ΟΙΚΟΝΟΜΙΚΟ
ΠΑΝΕΠΙΣΤΗΜΙΟ
ΑΘΗΝΩΝ

ATHENS UNIVERSITY
OF ECONOMICS
AND BUSINESS

ΣΧΟΛΗ
ΟΙΚΟΝΟΜΙΚΩΝ
ΕΠΙΣΤΗΜΩΝ
SCHOOL OF
ECONOMIC
SCIENCES

ΤΜΗΜΑ
ΟΙΚΟΝΟΜΙΚΗΣ
ΕΠΙΣΤΗΜΗΣ
DEPARTMENT OF
ECONOMICS

Department of Economics
Athens University of Economics and Business

WORKING PAPER no. 06-2021

A regularization approach for estimation and variable selection in high dimensional regression models

Yiannis Dendramis, Liudas Giraitis, George Kapetanios

March 2021

A regularization approach for estimation and variable selection in high dimensional regression models

Y. Dendramis L. Giraitis G. Kapetanios

January 17, 2020

Abstract

Model selection and estimation are important topics in econometric analysis which can become considerably complicated in high dimensional settings, where the set of possible regressors can become larger than the set of available observations. For large scale problems the penalized regression methods (e.g. Lasso) have become the de facto benchmark that can effectively trade off parsimony and fit. In this paper we introduce a regularized estimation and model selection approach that is based on sparse large covariance matrix estimation, introduced by Bickel and Levina (2008) and extended by Dendramis, Giraitis, and Kapetanios (2018). We provide asymptotic and small sample results that indicate that our approach can be an important alternative to the penalized regression. Moreover, we also introduce a number of extensions that can improve the asymptotic and small sample performance of the proposed method. The usefulness of what we propose is illustrated via Monte Carlo exercises and an empirical application in macroeconomic forecasting.

Keywords: large dimensional regression, sparse matrix, thresholding, shrinkage, model selection.

*Yiannis Dendramis acknowledges support from the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement No 708501.

1 Introduction

In this paper, we study the problem of model selection and estimation for high dimensional datasets. The recent advent of large datasets has made this problem particularly important as the information which is extracted from large datasets is a key to various scientific discoveries or policy suggestions. In large datasets, conventional statistical and econometric techniques such as regression estimation fail to work consistently due to the dimensionality of the examined economic relationships. For instance, in a linear economic relationship we frequently have T observations of a dependent variable (y) as a function of many potential predictors (p predictors). When the number of predictors (p) is large or larger than the temporal dimension (T) then a regression with all available covariates becomes extremely problematic if not impossible. In this paper we suggest new, theoretically valid techniques to handle situations like the former.

Two main strands have dominated the relevant literature, so far: Penalized regression methods (see e.g. Tibshirani (1996), Fan and Li (2001), Zhou and Hastie (2005), Lv and Fan (2009), Efron et al. (2004), Bickel et al. (2009), Candes and Tao (2007), Zhang (2010), Antoniadis and Fan (2001), Fan and Lv (2013), Fan and Tang (2013), Hunter and Li (2005), Zou and Zhang (2009), Zhang and Huang (2008), Fan et al. (2014)) and greedy methods (see e.g. Friedman (2001), Friedman et al. (2000), Buhlmann (2006), Bühlmann and Hothorn (2007), Hothorn et al. (2010)), whose origins come from the machine learning literature. In penalized regression methods all the available regressors are considered at the same time, while a penalty function shrinks the estimated parameter coefficients, though a choice of a tuning parameter. The main drawback of this approach is that the applied researcher has to choose a specific penalty function and a tuning parameter. Different penalty functions lead to diverse theoretical and small sample performance, while in applied research there is no way to choose in advance an optimal penalty function.

For greedy methods, the idea is to screen in advance some covariates which are considered as more likely to explain the dependent variable. This is done based on their individual ability of the regressors to explain the dependent variable. Machine learning methods such us boosting, regression trees, and step wise regressions are casted in this category. The main drawback of this strand of literature is the scarcity of theoretical results for most of the methods. Other important sequential methods for model selection are Fithian et al. (2015), Fithian et al. (2014), and Tibshirani et al. (2014).

In a recent paper, Fan and Lv (2008) have proposed a combination of the two previous strands. They propose a two step procedure in which, in the first step all regressors are ranked according to their absolute correlations with the dependent variable, and a fixed proportion of regressors is selected. This is referred to as sure independence screening (SIS). In the second step a penalized regression approach is applied on the selected subset of regressors. Similar variable screening mechanisms have been developed in Hall et al.

(2009), Hall and Miller (2009), Fan et al. (2011), Li et al. (2012b) Li et al. (2012a), Huang et al. (2008), Fan and Song (2010), and Fan et al. (2009).

Most of the above literature assumes regressors which are either independent and identically distributed (IID) or deterministic processes. Generalizations to stochastic regressors is not straightforward, requiring conditions such as the spark condition of Donoho and Elad (2003) and Zheng et al. (2014).

The main idea of our approach is to first propose a theoretically valid regularized ordinary least squares (OLS) estimator, on the full set of available regressors. This is done through a regularized, threshold estimator of covariance and precision matrices, that is applicable to large datasets.

Several regularization techniques for improved estimation of large covariance matrices have been proposed in the literature. These include the works of Chen et al. (2013), Zhou et al. (2010), Kolar and Xing (2011), Fan et al. (2013), Fan et al. (2016), Han and Liu (2017), Bickel and Levina (2008), Cai and Liu (2011), Ledoit and Wolf (2004), and Abadir et al. (2014). All this work assumes that the volatility process is a deterministic function or a constant. Stochastic volatility is considered in Bickel et al. (2013), and Dendramis et al. (2018) while for structure free estimators see Pourahmadi (2013) .

Our proposed estimator retains good theoretical properties, even in the case that the number of available regressors is large, allowing the extraction of true signal covariates. Crucially, all the available regressors are considered in a single regression step. To enhance the small sample performance of this estimator we provide extensions related to the hypothesis testing literature. To this end, and following the work of Chudik et al. (2018), we provide an asymptotically attractive testing procedure which can be used as an extended screening mechanism that can extract true signals, and significantly improve the small sample performance of the benchmark estimator. This is done through a Bonferroni type critical value, designed to minimize the Type I error of the multiple testing procedure. As a final step, all covariates that are considered as significant can be added as joint determinants of the dependent variable in a final multiple regression step. Similar multiple testing approaches in high dimensional settings have been also studied in Fan and Han (2017), Fan et al. (2012), Guo et al. (2019), van de Geer and Stucky (2015) and Wasserman and Roeder (2009).

Our approach provides some important advantages over the alternatives. First we do not need to rely and choose a penalty function from the large set of proposals which are available in the literature. Second we do not need to examine each regressor separately, as with greedy methods. Third, our method does not need standardized regressors, as in penalized regression methods. Fourth, it can be extended to an inferential procedure that can be directly linked to classical statistical analysis.

In its most general, our preferred approach is eclectic as it combines powerful features of existing methods such as screening and multiple testing, which have been shown to have excellent properties in high dimensional regression settings, with our newly proposed regularized OLS. The latter provides a crucial added element of jointly considering all the available regressors which allows for a more refined multiple testing analysis.

The paper is structured as follows: Section 2 introduces the basic estimator for the large dimensional regression problem and presents its asymptotic properties. Section 3 extends the basic estimator, by proposing a testing approach that is able to enhance the performance of the model selection and estimation mechanism. Theoretical results are also derived in this section. Section 4 presents a number of further extensions and improvements that can be useful in diverse real world situations. Section 5 presents in detail the cross validation schemes which are essential for the application of the proposed methodologies. Section 6 gives the details of the Monte Carlo experiments and the simulation results. Section 7 presents the empirical application, and Section 8 concludes. Technical proofs are relegated to appendices.

