sim$ i.i.d. $\mathcal{N}(0,1)$).

		$d = 16$														
Panel I																
	p/T	200					500					1000				
ϕ		0	0.5	0.8	0.9	0.95	0	0.5	0.8	0.9	0.95	0	0.5	0.8	0.9	0.95
LASSO/ GLS LASSO	100	0.983	1.158	1.759	2.407	2.964	0.998	1.229	1.663	2.251	3.356	0.998	1.221	1.934	2.444	3.384
GLS LASSO		(0.03)	(0.03)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
LASSO/ GLS LASSO	200	0.989	1.154	1.909	2.474	2.982	0.995	1.212	1.519	2.296	3.326	0.999	1.277	1.855	2.100	3.418
GLS LASSO		(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
LASSO/ GLS LASSO	500	0.986	1.202	1.969	2.414	2.879	0.992	1.101	1.583	2.479	3.169	1.000	1.238	1.437	2.058	3.525
GLS LASSO		(0.02)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Panel II																
DEBIASED LASSO/ DEBIASED GLS	100	1.004	1.274	2.022	2.736	3.507	1.001	1.294	2.109	2.957	3.995	1.000	1.289	2.114	3.016	4.190
DEBIASED GLS		(0.07)	(0.06)	(0.06)	(0.06)	(0.06)	(0.05)	(0.04)	(0.04)	(0.04)	(0.03)	(0.03)	(0.03)	(0.03)	(0.02)	(0.02)
DEBIASED LASSO/ DEBIASED GLS	200	1.007	1.281	2.028	2.759	3.515	1.002	1.299	2.114	2.940	3.991	1.001	1.294	2.135	3.047	4.201
DEBIASED GLS		(0.07)	(0.06)	(0.05)	(0.05)	(0.05)	(0.05)	(0.04)	(0.04)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.02)	(0.02)
DEBIASED LASSO/ DEBIASED GLS	500	1.008	1.261	2.020	2.794	3.471	1.003	1.309	2.103	2.913	4.026	1.001	1.302	2.172	3.055	4.194
DEBIASED GLS		(0.06)	(0.06)	(0.05)	(0.05)	(0.05)	(0.04)	(0.04)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.02)	(0.02)	(0.02)

Table E.49: See description of Table 1, for $|S_0| = 7,\ d = 4$, DGP 3 ($\boldsymbol{x}_t \sim t_d$ and $\varepsilon_t \sim$ i.i.d. $\mathcal{N}(0,1)$).

		$d = 4$														
Panel I																
	p/T	200					500					1000				
ϕ		0	0.5	0.8	0.9	0.95	0	0.5	0.8	0.9	0.95	0	0.5	0.8	0.9	0.95
LASSO/ GLS LASSO	100	0.993	1.223	1.749	2.171	2.573	0.994	1.183	1.896	2.598	3.129	0.999	1.238	1.731	2.431	3.458
GLS LASSO		(0.04)	(0.04)	(0.03)	(0.03)	(0.03)	(0.03)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
LASSO/ GLS LASSO	200	0.997	1.212	1.672	1.991	2.389	0.991	1.182	2.013	2.598	3.040	0.999	1.218	1.672	2.601	3.466
GLS LASSO		(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.01)	(0.01)	(0.01)
LASSO/ GLS LASSO	500	0.997	1.185	1.605	1.967	2.331	0.992	1.265	1.989	2.324	2.829	0.997	1.142	1.893	2.693	3.226
GLS LASSO		(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Panel II																
DEBIASED LASSO/ DEBIASED GLS	100	1.002	1.262	1.977	2.658	3.368	1.000	1.283	2.068	2.878	3.892	1.000	1.288	2.105	2.979	4.103
DEBIASED GLS		(0.10)	(0.09)	(0.08)	(0.08)	(0.08)	(0.06)	(0.06)	(0.05)	(0.05)	(0.05)	(0.04)	(0.04)	(0.03)	(0.03)	(0.03)
DEBIASED LASSO/ DEBIASED GLS	200	1.006	1.255	1.980	2.696	3.419	1.001	1.294	2.087	2.881	3.894	1.000	1.293	2.130	3.009	4.127
DEBIASED GLS		(0.09)	(0.08)	(0.07)	(0.07)	(0.07)	(0.06)	(0.05)	(0.05)	(0.05)	(0.05)	(0.04)	(0.04)	(0.03)	(0.03)	(0.03)
DEBIASED LASSO/ DEBIASED GLS	500	1.007	1.247	2.014	2.798	3.461	1.003	1.307	2.055	2.865	3.960	1.001	1.307	2.178	3.012	4.124
DEBIASED GLS		(0.08)	(0.07)	(0.07)	(0.07)	(0.07)	(0.06)	(0.05)	(0.05)	(0.05)	(0.05)	(0.04)	(0.04)	(0.03)	(0.03)	(0.03)

Table E.50: See description of Table 1, for $|S_0| = 7,\ d = 8$, DGP 3 ($\boldsymbol{x}_t \sim t_d$ and $\varepsilon_t \sim$ i.i.d. $\mathcal{N}(0,1)$).

		$d = 8$														
Panel I																
	p/T	200					500					1000				
ϕ		0	0.5	0.8	0.9	0.95	0	0.5	0.8	0.9	0.95	0	0.5	0.8	0.9	0.95
LASSO/ GLS LASSO	100	0.987	1.212	1.865	2.302	2.699	0.996	1.209	1.755	2.522	3.281	0.998	1.270	1.859	2.387	3.549
GLS LASSO		(0.04)	(0.03)	(0.03)	(0.03)	(0.03)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)
LASSO/ GLS LASSO	200	0.995	1.214	1.810	2.225	2.643	0.996	1.163	1.806	2.577	3.237	0.999	1.251	1.649	2.287	3.513
GLS LASSO		(0.03)	(0.03)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
LASSO/ GLS LASSO	500	0.992	1.234	1.731	2.133	2.553	0.991	1.174	2.012	2.636	3.022	0.998	1.205	1.532	2.389	3.354
GLS LASSO		(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Panel II																
DEBIASED LASSO/ DEBIASED GLS	100	1.002	1.257	1.959	2.617	3.302	1.001	1.285	2.084	2.899	3.882	1.000	1.285	2.100	2.981	4.117
DEBIASED GLS		(0.08)	(0.07)	(0.06)	(0.06)	(0.06)	(0.05)	(0.05)	(0.04)	(0.04)	(0.04)	(0.04)	(0.03)	(0.03)	(0.03)	(0.03)
DEBIASED LASSO/ DEBIASED GLS	200	1.005	1.259	1.953	2.609	3.284	1.001	1.292	2.106	2.890	3.840	1.001	1.289	2.117	3.013	4.117
DEBIASED GLS		(0.07)	(0.07)	(0.06)	(0.06)	(0.06)	(0.05)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.03)	(0.03)	(0.03)	(0.03)
DEBIASED LASSO/ DEBIASED GLS	500	1.007	1.239	1.885	2.598	3.268	1.002	1.312	2.151	2.867	3.863	1.001	1.298	2.170	3.067	4.083
DEBIASED GLS		(0.07)	(0.06)	(0.06)	(0.06)	(0.06)	(0.05)	(0.04)	(0.04)	(0.04)	(0.04)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)

Table E.51: See description of Table 1, for $|S_0| = 7,\ d = 16$, DGP 3 ($\boldsymbol{x}_t \sim t_d$ and $\varepsilon_t \sim$ i.i.d. $\mathcal{N}(0,1)$).