2 Estimation by regularization

Consider the case that we observe a sample from a dependent variable $\{y_t\}_{t=1,..,T}$ and a p-variate sample $\{x_t\}_{t=1,..,T}$ of possible regressors of y_t. Some of them truly explain the dependent y_t. Given the increased availability of large datasets, it is possible to have p and T increasingly large, and even $p >> T$. In this setting, the dependent variable y_t is assumed to be explained by a relatively small subset of regressors on the set $\{x_{it}\}_{i=1,..,p}$, where x_{it} is the t-th observation ($t \in \{1, 2, .., T\}$) for covariate i ($i \in \{1, 2, .., p\}$). In vector notation, the dependent variable $\boldsymbol{y}$ and the full set of possible regressors $\{\boldsymbol{x}_i\}_{i=1,..,p}$ are T dimensional column vectors, the $\{\boldsymbol{x}_t\}_{t=1,..,T}$ are p-dimensional row vectors, while the $T \times p$ matrix $\boldsymbol{x}$ includes all available potential covariates.

Our intention is to find the subset of covariates $\{\boldsymbol{x}_i\}_{i=1,..,p}$ that truly explain the dependent $\boldsymbol{y}$, and estimate their corresponding regression coefficients. This is classified as a model selection and estimation problem, for high-dimensional regression and focuses on the estimation of a sparse regression model of the form

$$y_t = \sum_{i=1}^{p} \beta_i x_{it} + u_t, t = 1, ., T, \tag{1}$$

where u_t is the error with mean zero and variance σ_u^2. Without loss of generality we assume that the first k regressors truly explain the dependent variable y_t ($\{\beta_i\}_{i=1}^k \neq 0$) and the last $p-k$ are poor noise variables ($\{\beta_i\}_{i=k+1}^p = 0$). Estimating the parameter vector β_i for each regressor i in model (1) by classical OLS is not a feasible option. This is because all the theoretical properties of OLS rely on the assumption that p is small and fixed. This problem can be further complicated when we allow for cross sectional correlation, that is frequently observed in large panels comprised of macroeconomic and/or financial series. Non-sparse correlation structures of $\boldsymbol{x}_t$ will inflate the correlation between the dependent variable $\boldsymbol{y}$ and a possible regressor $\boldsymbol{x}_i$, which does not truly explain $\boldsymbol{y}_t$, exacerbating the uncertainty about the true regression model.

Our approach starts from the simple fact that, under mild existence and regularity conditions, $\boldsymbol{\beta} = (\beta_1, ..., \beta_p)' = \boldsymbol{\Sigma}_x^{-1}\boldsymbol{\Sigma}_{xy}$, where $\boldsymbol{\Sigma}_x$ and $\boldsymbol{\Sigma}_{xy}$ denote the variance of $\boldsymbol{x}_t$ and the covariance of $\boldsymbol{x}_t$ and y_t, respectively. Therefore the large regression problem is converted into a large covariance matrix estimation problem, which is addressed by a large portion of existing literature. Our approach, relies on the estimation of the large dimensional covariance of the p_z $(= p+1)$-dimensional random vector $\boldsymbol{z}_t = (y_t, \boldsymbol{x}_t)$, denoted as $\boldsymbol{\Sigma}_z$, of dimension $p_z \times p_z$. When p_z increases with T, estimation of $\boldsymbol{\Sigma}_z$ is particularly demanding since the number of estimated objects rises as a square of the dimension of the dataset, leading to a large amount of aggregated estimation error. Naturally, the sample estimate $\widehat{\boldsymbol{\Sigma}}_z$ of $\boldsymbol{\Sigma}_z$, defined as $\widehat{\boldsymbol{\Sigma}}_z = T^{-1}\sum_{t=1}^T (\boldsymbol{z}_t - \overline{\boldsymbol{z}})' (\boldsymbol{z}_t - \overline{\boldsymbol{z}})$, $\overline{\boldsymbol{z}} = T^{-1}\sum_{t=1}^T \boldsymbol{z}_t$, performs very poorly in this case. Several regularization techniques for improved estimation of large covariance matrices have been proposed. For an excellent review of structure-based estimators of covariance and precision matrices see the paper by Fan et al. (2016).

We consider p_z-dimensional vectors $\boldsymbol{z}_t$ whose variance, $\boldsymbol{\Sigma}_z$, is assumed to belong to the following class of sparse covariance matrices

$$\boldsymbol{\Sigma}_z \in Q\left(n_{p_z}, K\right) = \left\{ \boldsymbol{\Sigma} : \sigma_{ii} \leq K, \sum_{i=1}^{p_z} 1\left\{\sigma_{ij} \neq 0\right\} \leq n_{p_z} \right\} \tag{2}$$

The sparsity parameter, n_{p_z}, is the maximum number of non zero row elements of $\boldsymbol{\Sigma}_z$ and it is assumed that it does not grow too fast with p_z. When $\boldsymbol{z}_t$ is iid and light tailed random variable, hard and adaptive thresholding introduced by Bickel and Levina (2008) and Cai and Liu (2011) are the two standard approaches to regularize the $\boldsymbol{\Sigma}_z$. For instance, regularization by hard thresholding is based on the idea of setting the elements of $\mathbf{\Sigma}_z$, whose absolute values are smaller than some value (a threshold), equal to zero. Regularising the sample covariance matrix $\widehat{\mathbf{\Sigma}}_z$ by hard thresholding yields the estimate

$$T_\lambda\left(\widehat{\mathbf{\Sigma}}_z\right) = \left(\widehat{\sigma}_{ij} I\left(|\widehat{\sigma}_{ij}| > \lambda_{ij}\right)\right) \tag{3}$$

with $\widehat{\sigma}_{ij} = \left[\widehat{\mathbf{\Sigma}}_z\right]_{ij}$, and $I\left(|\widehat{\sigma}_{ij}| > \lambda_{ij}\right) = 1$ when $|\widehat{\sigma}_{ij}| > \lambda_{ij}$ and 0 otherwise. Estimator (3) sets a lower bound for the elements of the estimated covariance $\widehat{\mathbf{\Sigma}}_z$. Elements of $\widehat{\mathbf{\Sigma}}_z$ bellow that bound are set equal to zero. In Bickel and Levina (2008) the λ_{ij} is universal for all *i*,*j*, with

$$\lambda_{ij} = \lambda = \kappa\sqrt{\frac{\log p_z}{T}}, \kappa > 0 \tag{4}$$

A universal thresholding rule essentially treats the problem as if all $\sigma_{ii} = K$, when selecting the λ. In adaptive thresholding, introduced by Cai and Liu (2011), the λ_{ij} depends on the i, j-th entry of the matrix $\widehat{\mathbf{\Sigma}}_z$, capturing the variability of individual variables, instead of selecting a universal upper bound. Entry-dependent thresholds that automatically adapt to the variability of the individual entries of the sample covariance matrix $\widehat{\mathbf{\Sigma}}_z$, are particularly useful when the diagonal elements of $\mathbf{\Sigma}_z$ vary over a wide range, and there is no a priori, obvious upper bound for them. In this case thresholding becomes

$$\begin{aligned} \lambda_{ij} &= \kappa\sqrt{\frac{\widehat{\theta}_{ij}\log p_z}{T}}, \kappa > 0 \\ \widehat{\theta}_{ij} &= \frac{1}{T}\sum_{t=1}^{T}\left[\left(z_{i,t} - \overline{z}_i\right)\left(z_{j,t} - \overline{z}_j\right) - \widehat{\sigma}_{ij}\right]^2 \end{aligned} \tag{5}$$

where the introduction of $\widehat{\theta}_{ij}$ accounts for the variability of the diagonal elements of $\mathbf{\Sigma}_z$.

Both universal and adaptive threshold λ_{ij} depends on a tuning parameter, κ, which does not affect the asymptotic performance of the estimator, but it can have significant impact on the small sample performance. In practice, κ can be chosen through cross validation, as discussed later in the paper. Procedures based on other thresholding operators can be defined (e.g. soft, lasso, etc), but they have similar properties to the hard thresholding, asymptotically, although they may differ in finite samples.

Extending the results on thresholding estimators, Dendramis et al. (2018) show that the nice theoretical results of Bickel and Levina (2008) are also valid when z_t is a stationary α-mixing process. Moreover, it is proven that z_t can be also allowed to be a heavy tailed random variable. The last result comes at a cost on the rate at which p_z is allowed to increase, compared to T.