									$d = 16$							
Panel I																
	p/T			200					500					1000		
ϕ		0	0.5	0.8	0.9	0.95	0	0.5	0.8	0.9	0.95	0	0.5	0.8	0.9	0.95
LASSO/ GLS LASSO	100	0.994	1.177	1.779	2.316	2.737	0.999	1.239	1.762	2.482	3.350	0.999	1.272	1.969	2.425	3.541
GLS LASSO		(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
LASSO/ GLS LASSO	200	0.989	1.210	1.862	2.291	2.709	0.994	1.185	1.677	2.463	3.214	0.999	1.271	1.768	2.280	3.530
GLS LASSO		(0.03)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
LASSO/ GLS LASSO	500	0.993	1.236	1.801	2.192	2.633	0.994	1.149	1.865	2.586	3.063	1.000	1.220	1.514	2.299	3.430
GLS LASSO		(0.02)	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Panel II																
DEBIASED LASSO/ DEBIASED GLS	100	1.003	1.257	1.959	2.605	3.278	1.001	1.287	2.090	2.903	3.875	1.000	1.284	2.102	2.980	4.121
DEBIASED GLS		(0.07)	(0.07)	(0.06)	(0.06)	(0.06)	(0.05)	(0.04)	(0.04)	(0.04)	(0.04)	(0.03)	(0.03)	(0.03)	(0.03)	(0.02)
DEBIASED LASSO/ DEBIASED GLS	200	1.005	1.254	1.949	2.579	3.262	1.001	1.290	2.110	2.898	3.828	1.001	1.289	2.112	3.006	4.124
DEBIASED GLS		(0.07)	(0.06)	(0.06)	(0.06)	(0.06)	(0.05)	(0.04)	(0.04)	(0.04)	(0.04)	(0.03)	(0.03)	(0.03)	(0.02)	(0.02)
DEBIASED LASSO/ DEBIASED GLS	500	1.009	1.241	1.857	2.506	3.175	1.002	1.304	2.196	2.897	3.826	1.001	1.298	2.148	3.085	4.093
DEBIASED GLS		(0.06)	(0.06)	(0.05)	(0.05)	(0.05)	(0.04)	(0.04)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.02)	(0.02)	(0.02)

Table E.52: See description of Table 2, for $|S_0| = 3,\ d = 4$, DGP 3 ($\boldsymbol{x}_t \sim t_d$ and $\varepsilon_t \sim$ i.i.d. $\mathcal{N}(0,1)$).

								Size adjusted power and standard errors								
								$d = 4$								
	p/T			200					500					1000		
ϕ		0	0.5	0.8	0.9	0.95	0	0.5	0.8	0.9	0.95	0	0.5	0.8	0.9	0.95
GLS LASSO	100	0.764	0.786	0.823	0.825	0.825	0.855	0.872	0.883	0.884	0.891	0.895	0.916	0.927	0.930	0.930
(s.e.)		(0.10)	(0.09)	(0.08)	(0.08)	(0.08)	(0.06)	(0.06)	(0.05)	(0.05)	(0.05)	(0.04)	(0.04)	(0.04)	(0.03)	(0.03)
LASSO		0.764	0.724	0.630	0.532	0.410	0.857	0.835	0.773	0.710	0.582	0.894	0.884	0.839	0.786	0.717
(s.e.)		(0.10)	(0.11)	(0.16)	(0.21)	(0.27)	(0.06)	(0.07)	(0.10)	(0.14)	(0.19)	(0.04)	(0.05)	(0.07)	(0.10)	(0.14)
GLS LASSO	200	0.785	0.823	0.839	0.852	0.854	0.875	0.899	0.903	0.911	0.908	0.915	0.921	0.933	0.937	0.941
(s.e.)		(0.10)	(0.09)	(0.08)	(0.08)	(0.08)	(0.06)	(0.06)	(0.05)	(0.05)	(0.05)	(0.04)	(0.04)	(0.03)	(0.03)	(0.03)
LASSO		0.791	0.753	0.677	0.568	0.441	0.878	0.856	0.780	0.719	0.639	0.916	0.898	0.849	0.813	0.730
(s.e.)		(0.10)	(0.11)	(0.16)	(0.21)	(0.27)	(0.06)	(0.07)	(0.10)	(0.14)	(0.19)	(0.04)	(0.05)	(0.07)	(0.10)	(0.14)
GLS LASSO	500	0.805	0.811	0.832	0.837	0.836	0.867	0.885	0.892	0.900	0.903	0.912	0.919	0.932	0.937	0.935
(s.e.)		(0.10)	(0.09)	(0.08)	(0.08)	(0.08)	(0.06)	(0.06)	(0.05)	(0.05)	(0.05)	(0.04)	(0.04)	(0.04)	(0.03)	(0.03)
LASSO		0.803	0.780	0.693	0.621	0.527	0.868	0.854	0.787	0.699	0.616	0.911	0.898	0.855	0.801	0.723
(s.e.)		(0.10)	(0.11)	(0.16)	(0.21)	(0.27)	(0.06)	(0.07)	(0.10)	(0.14)	(0.19)	(0.04)	(0.05)	(0.07)	(0.10)	(0.14)

Table E.53: See description of Table 2, for $|S_0| = 3,\ d = 8$, DGP 3 ($\boldsymbol{x}_t \sim t_d$ and $\varepsilon_t \sim$ i.i.d. $\mathcal{N}(0,1)$).

								Size adjusted power and standard errors								
								$d = 8$								
	p/T			200					500					1000		
ϕ		0	0.5	0.8	0.9	0.95	0	0.5	0.8	0.9	0.95	0	0.5	0.8	0.9	0.95
GLS LASSO	100	0.823	0.826	0.857	0.863	0.853	0.891	0.892	0.902	0.906	0.908	0.920	0.928	0.933	0.937	0.941
(s.e.)		(0.08)	(0.07)	(0.07)	(0.07)	(0.07)	(0.05)	(0.05)	(0.04)	(0.04)	(0.04)	(0.04)	(0.03)	(0.03)	(0.03)	(0.03)
LASSO		0.824	0.772	0.714	0.627	0.505	0.888	0.888	0.824	0.759	0.662	0.919	0.903	0.876	0.837	0.773
(s.e.)		(0.08)	(0.09)	(0.13)	(0.17)	(0.22)	(0.05)	(0.06)	(0.09)	(0.12)	(0.16)	(0.04)	(0.04)	(0.06)	(0.08)	(0.11)
GLS LASSO	200	0.834	0.856	0.877	0.874	0.884	0.900	0.907	0.924	0.919	0.928	0.923	0.932	0.944	0.945	0.949
(s.e.)		(0.08)	(0.07)	(0.07)	(0.07)	(0.07)	(0.05)	(0.05)	(0.04)	(0.04)	(0.04)	(0.04)	(0.03)	(0.03)	(0.03)	(0.03)
LASSO		0.836	0.790	0.717	0.628	0.533	0.900	0.886	0.825	0.767	0.684	0.922	0.913	0.883	0.832	0.776
(s.e.)		(0.08)	(0.09)	(0.13)	(0.17)	(0.22)	(0.05)	(0.06)	(0.09)	(0.12)	(0.16)	(0.04)	(0.04)	(0.06)	(0.08)	(0.11)
GLS LASSO	500	0.828	0.835	0.847	0.848	0.845	0.890	0.902	0.915	0.918	0.914	0.925	0.930	0.943	0.947	0.948
(s.e.)		(0.08)	(0.08)	(0.07)	(0.07)	(0.07)	(0.05)	(0.05)	(0.04)	(0.04)	(0.04)	(0.04)	(0.03)	(0.03)	(0.03)	(0.03)
LASSO		0.822	0.806	0.748	0.664	0.568	0.891	0.873	0.827	0.766	0.686	0.925	0.911	0.870	0.826	0.763
(s.e.)		(0.08)	(0.09)	(0.13)	(0.17)	(0.23)	(0.05)	(0.06)	(0.09)	(0.12)	(0.16)	(0.04)	(0.04)	(0.06)	(0.08)	(0.11)

Table E.54: See description of Table 2, for $|S_0| = 3,\ d = 16$, DGP 3 ($\boldsymbol{x}_t \sim t_d$ and $\varepsilon_t \sim$ i.i.d. $\mathcal{N}(0, 1)$).

Size adjusted power and standard errors																
$d = 16$																
	p/T	200					500					1000				
ϕ		0	0.5	0.8	0.9	0.95	0	0.5	0.8	0.9	0.95	0	0.5	0.8	0.9	0.95
GLS LASSO	100	0.829	0.851	0.860	0.856	0.855	0.881	0.895	0.912	0.917	0.916	0.922	0.923	0.932	0.936	0.934
(s.e.)		(0.07)	(0.07)	(0.06)	(0.06)	(0.06)	(0.05)	(0.04)	(0.04)	(0.04)	(0.04)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
LASSO		0.833	0.799	0.703	0.627	0.566	0.879	0.873	0.809	0.748	0.641	0.923	0.916	0.882	0.835	0.796
(s.e.)		(0.08)	(0.09)	(0.12)	(0.16)	(0.21)	(0.05)	(0.06)	(0.08)	(0.11)	(0.14)	(0.03)	(0.04)	(0.06)	(0.08)	(0.10)
GLS LASSO	200	0.847	0.859	0.877	0.882	0.886	0.918	0.921	0.930	0.936	0.936	0.936	0.943	0.949	0.951	0.955
(s.e.)		(0.08)	(0.07)	(0.06)	(0.06)	(0.06)	(0.05)	(0.04)	(0.04)	(0.04)	(0.04)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
LASSO		0.847	0.816	0.741	0.649	0.548	0.918	0.899	0.849	0.796	0.746	0.934	0.925	0.891	0.852	0.801
(s.e.)		(0.08)	(0.09)	(0.12)	(0.16)	(0.21)	(0.05)	(0.06)	(0.08)	(0.11)	(0.14)	(0.03)	(0.04)	(0.06)	(0.08)	(0.11)
GLS LASSO	500	0.834	0.836	0.857	0.858	0.859	0.905	0.915	0.923	0.929	0.929	0.922	0.932	0.943	0.949	0.951
(s.e.)		(0.08)	(0.07)	(0.06)	(0.06)	(0.07)	(0.05)	(0.04)	(0.04)	(0.04)	(0.04)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
		0.836	0.817	0.761	0.682	0.583	0.906	0.889	0.839	0.783	0.720	0.923	0.916	0.879	0.836	0.775
(s.e.)		(0.08)	(0.09)	(0.12)	(0.16)	(0.21)	(0.05)	(0.06)	(0.08)	(0.11)	(0.14)	(0.03)	(0.04)	(0.06)	(0.08)	(0.11)

Table E.55: See description of Table 2, for $|S_0| = 7,\ d = 4$, DGP 3 ($\boldsymbol{x}_t \sim t_d$ and $\varepsilon_t \sim$ i.i.d. $\mathcal{N}(0, 1)$).

Size adjusted power and standard errors																
$d = 4$																
	p/T	200					500					1000				
ϕ		0	0.5	0.8	0.9	0.95	0	0.5	0.8	0.9	0.95	0	0.5	0.8	0.9	0.95
GLS LASSO	100	0.721	0.725	0.734	0.733	0.719	0.761	0.762	0.761	0.753	0.748	0.731	0.737	0.740	0.749	0.749
(s.e.)		(0.10)	(0.09)	(0.09)	(0.09)	(0.09)	(0.06)	(0.06)	(0.05)	(0.05)	(0.05)	(0.04)	(0.04)	(0.04)	(0.03)	(0.04)
LASSO		0.723	0.677	0.580	0.464	0.428	0.762	0.741	0.676	0.602	0.529	0.730	0.737	0.726	0.727	0.625
(s.e.)		(0.10)	(0.12)	(0.16)	(0.21)	(0.28)	(0.06)	(0.07)	(0.11)	(0.14)	(0.19)	(0.04)	(0.05)	(0.07)	(0.10)	(0.14)
GLS LASSO	200	0.755	0.773	0.795	0.799	0.802	0.851	0.861	0.882	0.886	0.884	0.894	0.909	0.919	0.921	0.922
(s.e.)		(0.10)	(0.10)	(0.09)	(0.09)	(0.09)	(0.06)	(0.06)	(0.05)	(0.05)	(0.05)	(0.05)	(0.04)	(0.04)	(0.03)	(0.04)
LASSO		0.754	0.723	0.634	0.524	0.432	0.851	0.839	0.772	0.699	0.609	0.894	0.879	0.832	0.778	0.704
(s.e.)		(0.10)	(0.12)	(0.16)	(0.22)	(0.28)	(0.06)	(0.07)	(0.11)	(0.14)	(0.19)	(0.05)	(0.05)	(0.07)	(0.10)	(0.14)
GLS LASSO	500	0.776	0.781	0.799	0.792	0.796	0.861	0.877	0.888	0.890	0.891	0.903	0.911	0.925	0.926	0.925
(s.e.)		(0.11)	(0.10)	(0.09)	(0.09)	(0.10)	(0.06)	(0.06)	(0.05)	(0.05)	(0.06)	(0.05)	(0.04)	(0.04)	(0.04)	(0.04)
LASSO		0.774	0.742	0.660	0.529	0.421	0.861	0.844	0.779	0.702	0.605	0.903	0.888	0.840	0.784	0.708
(s.e.)		(0.11)	(0.12)	(0.17)	(0.22)	(0.28)	(0.06)	(0.07)	(0.11)	(0.14)	(0.19)	(0.05)	(0.05)	(0.07)	(0.10)	(0.14)

Table E.56: See description of Table 2, for $|S_0| = 7,\ d = 8$, DGP 3 ($\boldsymbol{x}_t \sim t_d$ and $\varepsilon_t \sim$ i.i.d. $\mathcal{N}(0, 1)$).