Given the theoretically valid regularized covariance matrix estimator of $\boldsymbol{z}_t$, for large p_z and T dimensions, the regularized estimator of $\boldsymbol{\beta}$ is defined as

$$\hat{\boldsymbol{\beta}}_\lambda = T_\lambda \left(\hat{\boldsymbol{\Sigma}}_x\right)^{-1} T_\lambda(\hat{\boldsymbol{\Sigma}}_{xy}) \tag{6}$$

with $T_\lambda \left(\hat{\boldsymbol{\Sigma}}_x\right)$, $T_\lambda(\hat{\boldsymbol{\Sigma}}_{xy})$ being the submatrice and subvector of thresholded estimate $T_\lambda \left(\widehat{\boldsymbol{\Sigma}}_z\right)$ that correspond to the regressors $\boldsymbol{x}_t$ covariance matrix, and covariance vector of y_t with $\boldsymbol{x}_t$, respectively. The T_λ (.) is defined in (3). As we will show in the next section, estimator (6), be theoretically valid at the large p, T setting.

3 Theoretical properties

In this section, we consider the asymptotic properties of $\hat{\boldsymbol{\beta}}_\lambda$ defined in (6). We show that under a mild set of assumptions, $\hat{\boldsymbol{\beta}}_\lambda$ can achieve consistency with optimal convergence rates. Consistency is subject to the following set of assumptions:

Assumption 1 *The error term u_t in model (1) is a martingale difference sequence (MDS) with respect to the filtration $F^u_{t-1} = \sigma(u_{t-1}, u_{t-2}, ..)$ with zero mean and a constant variance* $0 < \sigma^2 < C < 8$.

Assumption 2 *There exist sufficiently large positive constants D_0, D_1, D_2, D_3 and $s_u, s_z > 0$ such that z_{it} and u_t satisfy the following conditions*

$$Pr(|z_{it}| \geq \zeta) \leq D_0 \exp(-D_1 \zeta^{s_z}) \tag{7}$$

$$Pr(|u_t| \geq \zeta) \leq D_2 \exp(-D_3 \zeta^{s_u}) \tag{8}$$

This can be weakened to a polynomial rate, following Dendramis et al. (2018) but at a cost of smaller p.

Assumption 3 *Regressors are uncorrelated with the errors $E(x_{it}u_t|F_{t-1}) = 0$ for all $t = 1, 2, .., T, i = 1, .., p$ and $F_{t-1} = F^u_{t-1} \cup F^x_{t-1} = \sigma(u_{t-1}, u_{t-2}, ..) \cup \{\cup^p_{j=1} \sigma(x_{jt-1}, x_{jt-2}, ..)\}$.*

Assumption 4 *The number of true regressors k is finite.*

Assumption 5 $E(x_{it}x_{jt} - E(x_{it}x_{jt})|F_{t-1}) = 0$*, for all i, j.*

Assumption 1 allows for some dependence through e.g. an arch process but no serial correlation. Assumption 2 can be relaxed to allow the vector z_t to be heavy tailed distributed. This is a departure from the usual in the literature exponentially declining bound for the probability tails but this comes at the cost of smaller rates in dimension p_z. In Dendramis et al. (2018) both exponentially and polynomially declining bounds are studied. Assumption 3 is a MDS assumption on the $x_{it}u_t$ with respect to F_{t-1}. The requirement in Assumption 4, can be relaxed to $k < T$. The rest of the assumptions are technical requirements which are necessary for the proof of Theorem 1 and are analogous to the assumptions needed in the recent literature (see e.g. Chudik et al. (2018)).

Theorem 1 *Consider the DGP defined in (1), and suppose that assumptions 1-5 hold. Then for sufficiently large κ, the regularized estimator $\hat{\boldsymbol{\beta}}_\lambda$ satisfies*

$$\left\|\hat{\boldsymbol{\beta}}_\lambda - \boldsymbol{\beta}\right\| = O_p\left(n_{p_z}^{3/2}\lambda_{p_z}\right) \tag{9}$$

A probability bound for $\left\|\hat{\boldsymbol{\beta}}_\lambda - \boldsymbol{\beta}\right\|$ is given by

$$\Pr\left(\left\|\hat{\boldsymbol{\beta}}_\lambda - \boldsymbol{\beta}\right\| > n_{p_z}^{3/2}\lambda_{p_z}\right) \leq p_z^2 D_1 \exp\left(-D_2 T \lambda_{p_z}^2\right) + p_z D_3 \exp\left(-D_4 T \lambda_{p_z}\right) \tag{10}$$

for sufficiently large constants D_1, D_2, D_3, and D_4.

Theorem 1 states that under reasonable assumptions on the DGP of model (1), the proposed estimator converges to the true $\boldsymbol{\beta}$. This means that asymptotically we will be able to extract the true signals ($\beta_i \neq 0$) and discard the noise variables ($\beta_i = 0$). Theorem 1 holds for $n_{p_z}^{3/2}\lambda_{p_z} = o(1)$, which can happen when n_{p_z} is a fixed constant, or when it is an increasing sequence with $n_{p_z} < (T/\log p_z)^{1/3}$.

4 Methodological refinements using multiple testing

In order to further boost the performance of our basic regularized estimator $\hat{\boldsymbol{\beta}}_\lambda$ we introduce a testing procedure. To this end, given the estimate $\hat{\boldsymbol{\beta}}_\lambda$, for each covariate x_i, $i \in \{1, .., p\}$ we define the following ratio

$$\xi_{\lambda,i} = \frac{\hat{\beta}_{\lambda,i}}{\widehat{\gamma}_{\lambda,i}} \tag{11}$$

with $\widehat{\gamma}_{\lambda,i} = \sqrt{s^2\left[\left(T \times T_\lambda\left(\widehat{\boldsymbol{\Sigma}}_x\right)\right)^{-1}\right]_{ii}}$, $s^2 = \frac{1}{T}\sum_{t=1}^{T}\widehat{u}_t^2$, $\widehat{u}_t = y_t - x_t\widehat{\beta}_\lambda$ and $[\boldsymbol{A}]_{ij}$ is the *i*,*j*-th element of the matrix $\boldsymbol{A}$, for $i = 1, .., p$. The data implied parameter $\widehat{\gamma}_{\lambda,i}$ is a normalization constant and it is the vehicle that allow us to introduce the testing procedure. When p is small relative to T (e.g. when $p < T^{-1/2}$) the $\widehat{\gamma}_{\lambda,i}$ is the standard error of the estimate $\hat{\beta}_i$, but in general it is not. Then, the testing procedure is defined as

$$I\left(\widehat{\beta_{\lambda,i} \neq 0}\right) = \begin{cases} 1 \text{ when } \xi_{\lambda,i} > c_p \\ 0 \text{ otherwise} \end{cases} \tag{12}$$

with $c_p = \Phi^{-1}\left(1 - \frac{\alpha}{cp^{\pi}}\right)$, $c_p = O\left(\left(\pi \log(p)\right)^{1/2}\right) = o\left(T^{c_0}\right)$, $\forall c_0 > 0$, $\Phi^{-1}(.)$ is the quantile function of the normal distribution and α is the nominal size of the individual test, to be set by the researcher (e.g. $\alpha = 1\%$ or 5%), and the constants π, c satisfy $\pi > 0$, $c < \infty$. The critical value c_p is a Bonferroni type critical value controlling the family wise error of the multiple testing problem to be at the level of $\alpha\%$. The constants c, π do not affect the asymptotic results but in small samples the testing performance can be boosted by these parameters. In practise on can set them both 1, or choose them through cross validation. In our empirical and simulation exercise we set them both equal to 1. Of course, other c_p values can be used, such as those proposed by Holm (1979), Benjamini and Hochberg (1995), or more recently by Gavrilov et al. (2009), which are designed to control the family-wise error rate, and are expected to work similarly.

The introduced testing procedure works as follows: When $\xi_{\lambda,i}$ is bounded in T , then x_{it} is considered as a noise variable. In the opposite case, the $\xi_{\lambda,i}$ diverges in T, meaning that x_{it} has power in explaining the y_t, i.e. x_{it} can be considered as a true signal variable.