Size adjusted power and standard errors																
$d = 8$																
	p/T	200					500					1000				
ϕ		0	0.5	0.8	0.9	0.95	0	0.5	0.8	0.9	0.95	0	0.5	0.8	0.9	0.95
GLS LASSO	100	0.697	0.684	0.700	0.717	0.699	0.762	0.784	0.784	0.786	0.792	0.735	0.737	0.737	0.747	0.735
(s.e.)		(0.08)	(0.08)	(0.07)	(0.07)	(0.07)	(0.05)	(0.05)	(0.04)	(0.04)	(0.04)	(0.04)	(0.03)	(0.03)	(0.03)	(0.03)
LASSO		0.692	0.658	0.600	0.551	0.481	0.760	0.763	0.718	0.674	0.562	0.734	0.726	0.712	0.695	0.660
(s.e.)		(0.08)	(0.10)	(0.13)	(0.18)	(0.23)	(0.05)	(0.06)	(0.09)	(0.12)	(0.16)	(0.04)	(0.04)	(0.06)	(0.08)	(0.11)
GLS LASSO	200	0.808	0.817	0.836	0.843	0.838	0.884	0.896	0.907	0.910	0.911	0.914	0.924	0.934	0.934	0.935
(s.e.)		(0.08)	(0.08)	(0.07)	(0.07)	(0.08)	(0.05)	(0.05)	(0.04)	(0.04)	(0.05)	(0.04)	(0.03)	(0.03)	(0.03)	(0.03)
LASSO		0.809	0.782	0.687	0.598	0.514	0.883	0.870	0.810	0.744	0.652	0.914	0.903	0.867	0.817	0.755
(s.e.)		(0.09)	(0.10)	(0.14)	(0.18)	(0.23)	(0.05)	(0.06)	(0.09)	(0.12)	(0.16)	(0.04)	(0.04)	(0.06)	(0.08)	(0.11)
GLS LASSO	500	0.811	0.831	0.838	0.838	0.839	0.890	0.898	0.909	0.910	0.908	0.919	0.927	0.937	0.938	0.938
(s.e.)		(0.09)	(0.08)	(0.08)	(0.08)	(0.08)	(0.05)	(0.05)	(0.04)	(0.04)	(0.05)	(0.04)	(0.03)	(0.03)	(0.03)	(0.03)
LASSO		0.812	0.786	0.715	0.632	0.546	0.888	0.868	0.812	0.753	0.663	0.920	0.908	0.871	0.828	0.755
(s.e.)		(0.09)	(0.10)	(0.14)	(0.18)	(0.23)	(0.05)	(0.06)	(0.09)	(0.12)	(0.16)	(0.04)	(0.04)	(0.06)	(0.08)	(0.11)

Table E.57: See description of Table 2, for $|S_0| = 7,\ d = 16$, DGP 3 ($\boldsymbol{x}_t \sim t_d$ and $\varepsilon_t \sim$ i.i.d. $\mathcal{N}(0,1)$).

Size adjusted power and standard errors																
$d = 16$																
	p/T			200					500					1000		
ϕ		0	0.5	0.8	0.9	0.95	0	0.5	0.8	0.9	0.95	0	0.5	0.8	0.9	0.95
GLS LASSO	100	0.677	0.700	0.709	0.715	0.705	0.777	0.779	0.777	0.778	0.772	0.747	0.737	0.735	0.738	0.739
(s.e.)		(0.08)	(0.07)	(0.06)	(0.07)	(0.07)	(0.05)	(0.04)	(0.04)	(0.04)	(0.04)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
LASSO		0.676	0.675	0.630	0.559	0.488	0.776	0.765	0.719	0.691	0.638	0.746	0.747	0.753	0.752	0.719
(s.e.)		(0.08)	(0.09)	(0.12)	(0.16)	(0.21)	(0.05)	(0.06)	(0.08)	(0.11)	(0.15)	(0.03)	(0.04)	(0.06)	(0.08)	(0.11)
GLS LASSO	200	0.806	0.818	0.837	0.839	0.841	0.891	0.897	0.911	0.914	0.912	0.928	0.931	0.942	0.945	0.945
(s.e.)		(0.08)	(0.07)	(0.07)	(0.07)	(0.07)	(0.05)	(0.04)	(0.04)	(0.04)	(0.04)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
LASSO		0.807	0.788	0.700	0.623	0.534	0.891	0.869	0.817	0.754	0.683	0.928	0.915	0.874	0.829	0.774
(s.e.)		(0.08)	(0.09)	(0.13)	(0.17)	(0.21)	(0.05)	(0.06)	(0.08)	(0.11)	(0.15)	(0.03)	(0.04)	(0.06)	(0.08)	(0.11)
GLS LASSO	500	0.818	0.827	0.841	0.842	0.840	0.894	0.901	0.912	0.912	0.914	0.925	0.931	0.941	0.944	0.943
(s.e.)		(0.08)	(0.08)	(0.07)	(0.07)	(0.08)	(0.05)	(0.04)	(0.04)	(0.04)	(0.04)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
LASSO		0.818	0.796	0.722	0.662	0.555	0.893	0.878	0.830	0.770	0.695	0.924	0.913	0.876	0.833	0.772
(s.e.)		(0.08)	(0.09)	(0.13)	(0.17)	(0.22)	(0.05)	(0.06)	(0.08)	(0.11)	(0.15)	(0.03)	(0.04)	(0.06)	(0.08)	(0.11)

Table E.58: See description of Table 3, for $|S_0| = 3$. $d = 4$, DGP 3 ($\boldsymbol{x}_t \sim t_d$ and $\varepsilon_t \sim$ i.i.d. $\mathcal{N}(0,1)$).

$d = 4$																	
		p/T			200					500					1000		
ϕ			0	0.5	0.8	0.9	0.95	0	0.5	0.8	0.9	0.95	0	0.5	0.8	0.9	0.95
DEBIASED LASSO	AvgCov S_0	100	0.463	0.517	0.653	0.744	0.820	0.251	0.292	0.418	0.552	0.676	0.161	0.184	0.273	0.365	0.492
	AvgCov S_0^c		0.958	0.957	0.957	0.955	0.955	0.955	0.955	0.953	0.953	0.953	0.952	0.952	0.953	0.951	0.952
	AvgLength		0.285	0.325	0.458	0.607	0.790	0.177	0.204	0.291	0.396	0.537	0.125	0.144	0.206	0.281	0.387
	AvgLength^c		0.284	0.324	0.457	0.606	0.788	0.177	0.203	0.291	0.395	0.537	0.125	0.143	0.206	0.281	0.387
DEBIASED GLS	AvgCov S_0		0.460	0.428	0.390	0.390	0.408	0.251	0.230	0.205	0.200	0.203	0.161	0.149	0.130	0.125	0.126
	AvgCov S_0^c		0.957	0.960	0.963	0.965	0.970	0.954	0.956	0.957	0.960	0.965	0.951	0.953	0.953	0.956	0.961
	AvgLength		0.283	0.258	0.232	0.231	0.240	0.177	0.159	0.141	0.138	0.141	0.124	0.111	0.098	0.095	0.095
	AvgLength^c		0.282	0.258	0.232	0.230	0.240	0.177	0.159	0.141	0.138	0.141	0.124	0.111	0.098	0.095	0.095
DEBIASED LASSO	AvgCov S_0	200	0.489	0.548	0.675	0.774	0.832	0.246	0.284	0.398	0.531	0.669	0.150	0.180	0.271	0.377	0.499
	AvgCov S_0^c		0.965	0.965	0.964	0.964	0.961	0.959	0.959	0.958	0.958	0.958	0.954	0.956	0.957	0.956	0.955
	AvgLength		0.290	0.332	0.468	0.619	0.804	0.178	0.204	0.291	0.395	0.536	0.125	0.144	0.207	0.282	0.387
	AvgLength^c		0.289	0.331	0.466	0.617	0.802	0.177	0.204	0.291	0.395	0.536	0.125	0.144	0.206	0.282	0.387
DEBIASED GLS	AvgCov S_0		0.488	0.460	0.421	0.426	0.446	0.245	0.224	0.196	0.195	0.201	0.152	0.136	0.121	0.115	0.117
	AvgCov S_0^c		0.965	0.967	0.970	0.973	0.977	0.959	0.961	0.964	0.966	0.971	0.953	0.955	0.958	0.961	0.964
	AvgLength		0.288	0.263	0.238	0.239	0.251	0.177	0.160	0.143	0.140	0.144	0.125	0.112	0.098	0.096	0.097
	AvgLength^c		0.287	0.262	0.237	0.238	0.250	0.177	0.160	0.142	0.140	0.144	0.125	0.112	0.098	0.096	0.097
DEBIASED LASSO	AvgCov S_0	500	0.541	0.602	0.726	0.822	0.878	0.281	0.323	0.445	0.585	0.717	0.179	0.211	0.297	0.394	0.538
	AvgCov S_0^c		0.978	0.977	0.976	0.972	0.970	0.969	0.969	0.966	0.966	0.966	0.960	0.961	0.961	0.959	0.961
	AvgLength		0.307	0.349	0.488	0.643	0.829	0.180	0.207	0.294	0.397	0.536	0.125	0.145	0.208	0.283	0.389
	AvgLength^c		0.372	0.364	0.537	0.775	0.907	0.180	0.206	0.294	0.397	0.536	0.125	0.145	0.207	0.283	0.388
DEBIASED GLS	AvgCov S_0		0.540	0.517	0.482	0.486	0.511	0.282	0.262	0.233	0.232	0.236	0.179	0.157	0.138	0.136	0.137
	AvgCov S_0^c		0.978	0.978	0.981	0.983	0.985	0.969	0.971	0.973	0.975	0.979	0.960	0.961	0.965	0.968	0.971
	AvgLength		0.315	0.286	0.274	0.279	0.364	0.179	0.162	0.145	0.142	0.149	0.125	0.112	0.100	0.097	0.099
	AvgLength^c		0.335	0.344	0.309	0.311	0.368	0.179	0.162	0.145	0.142	0.149	0.125	0.112	0.100	0.097	0.099