In the large dimensional regression literature two critical measures of how reasonable a method can separate the true regressors from the noise variables are the true positive rate (TPR) and false positive rate (FPR) defined bellow

$$\begin{aligned} TPR_{p,T} &= \frac{\#\left\{i : \hat{\beta}_{\lambda,i} \neq 0 \text{ and } \beta_i \neq 0\right\}}{\#\left\{i : \beta_i \neq 0\right\}} \\ FPR_{p,T} &= \frac{\#\left\{i : \hat{\beta}_{\lambda,i} \neq 0 \text{ and } \beta_i = 0\right\}}{\#\left\{i : \beta_i = 0\right\}} \end{aligned} \tag{13}$$

The $TPR_{p,T}$ measures the percentage of true signals that we correctly identify as true signals (should converge to 1) while the $FPR_{p,T}$ measures the percentage of noise variables that we incorrectly identify as true signals (should converge to zero). We enrich our methodological contribution of Theorem 1 by proposing the following algorithm.

```
Algorithm 1: Support Recovery
  1. Estimate the parameter vector of model (1) by the basic estimator $\hat{\beta}_{\lambda,i}$ defined in (6);
  2. Filter the covariates $\{x_i\}_{i=1}^{p}$ that truly explain $y_t$ by the testing approach (12);
```

Theorem 2 proves that the previous two step procedure can result to $TPR_{p,T}$, $FPR_{p,T}$ that asymptotically converge to 1 and 0 respectively, meaning that for suitable rates of p and T we can correctly identify the true model.

Theorem 2 *Consider the DGP defined in (1), and suppose that assumptions 1-5 hold. Then, Algorithm 1 can identify the true model, or equivalently, for sufficiently large constants D_1, D_2, D_3, D_4, for the true positive rate we have that*

$$\begin{aligned}E\left[TPR_{p,T}\right] &= k^{-1}\sum_{i=1}^{k}\Pr\left[|\xi_{\lambda,i}| > c_T \mid \beta_i \neq 0\right] \\ &\geq 1 - O\left(p_z^2 D_1 \exp\left(-D_2 T \lambda_{p_z}^2\right) + p_z D_3 \exp\left(-D_4 T \lambda_{p_z}\right)\right)\end{aligned}$$

or equivalently $TPR_{p,T} \to_p 1$, and the false positive rate

$$\begin{aligned}E\left[FPR_{p,T}\right] &= (p-k)^{-1}\sum_{i=k+1}^{p}\Pr\left[|\xi_{\lambda,i}| > c_T \mid \beta_i = 0\right] \\ &= \frac{1}{p-k} O\left(p_z^2 D_1 \exp\left(-D_2 T \lambda_{p_z}^2\right) + p_z D_3 \exp\left(-D_4 T \lambda_{p_z}\right)\right) \\ &= O\left(p_z D_1 \exp\left(-D_2 T \lambda_{p_z}^2\right) + D_3 \exp\left(-D_4 T \lambda_{p_z}\right)\right)\end{aligned}$$

or equivalently $FPR_{p,T} \to_p 0$

As a final theoretical contribution we also provide theorem 3, which states that if the previously defined two step procedure is followed by a final regression step, the selected covariates and parameter vector converges to the true one, asymptotically. [GK take the algorithm mout of theorem 2 and put it separate before the theorem, YD not sure what you mean]

Theorem 3 *Consider the DGP defined in (1), and suppose that assumptions 1-5 hold. Then, Algorithm 2, for sufficiently large constants D_1, D_2, D_3, D_4, results to consistent parameter estimates,*

```
Algorithm 2: Filtering and Estimation
1. Estimate the parameter vector of model (1) by the basic estimator $\hat{\beta}_{\lambda,i}$ defined in (6);
2. Filter the covariates $\{\boldsymbol{x}_i\}_{i=1}^{p}$ that truly explain $y_t$ by the testing approach (12). Let $\boldsymbol{x}_{t,-J}$ denote the $\boldsymbol{x}_t$, excluding columns $J = \left\{j : I\left(\widehat{\beta_{\lambda,j} \neq 0}\right) = 0\right\}$. These are considered as the important covariates of model (1).;
3. Run a final regression on the $\widetilde{\boldsymbol{x}}_{t,-J}$ covariates, $\hat{\boldsymbol{\beta}}_{\lambda,f} = \left(\boldsymbol{x}'_{T,-J}\boldsymbol{x}_{T,-J}\right)^{-1}\boldsymbol{x}_{T,-J}\boldsymbol{y}$;
```

with rates of convergence given by

$$\left\|\hat{\boldsymbol{\beta}}_{\lambda,2s}-\boldsymbol{\beta}\right\| = O_p\left(\frac{1}{\sqrt{T}}\right)+O_p\left(p_z^2 D_1 \exp\left(-D_2 T\lambda_{p_z}^2\right)+p_z D_3 \exp\left(-D_4 T\lambda_{p_z}\right)\right)$$

5 Further extensions

5.1 The sure screening extension

The Sure Independence Screening (SIS) developed by (Fan and Lv (2008)) is a simple, powerful method for variable selection when p and T are large. According to this, marginal correlations of potential covariates $\{\boldsymbol{x}_i\}_{i=1}^{p}$ with the dependent variable y_t are ranked in ascending order. The k^* covariates with the highest absolute correlation are considered as the most likely true regressors of dependent variable y_t, with $k^* << p$. As a final step the researcher can use a penalized regression method (e.g. Lasso) to estimate the parameter vector $\boldsymbol{\beta}$ in (1). In practice the researcher has to choose the threshold k^*, which sets an upper limit on the number of likely covariates and the final step penalized regression method.

In our framework the SIS can result to the dimensionality reduction which is important for better small sample performance. To this end, we first screen the important variables by SIS for a predetermined level of k^*. Then, we apply the basic estimator $\hat{\beta}_{\lambda,i}$ defined in (6), on the selected k^* regressors from the set $\{\boldsymbol{x}_i\}_{i=1}^{p}$ and the dependent variable y_t. Instead of this, one can also apply the developed testing methodology (see Theorem 2) and the final regression step (see Theorem 3).

5.2 Non sparse covariance matrix estimator

The proposed basic estimator $\hat{\beta}_{\lambda,i}$ defined in (6), depends crucially, on the large dimensional covariance matrix of $\boldsymbol{z}_t = (y_t, \boldsymbol{x}_t)$ which is assumed to be sparse. Sparsity of $\boldsymbol{\Sigma}_z$, is an essential assumption, necessary for acceptable rates of convergence of $\hat{\beta}_{\lambda,i}$. In the literature there have been proposed other consistent covariance matrix estimators which do not need the condition (2) for $\boldsymbol{\Sigma}_z$. This is an important development as there are applications in economics in which the sparsity belief might be inappropriate, owing to the presence of common factors of $\boldsymbol{z}_t$. In (Fan et al. (2013)) the leading assumption for $\boldsymbol{\Sigma}_z$ is that the first q eigenvalues are spiked enough and grow fast, at a rate $O(p)$. Then, the common and idiosyncratic components of $\boldsymbol{z}_t$ can be identified, and in addition, principal component analysis (PCA) on the sample covariance matrix can consistently estimate the space spanned by the eigenvectors of $\boldsymbol{\Sigma}_z$. Let $\widehat{\mu}_1 \geq \widehat{\mu}_2 \geq ..\widehat{\mu}_p$ be the ordered eigenvalues of the sample covariance matrix $\widehat{\Sigma}_z$ and $\{\widehat{\boldsymbol{v}}_j\}_{j=1}^{p}$ be their corresponding eigenvectors. The sample covariance of $\boldsymbol{z}_t$ has the following spectral decomposition

$$\widehat{\Sigma}_z = \sum_{i=1}^{q} \widehat{\mu}_i \widehat{\boldsymbol{v}}_i \widehat{\boldsymbol{v}}_i' + \widehat{\boldsymbol{R}}_q \tag{14}$$

where $\widehat{\boldsymbol{R}}_q = \sum_{i=q+1}^{p} \widehat{\mu}_i \widehat{\boldsymbol{v}}_i \widehat{\boldsymbol{v}}_i' = \left(\widehat{r}_{ij}\right)_{p\times p}$ is the principal orthogonal complement, and q is the number of diverging eigenvalues of $\boldsymbol{\Sigma}_z$. When $\widehat{\boldsymbol{R}}_q$ is sparse the estimator of $\boldsymbol{\Sigma}_z$ is then defined as

$$\widehat{\boldsymbol{\Sigma}}_{z,q} = \sum_{i=1}^{q} \widehat{\mu}_i \widehat{\boldsymbol{v}}_i \widehat{\boldsymbol{v}}_i' + T_\lambda\left(\widehat{\boldsymbol{R}}_q\right) \tag{15}$$

with $T_\lambda(.)$ being the regularized estimator defined in (3) and the threshold λ_{ij} defined according to (4) or (5). In this setting, for a given value of q the basic estimator $\hat{\boldsymbol{\beta}}_\lambda$ defined in (6), can be trivially adjusted for the large covariance estimator $\widehat{\boldsymbol{\Sigma}}_{z,q}$, given in (15) by replacing the (3) with (15). When q is unknown, then it can be calibrated via cross validation, or use information criteria for the optimal $\widehat{q}$ (see e.g. Bai and Ng (2002)).