Table E.59: See description of Table 3, for $|S_0| = 3$. $d = 8$, DGP 3 ($\boldsymbol{x}_t \sim t_d$ and $\varepsilon_t \sim$ i.i.d. $\mathcal{N}(0,1)$).

			$d=8$														
		p/T			200					500					1000		
ϕ			0	0.5	0.8	0.9	0.95	0	0.5	0.8	0.9	0.95	0	0.5	0.8	0.9	0.95
DEBIASED LASSO	AvgCov S_0	100	0.789	0.812	0.867	0.892	0.908	0.602	0.661	0.788	0.859	0.906	0.432	0.485	0.652	0.775	0.854
	AvgCov S_0^c		0.957	0.957	0.957	0.955	0.952	0.955	0.954	0.953	0.953	0.953	0.952	0.952	0.952	0.952	0.952
	AvgLength		0.278	0.318	0.448	0.594	0.772	0.175	0.202	0.288	0.392	0.532	0.124	0.143	0.205	0.280	0.385
	AvgLength^c		0.278	0.317	0.447	0.594	0.772	0.175	0.202	0.288	0.392	0.532	0.124	0.143	0.205	0.280	0.385
DEBIASED GLS	AvgCov S_0		0.790	0.767	0.738	0.746	0.764	0.603	0.565	0.514	0.502	0.521	0.429	0.394	0.350	0.337	0.338
	AvgCov S_0^c		0.956	0.959	0.961	0.965	0.970	0.954	0.955	0.955	0.958	0.965	0.952	0.952	0.953	0.956	0.960
	AvgLength		0.276	0.252	0.226	0.225	0.236	0.175	0.157	0.139	0.135	0.140	0.124	0.111	0.097	0.094	0.094
	AvgLength^c		0.276	0.252	0.226	0.225	0.235	0.175	0.157	0.139	0.135	0.140	0.124	0.111	0.097	0.094	0.094
DEBIASED LASSO	AvgCov S_0	200	0.807	0.835	0.885	0.907	0.925	0.612	0.656	0.784	0.863	0.908	0.427	0.498	0.655	0.773	0.859
	AvgCov S_0^c		0.964	0.964	0.963	0.961	0.959	0.958	0.957	0.956	0.957	0.956	0.954	0.955	0.955	0.954	0.955
	AvgLength		0.280	0.321	0.451	0.597	0.776	0.176	0.202	0.289	0.392	0.531	0.124	0.143	0.205	0.280	0.385
	AvgLength^c		0.280	0.320	0.451	0.597	0.775	0.176	0.202	0.289	0.392	0.531	0.124	0.143	0.205	0.280	0.385
DEBIASED GLS	AvgCov S_0		0.804	0.783	0.760	0.767	0.779	0.610	0.571	0.525	0.526	0.541	0.427	0.390	0.346	0.337	0.340
	AvgCov S_0^c		0.963	0.966	0.970	0.973	0.976	0.958	0.960	0.960	0.963	0.970	0.954	0.955	0.957	0.959	0.964
	AvgLength		0.278	0.255	0.230	0.233	0.243	0.176	0.158	0.140	0.137	0.143	0.124	0.111	0.098	0.095	0.096
	AvgLength^c		0.278	0.254	0.230	0.233	0.242	0.175	0.158	0.140	0.137	0.143	0.124	0.111	0.098	0.095	0.096
DEBIASED LASSO	AvgCov S_0	500	0.835	0.855	0.919	0.935	0.938	0.661	0.713	0.817	0.883	0.920	0.459	0.516	0.679	0.797	0.868
	AvgCov S_0^c		0.975	0.974	0.974	0.970	0.967	0.966	0.965	0.963	0.964	0.962	0.960	0.960	0.957	0.957	0.958
	AvgLength		0.285	0.325	0.457	0.603	0.780	0.177	0.203	0.290	0.391	0.528	0.124	0.143	0.206	0.281	0.386
	AvgLength^c		0.285	0.325	0.456	0.602	0.779	0.177	0.203	0.290	0.391	0.527	0.124	0.143	0.206	0.281	0.386
DEBIASED GLS	AvgCov S_0		0.834	0.810	0.793	0.802	0.814	0.663	0.626	0.572	0.577	0.591	0.459	0.418	0.380	0.366	0.379
	AvgCov S_0^c		0.975	0.976	0.980	0.982	0.984	0.966	0.968	0.968	0.973	0.977	0.960	0.962	0.964	0.966	0.971
	AvgLength		0.283	0.258	0.237	0.242	0.252	0.176	0.160	0.140	0.142	0.148	0.124	0.112	0.099	0.095	0.099
	AvgLength^c		0.283	0.258	0.237	0.242	0.252	0.176	0.160	0.140	0.142	0.147	0.124	0.112	0.099	0.095	0.099

Table E.60: See description of Table 3, for $|S_0| = 3$. $d = 16$, DGP 3 ($\boldsymbol{x}_t \sim t_d$ and $\varepsilon_t \sim$ i.i.d. $\mathcal{N}(0,1)$).

			$d=16$														
		p/T			200					500					1000		
ϕ			0	0.5	0.8	0.9	0.95	0	0.5	0.8	0.9	0.95	0	0.5	0.8	0.9	0.95
DEBIASED LASSO	AvgCov S_0	100	0.886	0.890	0.918	0.927	0.938	0.843	0.858	0.908	0.932	0.937	0.766	0.799	0.875	0.910	0.928
	AvgCov S_0^c		0.956	0.956	0.957	0.953	0.952	0.953	0.953	0.953	0.954	0.953	0.951	0.952	0.952	0.952	0.952
	AvgLength		0.277	0.317	0.447	0.593	0.771	0.175	0.202	0.288	0.392	0.531	0.124	0.143	0.205	0.280	0.385
	AvgLength^c		0.277	0.317	0.447	0.593	0.771	0.175	0.201	0.288	0.391	0.531	0.124	0.143	0.205	0.280	0.385
DEBIASED GLS	AvgCov S_0		0.885	0.878	0.872	0.879	0.890	0.842	0.820	0.795	0.795	0.811	0.767	0.724	0.679	0.673	0.676
	AvgCov S_0^c		0.956	0.959	0.961	0.966	0.970	0.953	0.954	0.955	0.958	0.964	0.951	0.952	0.953	0.955	0.959
	AvgLength		0.276	0.252	0.225	0.226	0.234	0.175	0.157	0.139	0.135	0.139	0.124	0.111	0.097	0.094	0.094
	AvgLength^c		0.276	0.251	0.225	0.225	0.234	0.175	0.157	0.138	0.135	0.139	0.124	0.111	0.097	0.094	0.094
DEBIASED LASSO	AvgCov S_0	200	0.897	0.912	0.921	0.926	0.933	0.838	0.864	0.906	0.925	0.932	0.753	0.799	0.874	0.912	0.931
	AvgCov S_0^c		0.963	0.963	0.963	0.961	0.958	0.957	0.956	0.956	0.956	0.955	0.953	0.954	0.953	0.953	0.953
	AvgLength		0.279	0.320	0.450	0.596	0.774	0.176	0.202	0.289	0.392	0.531	0.124	0.143	0.205	0.280	0.385
	AvgLength^c		0.279	0.319	0.449	0.595	0.773	0.176	0.202	0.289	0.392	0.531	0.124	0.143	0.205	0.280	0.385
DEBIASED GLS	AvgCov S_0		0.897	0.884	0.877	0.886	0.890	0.840	0.830	0.804	0.809	0.819	0.753	0.728	0.687	0.676	0.689
	AvgCov S_0^c		0.962	0.965	0.969	0.973	0.976	0.957	0.959	0.959	0.963	0.969	0.953	0.955	0.956	0.958	0.964
	AvgLength		0.277	0.253	0.229	0.232	0.242	0.175	0.158	0.139	0.137	0.143	0.124	0.111	0.098	0.094	0.096
	AvgLength^c		0.277	0.252	0.229	0.231	0.241	0.175	0.158	0.139	0.137	0.143	0.124	0.111	0.098	0.094	0.096
DEBIASED LASSO	AvgCov S_0	500	0.908	0.906	0.927	0.930	0.935	0.862	0.884	0.917	0.937	0.938	0.772	0.815	0.885	0.919	0.935
	AvgCov S_0^c		0.975	0.975	0.974	0.969	0.967	0.964	0.962	0.963	0.963	0.960	0.959	0.958	0.956	0.957	0.957
	AvgLength		0.282	0.323	0.453	0.599	0.774	0.176	0.203	0.289	0.391	0.527	0.124	0.143	0.206	0.281	0.386
	AvgLength^c		0.282	0.323	0.453	0.599	0.774	0.176	0.203	0.289	0.391	0.527	0.124	0.143	0.206	0.281	0.385
DEBIASED GLS	AvgCov S_0		0.902	0.887	0.891	0.898	0.901	0.859	0.844	0.828	0.833	0.846	0.771	0.743	0.709	0.704	0.724
	AvgCov S_0^c		0.974	0.976	0.980	0.982	0.983	0.964	0.966	0.967	0.972	0.975	0.959	0.961	0.963	0.965	0.972
	AvgLength		0.281	0.257	0.237	0.238	0.251	0.176	0.159	0.140	0.142	0.146	0.124	0.112	0.098	0.095	0.100
	AvgLength^c		0.281	0.257	0.237	0.238	0.251	0.176	0.159	0.140	0.142	0.146	0.124	0.112	0.098	0.095	0.100

Table E.61: See description of Table 3, for $|S_0| = 7$. $d = 4$, DGP 3 ($\boldsymbol{x}_t \sim t_d$ and $\varepsilon_t \sim$ i.i.d. $\mathcal{N}(0,1)$).

			$d=4$														
		p/T	200					500					1000				
ϕ			0	0.5	0.8	0.9	0.95	0	0.5	0.8	0.9	0.95	0	0.5	0.8	0.9	0.95
DEBIASED LASSO	AvgCov S_0	100	0.473	0.533	0.668	0.758	0.826	0.249	0.284	0.409	0.546	0.671	0.162	0.186	0.271	0.366	0.494
	AvgCov S_0^c		0.956	0.955	0.954	0.955	0.955	0.950	0.951	0.953	0.952	0.953	0.952	0.950	0.953	0.952	0.951
	AvgLength		0.295	0.336	0.471	0.624	0.810	0.179	0.206	0.295	0.400	0.542	0.125	0.144	0.207	0.282	0.388
	AvgLength^c		0.294	0.335	0.471	0.623	0.809	0.179	0.206	0.294	0.400	0.542	0.125	0.144	0.207	0.282	0.388
DEBIASED GLS	AvgCov S_0		0.472	0.445	0.418	0.419	0.440	0.249	0.225	0.202	0.204	0.209	0.162	0.146	0.129	0.127	0.131
	AvgCov S_0^c		0.955	0.957	0.963	0.967	0.970	0.950	0.953	0.955	0.962	0.967	0.951	0.951	0.951	0.956	0.964
	AvgLength		0.293	0.270	0.250	0.252	0.264	0.179	0.162	0.144	0.146	0.150	0.125	0.113	0.099	0.097	0.100
	AvgLength^c		0.293	0.269	0.250	0.252	0.264	0.179	0.162	0.144	0.146	0.150	0.125	0.113	0.099	0.097	0.100
DEBIASED LASSO	AvgCov S_0	200	0.523	0.584	0.723	0.805	0.857	0.267	0.305	0.427	0.556	0.688	0.168	0.191	0.277	0.386	0.514
	AvgCov S_0^c		0.969	0.968	0.967	0.966	0.965	0.958	0.960	0.960	0.958	0.959	0.954	0.953	0.957	0.956	0.955
	AvgLength		0.304	0.347	0.484	0.638	0.825	0.181	0.209	0.298	0.403	0.546	0.126	0.145	0.208	0.284	0.390
	AvgLength^c		0.304	0.346	0.483	0.637	0.824	0.181	0.208	0.297	0.403	0.546	0.126	0.145	0.208	0.284	0.390
DEBIASED GLS	AvgCov S_0		0.525	0.498	0.478	0.477	0.496	0.266	0.243	0.225	0.224	0.229	0.168	0.150	0.132	0.131	0.134
	AvgCov S_0^c		0.968	0.970	0.974	0.977	0.979	0.958	0.960	0.964	0.969	0.973	0.954	0.954	0.956	0.963	0.968
	AvgLength		0.302	0.281	0.264	0.268	0.281	0.181	0.164	0.148	0.151	0.155	0.126	0.113	0.100	0.100	0.103
	AvgLength^c		0.302	0.281	0.264	0.267	0.281	0.181	0.163	0.148	0.150	0.154	0.126	0.113	0.100	0.100	0.103
DEBIASED LASSO	AvgCov S_0	500	0.593	0.653	0.775	0.848	0.889	0.298	0.348	0.487	0.610	0.731	0.182	0.205	0.303	0.406	0.540
	AvgCov S_0^c		0.985	0.984	0.982	0.979	0.975	0.970	0.972	0.971	0.969	0.970	0.960	0.961	0.964	0.962	0.961
	AvgLength		0.331	0.409	0.531	0.693	0.998	0.185	0.213	0.302	0.407	0.548	0.127	0.147	0.210	0.287	0.393
	AvgLength^c		0.343	0.401	0.560	0.710	0.911	0.185	0.213	0.302	0.407	0.548	0.127	0.147	0.210	0.286	0.393
DEBIASED GLS	AvgCov S_0		0.591	0.574	0.555	0.558	0.572	0.298	0.276	0.253	0.252	0.263	0.182	0.162	0.149	0.148	0.149
	AvgCov S_0^c		0.984	0.984	0.986	0.988	0.989	0.970	0.973	0.977	0.980	0.982	0.960	0.962	0.965	0.972	0.974
	AvgLength		0.387	0.324	0.411	0.317	0.675	0.185	0.167	0.155	0.155	0.161	0.127	0.114	0.102	0.104	0.105
	AvgLength^c		0.342	0.355	1.035	0.609	0.371	0.185	0.167	0.155	0.155	0.161	0.127	0.114	0.102	0.104	0.105

Table E.62: See description of Table 3, for $|S_0| = 7$. $d = 8$, DGP 3 ($\boldsymbol{x}_t \sim t_d$ and $\varepsilon_t \sim$ i.i.d. $\mathcal{N}(0,1)$).