Preliminary simulations for this extension do not suggest significant improvements over the basic estimator and testing as it is defined in sections 2 and 3.

5.3 Time varying parameter extension

Interestingly, our approach for estimating the large parameter vector can be extended to the case of time varying parameter models. Assuming a time varying parameter structure, for the dependent $\{y_t\}_{t=1,..,T}$ and the set of possible regressors $\{\boldsymbol{x}_t\}_{t=1,..,T}$, such as

$$y_t = \sum_{i=1}^{p} \beta_{it} x_{it} + u_t, u_t \sim N\left(0, \sigma_u^2\right), t = 1, ., T \tag{16}$$

where $\{\beta_i\}_{i=1}^{k} \neq 0$, $\{\beta_i\}_{i=k+1}^{p} = 0$, and p,T are large, as in our basic model (1), suggests an additional difficulty: the large parameter vector $\boldsymbol{\beta}_t = \left(\beta_{1t}, \beta_{2t}, .., \beta_{pt}\right)$ is non constant over time. Following our previous exposition, under mild existence and regularity conditions, we are able to define true parameter vector as $\boldsymbol{\beta}_t = \boldsymbol{\Sigma}_{x,t}^{-1}\boldsymbol{\Sigma}_{xy,t}$, where $\boldsymbol{\Sigma}_{x,t}$ and $\boldsymbol{\Sigma}_{xy,t}$ denote the time varying variance of $\boldsymbol{x}_t$ and the time varying covariance of $\boldsymbol{x}_t$ and y_t, respectively. Now, the problem of estimating a large time varying parameter vector, is equivalent to the estimation of a large time varying covariance matrix $\boldsymbol{\Sigma}_{z,t} = cov\left(\boldsymbol{z}_t\right)$, for $\boldsymbol{z}_t = \left(y_t, \boldsymbol{x}_t\right)$ which has been particularly addressed in Dendramis et al. (2018). In this paper, it is shown that a large $\boldsymbol{\Sigma}_{zt}$ that belongs to the class of sparse large dimensional matrices, as it is given in (2), for each time t, can be be estimated for cases in which p and T are increasingly large, with even $p > T$. Then, the testing results of Theorem 2 and 3 can be also altered to allow for large time varying covariance cases. The latter extensions constitute topics in our current research agenda.

6 Implementation and Cross Validation Schemes

The application of the proposed methods require the selection of a number of tuning parameters. For instance, the basic estimator $\hat{\boldsymbol{\beta}}_\lambda$ given in (1) depends on the value of κ, while extensions of it, adapted to the (15) or (19), require selection of additional tuning parameters. Moreover in the developed testing approach one can also calibrate the π parameter, of the Bonferroni critical value (see (12)), as it is often done in the literature, in order obtain better small sample performance.

As is well known, estimating a model as well as calibrating tuning parameters, on the same dataset leads to over-fitting. Cross validation (CV) is designed to address this issue, starting from the remark that evaluating a model on new data yields a better estimate of its performance (see Stone (1977) or Geisser (1975)). The main idea in CV is to split data once, or several times, for estimating the outcome of each possible value for the tuning parameter. Part of data (the training sample) is used for model training, and the remaining part (the validation sample) is used for estimating its performance. The validation sample plays the role of new data. Then, CV selects the tuning parameter with the best performance on the validation sample. CV avoids over-fitting because the training sample is independent from the validation sample (at least when data are i.i.d.). CV assumes that data are identically distributed, and training and validation samples are independent, although both assumptions can be relaxed. When data are dependent, CV can be modified, in order to account for this. Interested readers are referred to the review by Opsomer et al. (2001) on model selection in non-parametric regression with dependent data.

To obtain values for all these parameters we propose a simple CV scheme, whose the CV objective function focuses directly on the linear regression model (1). This is done through K-fold cross validation. According to this, the original sample is randomly partitioned into K roughly equally sized subsamples. Let ζ_i, $i \in \{1, .., K\}$ denote the relevant subsamples. Pick up a partition ζ_i, and use the $K-1$ partitions ζ_j, $j \neq i$ for $j \in \{1, .., K\}$ to estimate the regression coefficient, $\widehat{\beta}_{\gamma,j}$, where γ denotes the tuning parameter (e.g. κ or π). The discarded partition ζ_i is then used as a testing sample, on which we compute the average squared error

$$e_{i,\gamma} = \frac{1}{\{\#\text{ obs. in part. } \zeta_i\}} \sum_{t \in \zeta_i} \left(y_t - \boldsymbol{x}_t \widehat{\beta}_{\gamma,j} \right)^2 \tag{17}$$

This is repeated for all partitions ζ_i, $i \in \{1, .., K\}$. The optimal tuning parameter is obtained as the one the minimizes the average error on all the validation samples considered, that is

$$\widehat{\gamma} = \arg\min_{\gamma \in \Gamma} \frac{1}{K} \sum_{i=1}^{K} e_{i,\gamma} \tag{18}$$

over a suitable parameter vector space Γ for the vector of interest, γ. When $K = T - 1$, with T being the total number of observations, we have the leave on out cross validation. For instance, in the basic estimator $\hat{\boldsymbol{\beta}}_\lambda$ we set $\gamma \equiv \lambda$ and the parameter space Γ includes κ-points which imply invertibility for $T_\lambda \left(\hat{\boldsymbol{\Sigma}}_x \right)$ (see also discussion in section 7).

7 Computational aspects and invertibility conditions

The proposed basic estimator $\hat{\boldsymbol{\beta}}_\lambda$ involves several computational challenges. The main challenge involves the estimation of the large inverse covariance matrix $T_\lambda \left(\hat{\boldsymbol{\Sigma}}_x \right)^{-1}$. When $p > T$ the $T_\lambda \left(\hat{\boldsymbol{\Sigma}}_x \right)$ does not necessitate positive definiteness (PD) and thus existence of $T_\lambda \left(\hat{\boldsymbol{\Sigma}}_x \right)^{-1}$. When the tuning parameter κ of the threshold λ_{ij} in (4) or (5) is unknown, and needs to be calibrated via a cross validation method, then conditions on the minimum eigenvalues of $T_\lambda \left(\hat{\boldsymbol{\Sigma}}_x \right)$ can be adopted ad hoc. For instance, for a given grid of points for $\kappa \in \{\kappa_1, \kappa_2, .., \kappa_N\}$, one can consider only the points which result to PD matrices $T_\lambda \left(\hat{\boldsymbol{\Sigma}}_x \right)$. This will ensure existence of $T_\lambda \left(\hat{\boldsymbol{\Sigma}}_x \right)^{-1}$. In the case that κ is known or when there is no a priori reason to discard specific values for κ, an alternative that results to PD $T_{\lambda}\left(\hat{\mathbf{\Sigma}}_{x}\right)$ is to consider convex combination of $T_{\lambda}\left(\hat{\mathbf{\Sigma}}_{x}\right)$ and a well-defined target matrix which is known to result to PD matrix. In the literature there have been proposed simple target matrices like the identity matrix I_p, or the diagonal matrix $diag\left(\hat{\mathbf{\Sigma}}_{x}\right)$. The shrinking to a target matrix approach is formally given as

$$T_{\lambda,s}\left(\hat{\mathbf{\Sigma}}_{x}\right) = \rho_s \mathbf{\Sigma}_{tar} + (1-\rho_s)\, T_{\lambda}\left(\hat{\mathbf{\Sigma}}_{x}\right) \tag{19}$$

with $\mathbf{\Sigma}_{tar}$ being the target matrix, and $\rho_s \in (\rho_{inv}, 1)$, $0 < \rho_{inv} < 1$, with ρ_{inv} being the minimum shrinkage parameter that ensures the PD of $T_{\lambda,shr}\left(\hat{\mathbf{\Sigma}}_{x}\right)$.