			$d=8$														
		p/T	200					500					1000				
ϕ			0	0.5	0.8	0.9	0.95	0	0.5	0.8	0.9	0.95	0	0.5	0.8	0.9	0.95
DEBIASED LASSO	AvgCov S_0	100	0.805	0.832	0.886	0.909	0.917	0.609	0.669	0.791	0.862	0.908	0.435	0.499	0.656	0.773	0.852
	AvgCov S_0^c		0.959	0.956	0.956	0.958	0.956	0.955	0.954	0.953	0.953	0.953	0.951	0.951	0.953	0.952	0.951
	AvgLength		0.287	0.327	0.460	0.609	0.789	0.178	0.204	0.292	0.396	0.537	0.125	0.144	0.206	0.281	0.387
	AvgLength^c		0.287	0.327	0.460	0.609	0.789	0.178	0.204	0.292	0.396	0.537	0.125	0.144	0.206	0.281	0.387
DEBIASED GLS	AvgCov S_0		0.802	0.788	0.767	0.772	0.790	0.606	0.566	0.523	0.522	0.533	0.434	0.394	0.350	0.343	0.352
	AvgCov S_0^c		0.957	0.960	0.963	0.968	0.972	0.954	0.955	0.958	0.963	0.967	0.951	0.952	0.953	0.959	0.965
	AvgLength		0.285	0.263	0.241	0.244	0.257	0.177	0.160	0.143	0.143	0.147	0.125	0.112	0.098	0.097	0.099
	AvgLength^c		0.285	0.263	0.241	0.243	0.257	0.177	0.160	0.143	0.143	0.147	0.125	0.112	0.098	0.097	0.099
DEBIASED LASSO	AvgCov S_0	200	0.820	0.845	0.892	0.919	0.930	0.614	0.672	0.796	0.863	0.902	0.432	0.496	0.656	0.772	0.849
	AvgCov S_0^c		0.969	0.968	0.967	0.967	0.965	0.961	0.960	0.957	0.958	0.959	0.954	0.955	0.956	0.954	0.955
	AvgLength		0.293	0.334	0.468	0.617	0.799	0.179	0.206	0.293	0.398	0.538	0.125	0.144	0.207	0.282	0.388
	AvgLength^c		0.293	0.334	0.468	0.617	0.799	0.179	0.206	0.293	0.398	0.538	0.125	0.144	0.207	0.282	0.388
DEBIASED GLS	AvgCov S_0		0.820	0.802	0.791	0.797	0.817	0.613	0.580	0.546	0.546	0.570	0.432	0.392	0.352	0.353	0.354
	AvgCov S_0^c		0.968	0.970	0.973	0.977	0.980	0.960	0.962	0.967	0.970	0.974	0.953	0.955	0.958	0.965	0.967
	AvgLength		0.291	0.271	0.251	0.257	0.270	0.178	0.161	0.147	0.146	0.152	0.125	0.112	0.099	0.100	0.100
	AvgLength^c		0.291	0.271	0.251	0.257	0.270	0.178	0.161	0.147	0.146	0.152	0.125	0.112	0.099	0.100	0.100
DEBIASED LASSO	AvgCov S_0	500	0.868	0.888	0.915	0.932	0.938	0.663	0.729	0.836	0.893	0.927	0.459	0.533	0.700	0.801	0.874
	AvgCov S_0^c		0.984	0.983	0.981	0.979	0.974	0.971	0.971	0.967	0.969	0.968	0.960	0.962	0.961	0.959	0.961
	AvgLength		0.302	0.343	0.479	0.630	0.809	0.181	0.208	0.296	0.399	0.538	0.126	0.145	0.208	0.284	0.390
	AvgLength^c		0.301	0.343	0.479	0.629	0.809	0.181	0.208	0.296	0.399	0.538	0.126	0.145	0.208	0.284	0.390
DEBIASED GLS	AvgCov S_0		0.867	0.850	0.842	0.850	0.868	0.663	0.637	0.611	0.606	0.636	0.459	0.425	0.394	0.390	0.395
	AvgCov S_0^c		0.983	0.983	0.985	0.988	0.989	0.971	0.974	0.977	0.979	0.983	0.960	0.963	0.969	0.972	0.975
	AvgLength		0.300	0.281	0.263	0.272	0.288	0.180	0.164	0.151	0.150	0.160	0.126	0.113	0.103	0.102	0.103
	AvgLength^c		0.299	0.281	0.263	0.272	0.288	0.180	0.164	0.151	0.150	0.160	0.126	0.113	0.103	0.101	0.103

Table E.63: See description of Table 3, for $|S_0| = 7$. $d = 16$, DGP 3 ($\boldsymbol{x}_t \sim t_d$ and $\varepsilon_t \sim$ i.i.d. $\mathcal{N}(0, 1)$).

			$d = 16$														
		p/T			200					500					1000		
ϕ			0	0.5	0.8	0.9	0.95	0	0.5	0.8	0.9	0.95	0	0.5	0.8	0.9	0.95
DEBIASED LASSO	AvgCov S_0	100	0.895	0.902	0.923	0.931	0.934	0.841	0.866	0.909	0.930	0.945	0.760	0.802	0.877	0.908	0.925
	AvgCov S_0^c		0.958	0.956	0.957	0.958	0.956	0.954	0.954	0.953	0.953	0.954	0.951	0.952	0.952	0.952	0.952
	AvgLength		0.286	0.326	0.459	0.607	0.786	0.177	0.204	0.291	0.396	0.537	0.125	0.144	0.206	0.281	0.387
	AvgLength^c		0.286	0.326	0.459	0.607	0.786	0.177	0.204	0.291	0.396	0.537	0.125	0.144	0.206	0.281	0.387
DEBIASED GLS	AvgCov S_0		0.891	0.881	0.875	0.882	0.898	0.838	0.820	0.803	0.801	0.815	0.760	0.723	0.684	0.683	0.692
	AvgCov S_0^c		0.957	0.960	0.963	0.968	0.972	0.954	0.955	0.958	0.963	0.968	0.951	0.953	0.954	0.960	0.965
	AvgLength		0.284	0.262	0.239	0.243	0.256	0.177	0.159	0.143	0.142	0.146	0.125	0.112	0.098	0.097	0.099
	AvgLength^c		0.284	0.262	0.239	0.243	0.256	0.177	0.159	0.143	0.142	0.146	0.125	0.112	0.098	0.097	0.099
DEBIASED LASSO	AvgCov S_0	200	0.906	0.909	0.922	0.933	0.940	0.858	0.881	0.911	0.925	0.935	0.764	0.804	0.878	0.911	0.928
	AvgCov S_0^c		0.968	0.967	0.967	0.966	0.964	0.959	0.959	0.957	0.958	0.958	0.953	0.954	0.954	0.953	0.955
	AvgLength		0.291	0.333	0.466	0.615	0.795	0.179	0.205	0.293	0.397	0.538	0.125	0.144	0.207	0.282	0.387
	AvgLength^c		0.291	0.333	0.466	0.615	0.795	0.179	0.205	0.293	0.397	0.538	0.125	0.144	0.207	0.282	0.387
DEBIASED GLS	AvgCov S_0		0.901	0.891	0.884	0.893	0.906	0.856	0.840	0.824	0.825	0.843	0.765	0.730	0.696	0.700	0.704
	AvgCov S_0^c		0.967	0.969	0.972	0.977	0.980	0.959	0.961	0.965	0.968	0.973	0.953	0.955	0.959	0.965	0.968
	AvgLength		0.289	0.270	0.249	0.256	0.270	0.178	0.161	0.146	0.145	0.153	0.125	0.112	0.100	0.099	0.100
	AvgLength^c		0.289	0.269	0.249	0.256	0.270	0.178	0.161	0.146	0.145	0.153	0.125	0.112	0.100	0.099	0.100
DEBIASED LASSO	AvgCov S_0	500	0.910	0.916	0.924	0.929	0.934	0.869	0.897	0.923	0.936	0.946	0.788	0.833	0.894	0.918	0.936
	AvgCov S_0^c		0.983	0.982	0.981	0.978	0.974	0.970	0.970	0.966	0.968	0.967	0.960	0.961	0.960	0.958	0.960
	AvgLength		0.299	0.341	0.476	0.625	0.803	0.180	0.207	0.295	0.398	0.536	0.126	0.145	0.208	0.284	0.389
	AvgLength^c		0.299	0.341	0.476	0.625	0.803	0.180	0.207	0.295	0.398	0.536	0.126	0.145	0.208	0.284	0.389
DEBIASED GLS	AvgCov S_0		0.906	0.894	0.892	0.905	0.917	0.869	0.861	0.849	0.855	0.875	0.787	0.754	0.731	0.722	0.733
	AvgCov S_0^c		0.983	0.983	0.985	0.987	0.989	0.970	0.973	0.975	0.979	0.982	0.960	0.963	0.969	0.970	0.975
	AvgLength		0.297	0.277	0.261	0.270	0.285	0.180	0.164	0.148	0.150	0.160	0.126	0.113	0.103	0.100	0.104
	AvgLength^c		0.297	0.277	0.261	0.270	0.285	0.180	0.164	0.148	0.150	0.160	0.126	0.113	0.103	0.100	0.104

F Empirical Study Supplement

This Supplement provides a detailed discussion of the models examined in Section 7 of the main paper. The models are described below:

1. GLS LASSO:

$$\widehat{y}^{(\text{GLS})}_{\text{IP},t+1|t} = \widetilde{\mu}_{\text{IP},t} + \sum_{i=1}^{q} \widehat{\phi}_i y_{\text{IP},t-i+1} + \widetilde{\boldsymbol{x}}_t' \widehat{\boldsymbol{\beta}}_{\text{GLS}}, \tag{F.1}$$

   where $\widetilde{\mu}_{\text{IP},t}$ is the sample average of $\widetilde{y}_{\text{IP},t}$ and $\widetilde{y}_{\text{IP},t} = y_{\text{IP},t} - \sum_{i=1}^{q} \widehat{\phi}_i y_{\text{IP},t-i}$, $\widetilde{\boldsymbol{x}}_t = \boldsymbol{x}_t - \sum_{i=1}^{q} \widehat{\phi}_i \boldsymbol{x}_{t-i}$, $\widehat{\boldsymbol{\beta}}_{\text{GLS}}$ is a (121×1) vector of GLS LASSO estimated coefficients, $\widehat{\boldsymbol{\phi}} = (\widehat{\phi}_1, \dots, \widehat{\phi}_q)'$ are OLS estimates, such that $\widehat{\boldsymbol{\phi}} = (\widehat{\boldsymbol{e}}_{t-1}' \widehat{\boldsymbol{e}}_{t-1})^{-1} \widehat{\boldsymbol{e}}_{t-1}' \widehat{\boldsymbol{e}}_t$, $\widehat{\boldsymbol{e}}_t = y_{\text{IP},t} - \boldsymbol{x}_t' \widehat{\boldsymbol{\beta}}_{\text{GLS}}$. The order q is selected using information criteria, as discussed in Section 2, with the maximum number of lags set equal to $l_{\max} = 20$. The selection of the regularisation hyper-parameter is obtained via hv-block cross validation. In Section 5 we discuss how this cross-validation methods is implemented and also our choices for the relevant hyper-parameters required (h and v). This is the cross validation method that we use to select the regularisation hyper-parameters for all LASSO based methods presented below.

2. LASSO:

$$\widehat{y}^{(\text{LASSO})}_{\text{IP},t+1|t} = \widehat{\mu}_{\text{IP},t} + \boldsymbol{x}_t' \widehat{\boldsymbol{\beta}}_{\text{LASSO}},$$

   where $\widehat{\mu}_{\text{IP},t}$ is the sample average of $y_{\text{IP},t}$, $\widehat{\boldsymbol{\beta}}_{\text{LASSO}}$ is a (121×1) vector of LASSO estimated coefficients.

3. ARDL (q, r):

$$\widehat{y}^{(\text{ARDL})}_{\text{IP},t+1|t} = \widehat{\mu}_{\text{IP},t} + \boldsymbol{z}_{1,t}' \widehat{\boldsymbol{\beta}}_{\text{ARDL}},$$

   where $\widehat{\mu}_{\text{IP},t}$ is the sample average of $Y_{\text{IP},t}$, $\boldsymbol{z}_{1,t} = (y_{\text{IP},t}, \dots, y_{\text{IP},t-p}, \boldsymbol{x}_t, \dots, \boldsymbol{x}_{t-r})$ and $\widehat{\boldsymbol{\beta}}_{\text{ARDL}}$ is a $((q \times p) + r)$ vector of LASSO estimated coefficients of the distributed lags, that depends on the orders (p, r).

4. AR(q) model:

$$\widehat{y}^{(\text{AR})}_{\text{IP},t+1|t} = \widehat{\rho}_0 + \widehat{\rho}_1 y_{\text{IP},t} + \widehat{\rho}_2 y_{\text{IP},t-1} + \dots + \widehat{\rho}_q y_{\text{IP},t-q+1},$$

   $\widehat{\rho}_0, \widehat{\rho}_1, \dots, \widehat{\rho}_q$ are OLS estimates, the order q is selected using information criteria, with the maximum number of lags set equal to $l_{\max} = 20$.

5. FA-AR(q):

$$\widehat{y}^{(\text{FA- AR})}_{\text{IP},t+1|t} = \widehat{\rho}_0 + \widehat{\rho}_1 y_{\text{IP},t} + \widehat{\rho}_2 y_{\text{IP},t-1} + \dots + \widehat{\rho}_q y_{\text{IP},t-q+1} + \widehat{\boldsymbol{\lambda}}' \widehat{\boldsymbol{f}}_t,$$

   where $\widehat{\boldsymbol{f}}_t$ is an $(m \times 1)$ vector of the estimated unobserved common factors extracted from variables in $\boldsymbol{z}^{\text{s}}_{2,t}$, where $\boldsymbol{z}^{\text{s}}_{2,t}$ is the standardised version of $\boldsymbol{z}_{2,t} = (y_{\text{IP},t}, \boldsymbol{x}_t)$, obtained by subtracting its sample mean and dividing each series by its sample standard deviation. We set the number of factors, m, to 1. The unobserved factor, is estimated using the method of principal components, and $\widehat{\rho}_0, \ \widehat{\rho}_1, \ldots, \widehat{\rho}_q$ are OLS estimates. We select the order q using BIC, with the maximum number of lags set equal to $l_{\max} = 20$.