Moreover, the problem of inverting a large covariance matrix, which is essential for the proposed $\hat{\boldsymbol{\beta}}_{\lambda}$ can be tackled through another interesting direction. Cai and Liu (2011) suggest a constrained l_1 minimization method for estimating a sparse inverse covariance matrix. The authors provide interesting theoretical and simulation results for data which are iid random variables, while they show that their method is fast and easily implemented by linear programming.

Preliminary simulations for the (19) and Cai and Liu (2011) extensions do not suggest significant improvements over the basic estimator and testing as it is defined in sections 2 and 3.

8 A Monte Carlo study

To examine the small sample performance of our methodological contributions we carry out an extensive Monte Carlo study. To do so, we allow for a wide set of covariate designs, which are considered as plausible for macroeconomic and/or financial time series. Then, we examine the ability of each method to determine the true regressors, and the true parameter vector.

The literature on large dimensional regression has been dominated by penalized regression methods. Top among them are the Lasso and Adaptive Lasso methods, whose theoretical properties have been extensively analysed in e.g. Zou (2006), Zhao and Yu (2006), and Meinshausen and Bühlmann (2006). These are considered as refinements of multiple regression, where the vector of regression coefficients is obtained as a solution to the following optimization problem

$$\widehat{\beta} = \arg\min_{\beta} \sum_{t=1}^{T} \left(y_t - \sum_{i=1}^{p} \beta_i x_{it} \right)^2 + P(\beta, \lambda) \tag{20}$$

where $P(\boldsymbol{\beta}, \lambda)$ is a function of the tuning parameter λ that penalizes the complexity of coefficient $\boldsymbol{\beta}$. In empirical applications, the tuning parameter λ is calibrated by cross validation. For Lasso regression, $P(\boldsymbol{\beta}, \lambda)$ is the L_1 norm of the vector $\boldsymbol{\beta}$. The L_1 geometry of the solution implies that Lasso is performing variable selection in the sense that an estimated component can be exactly zero. For adaptive lasso (see e.g. Zhou and Hastie (2005)) the L_1 norm is replaced by a re-weighted version which diminishes the estimation bias implied by Lasso.

In our simulation experiment we generate the data from model (1) assuming that y_t is generated by the first 4 columns of $\boldsymbol{x}_t$ (i.e. $k = 4$) and $\beta_i = 1$, for $i = 1, .., 4$. Since most economic data support the inclusion of a constant, in model (1), we also set $\beta_0 = 1$. In all methods the inclusion of the constant is assumed to be known, that is, for all methods the TPR and FPR do not account for the constant but only for the rest of the p possible regressors.

The tuning parameters of the methods are calibrated using the CV methods presented in the previous section. The Lasso and Adaptive Lasso parameters are calibrated using 10-fold cross validation, as this is considered as a benchmark, in the literature, while for the rest of the proposed methods, $T-1$-fold cross validation is used, which is also known as leave one out cross validation. The Adaptive Lasso method is applied, as described in section 2.8.4 of Buhlmann and van de Geer (2011).

We then use 500 simulation experiments to measure average values of TPR and FPR defined in (13), average RMSE of the parameter vector defined as

$$RMSE = \sqrt{\frac{1}{p}\sum_{i=1}^{p}\left(\widehat{\beta}_i - \beta_i\right)^2}, \tag{21}$$

average probability of observing exactly the true model, defined as

$$TM = I\left(\left\{\sum_{i=1}^{k} I\left(\widehat{\beta}_i \neq 0\right) = k\right\} \cap \left\{\sum_{i=k+1}^{p} I\left(\widehat{\beta}_i \neq 0\right) = 0\right\}\right) \tag{22}$$

with $I\left(\widehat{\beta}_i \neq 0\right) = 1$ when $\widehat{\beta}_i \neq 0$ and zero otherwise, and average out of sample mean squared forecast error, computed at the $T+1$-*th* simulated observation and defined as

$$RMSFE = \sqrt{\left(y_{T+1} - \widehat{y}_{T+1}\right)^2} \tag{23}$$

For all these measures of performance the large set of covariates x_t are generated according to a wide range of distributional designs, presented bellow, and a rich set of T, p, R^2 parameters.

8.1 Designs for regressors

8.1.1 IID covariates

In this naive case we assume that covariates are generated from $x_t \sim N\left(0, I_p\right)$, for $t = 1, .., T$. This is the simplest case that we examine in which the regressors are uncorrelated.

8.1.2 Temporally uncorrelated, weakly collinear covariates

For the temporally uncorrelated, weakly collinear covariates we assume the true signals are generated by

$$x_{it} = \left(\varepsilon_{it} + g_t\right) / \sqrt{2}, \text{ for } i = 1, .., 4, \tag{24}$$

and the noise variables by

$$\begin{aligned} x_{5t} &= \varepsilon_{5t} \\ x_{it} &= \left(\varepsilon_{it} + \varepsilon_{i-1t}\right) / \sqrt{2}, \text{ for } i > 5, g_t \sim NIID\left(0,1\right), \varepsilon_{it} \sim NIID\left(0,1\right) \end{aligned}$$

In this case, there is correlation among the true signal variables, and among the noisy variables, but there is no correlation between the true signal and the noisy regressros. This implies a 50% correlation among the true signal variables.

8.1.3 Pseudo signal variables

For the pseudo signal variable case we assume that the true signals are generated by

$$x_{it} = \left(\varepsilon_{it} + g_t\right) / \sqrt{2}, \text{ for } i = 1, .., 4 \tag{25}$$

and the noise variables are generated by

$$\begin{aligned} x_{5t} &= \varepsilon_{5t} + \kappa x_{1t},\ x_{6t} = \varepsilon_{6t} + \kappa x_{2t} \text{ for } \kappa = 1.33 \\ x_{it} &= \left(\varepsilon_{it} + \varepsilon_{i-1t}\right) / \sqrt{2}, \text{ for } i > 6, g_t \sim N\left(0,1\right), \varepsilon_{it} \sim N\left(0,1\right) \end{aligned}$$

Now, we allow for all types of correlations between the variables. Correlations among the true signals, correlation among the noisy variables and correlation between the true signals and the noisy variables. This process implies that there are two pseudosignal variables, namelly, the x_{5t} and x_{6t}, with a correlation among signals and pseudosignals of 80%.

8.1.4 Strongly collinear noise variables

For the strongly collinear noise variables, using a persistent and unobserved common factor, we assume that the true signals are generated as

$$x_{it} = (\varepsilon_{it} + g_t) / \sqrt{2}, \text{ for } i = 1, .., 4, \tag{26}$$

while the noise variables are generated as

$$\begin{aligned} x_{5t} &= (\varepsilon_{5t} + b_i f_t) / \sqrt{3} \\ x_{it} &= \left((\varepsilon_{it} + \varepsilon_{i-1t}) / \sqrt{2} + b_i f_t \right) / \sqrt{3}, \text{ for } i > 5, b_i \sim N(1,1) \\ f_t &= 0.95 f_{t-1} + \sqrt{1 - 0.95^2} v_t, v_t \sim N(0,1) \end{aligned}$$

8.1.5 Temporally correlated and weakly collinear covariates

For the temporally correlated and weakly collinear covariate case, we assume that the true signals are generated as

$$x_{it} = (\varepsilon_{it} + g_t) / \sqrt{2}, \text{ for } i = 1, .., 4 \tag{27}$$

while the noise variables are generated as

$$\begin{aligned} x_{5t} &= \varepsilon_{5t} \\ x_{it} &= (\varepsilon_{it} + \varepsilon_{i-1t}) / \sqrt{2}, \text{ for } i > 5 \\ \varepsilon_{it} &= \rho \varepsilon_{it-1} + \sqrt{1 - \rho^2} v_{it}, v_t \sim N(0,1), \end{aligned}$$

where we set $\rho = 0.5$ for all i.

8.1.6 All covariates are collinear with the true signals

Finally, we examine the case in which all variables are collinear with the true signals

$$\boldsymbol{x}_t \sim N\left(0, \boldsymbol{\Sigma}_p\right)$$

with the i, j-th element of Σ_p defined as $\sigma_{ij} = 0.5^{|i-j|}, 1 \leq i, j \leq p$.

8.2 Summary of simulation results

In tables 1 to 6 we present simulation results for the DGPs of the covariate vector, $\boldsymbol{x}_t$, which are presented in the previous section. We report results for the $RMSE$ of the parameter vector β, the $RMSFE$ of the one step ahead forecast, the TPR, FPR as these are defined in equations (13), as well as the TM, which measures the number of times that a method exactly detects the correct model, that is TPR=1 and FPR=0, at the same time. In our comparisons, we focus on Lasso and adaptive Lasso since these are the main penalized regression methods used in the literature and also because they tend to perform better than other important methods like Boosting or other penalized regression methods, as it is highlighted in the literature.

Comparing Lasso with adaptive Lasso, it is evident that the former performs overwhelmingly better than the latter in the majority of the experiments, in terms of $RMSE$ of the parameter vector $\boldsymbol{\beta}$. The very slight decreases in the FPR which are delivered by the adaptive Lasso method, come at the significant cost of expanding $RMSE$. This implies that if our main aim is model selection, adaptive Lasso might be more suitable than Lasso. On the other hand, the method that minimizes the small sample bias, in estimation vector $\boldsymbol{\beta}$, is the Lasso. For this reason we choose Lasso as our benchmark model for comparison. In all tables the $RMSE$ of the parameter vector $\boldsymbol{\beta}$, and the $RMSFE$ of the one step ahead forecast, are reported as ratios from the Lasso performance. This means that values higher 1 imply that the Lasso performs better in these metrics than the method which is compared with.

Focusing, first, on the uncorrelated covariates case, we see that the probability that the Lasso and/or adaptive Lasso methods identify the truth is very close to zero. This result is not surprising as it is known that these two methods are choosing larger than the truth models. Contrarily, we see that our basic regularized estimator (R-BL and R-CL in the tables) performs satisfactory compared to Lasso and adaptive Lasso. For large and medium R^2 $(= 0.9, 0.6)$ as well as large and medium T $(T = 200, 300)$ the improvements in $RMSE$ over the benchmark are between 15% and 40%. The stronger the signal (higher R^2) and larger the T, the more significant is the improvement. It is reasonable to believe that the power of our set of proposals, increases with the T and the R^2. This can be also seen by the better TPR, FPR, TM measures, which are presented in the tables and support our theoretical findings of Theorem 1. When we enhance this estimator by the sure screening approach, the observed improvements are slight and non significant, indicating that in the uncorrelated covariates case, the basic estimator does not need the support of the sure screening enhancement to work satisfactory. Adaptive thresholding seems to be important, in this case, even though all $\sigma_{ii} = 1$, suggesting that this type of thresholding should not be necessary. The R-CL delivers larger improvements over the Lasso, than the one that is provided by R-BL, in terms of $RMSE$, although when the signal deteriorates (the R^2 diminishes), this method (R-CL) fails to result to FPR which is close to zero. The multiple testing enhancements that we consider, also provide large gains. The three step procedure (R-CL-t-r or R-BL-t-r) are the best performing methods with gains in $RMSE$ of order 40-50% over the Lasso method and this is partially affected by the choice of R^2 and T. This is also visible in the TPR, FPR metrics, which indicate that the model selection is also improved considerably, compared to the basic estimator. The adaptive thresholding (CL) now does not provide significant gains over the global thresholding (BL). This supports the view that testing, improves considerably the screening mechanism, making the specific thresholding approach not so important. The performance of the two step approach (R-CL-t or R-BL-t) is also significantly better than the Lasso method and in between the other two proposals considered. Again, entry specific thresholding is not so important while the sure screening extension is not important, also here. Overall, in this experiment there is a plethora of proposed methods that perform better than the Lasso and adaptive Lasso, with the R-CL-t-r or R-BL-t-r being the most successful.

Moving now to less sparse covariance matrices for the vector $\boldsymbol{x}_t$, as the one that is considered in the case of temporally uncorrelated and weakly collinear covariates, it is visible that the basic estimator (R-BL and R-CL) provides improvements over the benchmark only when the adaptive threshold is considered, while in both cases these deteriorate, compared to the uncorrelated covariates case, that we considered before. Now, the R-CL provides advancements in $RMSE$ of order 15-20% over the Lasso, while the sure screening does not affect significantly the performance in both R-CL and R-BL. Moving to the multiple testing enhancements, it is visible that the three step approach (R-CL-t-r or R-BL-t-r) is the best performing method again. Now the R-BL-t-r seems to provide some gains over the R-CL-t-r, in the case of medium and low $R^2 (= 0.3, 0.6)$, while comparing R-CL-t-r with R-CL we see that the source of this deterioration is the testing and not the estimation of $\boldsymbol{\beta}$, as the $RMSE$ of R-CL is lower than that of R-CL-t-r. Remarkably, this is not the case for R-BL-t-r and R-BL, where testing improves the performance of the basic estimator, as expected, in both model selection (TPR-FPR) and estimation bias ($RMSE$). This finding suggests that even when σ_{ii}'s are not constant across i, the estimation error that is induced because of the adaptive thresholding, may cancel any gains from it. This can result to poor estimates of $\boldsymbol{\beta}$, or poor estimates of its variance, as it is the case here, deteriorating the testing results.

In the case that the $\boldsymbol{x}_t$ includes pseudo signal variables, as in the case of pseudo signal variables, the proposed methods that outperform the Lasso methods, shrink considerably. Now, the that performs better than the Lasso, in terms of $RMSE$, is the R-BL-t-r. The highest the T and R^2 the better the performance. The maximum improvement is of order 10%, which is considerably lower than the improvements that we considered in the previous cases. Remarkably, while Lasso delivers TPR equal to 1 and FPR above zero, our proposal provides its gains through a reverse behavior: TPR below 1 and FPR close to zero, meaning that with R-BL-t-r it is more likely that you miss some true covariates, but you will discard all the false covariates of the model, in contrast to the Lasso's behavior. Here, it is also evident, that the small T $(= 150)$ needs the enhancement of the sure screening in order to beat the Lasso, in terms of $RMSE$. This is one of the few cases, in which the sure screening is necessary in order to enhance the performance of our proposals.

Focusing in the case in which all variables are collinear with the true signals, we see that most of the results discussed in the previous case, hold also here. The R-BL-t-r is the most credible method for parameter estimation and variable selection. Again, it might miss some of the true regressors, but it is more unlikely than it is with Lasso, to choose false covariates, as true regressors. The sure screening enhancement, in this case does not provide significant gains over the basic R-BL-t-r.

Finally, the good results of R-BL-t-r or S-R-BL-t-r in terms of $RMSE$, remain in strongly collinear noise variables due to a persistent unobserved common factor case, and the temporally correlated and weakly collinear covariates case. In the latter, the adaptive regularization, version of the regularized estimator seems to work well, as it has also be seen in the case of temporally uncorrelated and weakly collinear covariates.

Overall the basic estimator R-CL or R-BL provide gains over the benchmark in our simplest cases, while the multiple testing enhancements, results to significant improvements in most of the cases considered. When the covariance of $\boldsymbol{x}_t$ becomes less sparse, the adaptive thresholding gains are cancelled, probably due to the increased estimation error that it induces. The sure screening approach does not provide significant gains here. While, this approach has been considered as a good enhancement for the penalized regression methods, this is not true for our regularization and testing proposals. This might attributed to the fact that in our proposals, we consider explicitly the structure of the covariance matrix, accounting for zero elements in the estimation and testing procedure, making this type of screening unnecessary.

9 Empirical illustration

As an empirical illustration, we perform an out of sample forecasting evaluation of our models, for five key macroeconomic series using a large set of available covariates. We consider the monthly data from the FRED-MD database which consists of a dataset of 124 macroeconomic and financial series of the US economy. With this set of possible regressors augmented with 4 lags of the dependent variable, we forecast 5 key macroeconomic series, namely, industrial production index, housing price index, inflation, unemployment and the unfilled orders for durable goods (see table 7 for the list of forecasted variables). Our dataset starts at 01-Jan-1993 and ends at 01-Dec-2015. For each dependent variable we use a rolling estimation window of 140 monthly observations, to forecast the last 136 monthly observations of our sample. That is, the forecasting evaluation starts at 01-Sep-2004. To assess whether the relative performance of a model is stable over time, we also evaluate models over the recent recession event period. This will indicate if some models can perform better during crisis periods.

As in the Monte Carlo section, the tuning parameter for our methods, is also calibrated by the leave one out cross validation, while now since our main aim is forecasting we do this over the last 12 observations, of each estimation window considered. For Lasso and Adaptive Lasso we consider again the 10-fold cross validation. We compare our results with the autoregressive model of order d, (AR(d)), where d is selected by the Bayesian information criterion (BIC), and the factor augmented AR(d), (FAAR(d)), where the number of factors is chosen by the Bai and Ng (2002) information criterion. In the latter, the factors are extracted through principal component analysis (PCA) on the large set of available regressors.

The forecasting performance of the alternative models is evaluated relative to that of the benchmark (the AR(d) model) using the relative root mean squared forecast error (r-RMSFE). For each model m and target series s, it is:

$$r\text{-}RMSFE_{(m,s)} = \frac{\sqrt{\sum_{t=t_0}^{T}\left(e_t^{(m,s)}\right)^2}}{\sqrt{\sum_{t=t_0}^{T}\left(e_t^{(AR(d),s)}\right)^2}}, \tag{28}$$

where $e_t^{(m,s)} = y_t^{(s)} - \widehat{y}_t^{(m,s)}$ is the 1-step ahead forecast error of model m for series s, and $e_t^{(AR(d),s)} = y_t^{(s)} - \widehat{y}_t^{(AR(d),s)}$ is the counterpart for the benchmark $AR\,(d)$ model. When the $r\text{-}RMSFE_{(m,s)}$ is less than one, model m out performs the benchmark $AR\,(d)$ for macroeconomic variable s. To assess the significance of the forecast evaluation, we use the common

Diebold and Mariano (1995) test [1] as well as the forecast fluctuation test developed by Giacomini and Rossi (2010).[2]

A number of interesting conclusions could be drawn from table 8 and figures 1 to 5. In all series examined our proposals are either the best performing methods or among the 3 best performing ones. Focusing one the full sample evaluation, we see that in HOUSTNE series the R-BL-t-r is the best performing one, while in all other cases the evaluation of our methods differs slightly from the best performing one. In most of the cases, this is the Lasso method, supporting the view that this is the leading benchmark in the large dimensional regression literature. Unsurprisingly, the proposed three step procedure (R-BL-t-r or R-CL-t-r) seems to be the most reliable and steady performing method, from our proposals. They perform remarkably in UNRATE, AMDMNOx, and of course, HOUSTNE, with or without the sure screening extension. This is in accordance with our monte Carlo findings, were we have seen that when the Σ_z becomes less sparse, this method performs remarkably better than our other proposals. The only exemption is the INDPRO in which the sure screening extension of the R-BL-t enables it to be in the top 3 performing methods. Interestingly, in this full sample evaluation, the AR(d) and the Factor Augmented Ar(d) do not qualify in the best three methods, in any of the five series examined. Concluding, in this full sample evaluation the penalized regression are the most successful methods, while their performance is comparable to the R-BL-t-r or R-CL-t-r, which perform steadily well. The sure screening affects significantly the results only in the INDPRO case.

Turning into the crisis period evaluation interval we see that our proposals qualify as the best performing in more cases. Now the Lasso and Adaptive Lasso fails to enter in the top three performing methods in 4 out of 5 series examined, while our proposals with or without the sure screening extension qualify first in the 4 out of 5 cases. Again, the R-BL-t-r or R-CL-t-r are the most reliable and steady performing methods.

10 Conclusion

In this paper, we have considered the problem of model specification and estimation where there is a large set of potential predictors, for a dependent variable y_t. We provide alteranatives to and variations of penalized regression and greedy methods, which are related to

[1] Since our out of sample period contains a large number of observations, the small sample correction of the DM test proposed by Harvey et al. (1997) is not necessary, as it would lead to very minor modifications of the original DM statistic.

[2] In the empirical exercise we set the rolling windows size at $\nu = 50$, as it is recommended by Giacomini and Rossi (2010).

statistical inference. A fundamental component of our theoretical contribution is sparsity of the covariance matrix of the regressor matrix. Our theoretical, simulation and empirical results provide evidence in favour of the methods we propose, compared to existing methods.

Letter	Explanation
S	SIS
R-BL	Regularized estimator with Bickel and Levina
R-CL	Regularized estimator with Cai and Liu
t	testing procedure
r	final regression on important covariates

Table A: Method names in the tables. Each row stands for a method. Two or more rows result to a combined method. E.g. S-R-BL-t-r is a sure screening method (S), combined the basic regularized estimator (R-BL), with testing (t) and a final regression (r). R-BL-t-r is the basic regularized estimator (R-BL), with testing (t) and a final regression (r). R-BL-t is the basic regularized estimator. (R-BL), with testing (t). R-BL is the basic regularized estimator (R-BL).

T=200, p=200	Lasso	A-Lasso	S-R-BL-t	S-R-BL-t-r	S-R-BL	S-R-CL-t	S-R-CL-t-r	S-R-CL	R-BL-t	R-BL-t-r	R-BL	R-CL-t	R-CL-t-r	R-CL
$R^2 = 0.7$														
TPR	1	1	1	1	1	1	1	1	1	1	1	1	1	1
FPR	0.08	0.08	0	0	0	0	0.01	0	0	0	0	0	0	0.08
RMSE(β)	1	1.32	0.65	0.55	0.82	0.7	0.68	0.68	0.67	0.53	0.82	0.67	0.55	0.7
TrueModel	0	0	0.98	0.67	0.72	0.88	0.35	0.94	0.98	0.75	0.72	0.95	0.56	0.84
RMFSE	1	1.05	0.97	0.98	0.99	0.98	1	0.97	0.97	0.98	0.99	0.98	0.98	0.98
$R^2 = 0.5$														
TPR	1	1	0.99	1	0.97	1	1	1	0.99	0.99	0.97	0.99	0.99	0.99
FPR	0.08	0.08	0	0	0	0	0.01	0	0	0	0	0	0	0.27
RMSE(β)	1	1.34	0.67	0.6	0.87	0.76	0.73	0.73	0.67	0.57	0.87	0.69	0.59	0.75
TrueModel	0.01	0.01	0.82	0.6	0.52	0.63	0.29	0.7	0.82	0.66	0.52	0.79	0.57	0.55
RMFSE	1	1.06	0.98	0.97	0.99	0.97	0.97	0.98	0.98	0.97	0.99	0.99	0.97	1
$R^2 = 0.3$														
TPR	0.99	0.99	0.86	0.89	0.87	0.89	0.9	0.89	0.77	0.78	0.87	0.78	0.8	0.88
FPR	0.07	0.07	0	0.01	0.01	0	0.01	0.01	0	0	0.01	0	0	0.59
RMSE(β)	1	1.26	0.98	0.88	1.1	0.96	0.87	1.04	1.04	0.98	1.1	1.03	0.94	1.06
TrueModel	0	0	0.24	0.21	0.17	0.25	0.2	0.21	0.21	0.21	0.17	0.24	0.24	0.07
RMFSE	1	1.04	1	0.99	1.01	1.01	1	1	1	1	1.01	1	0.99	1.01
T=300, p=200														
$R^2 = 0.7$														
TPR	1	1	1	1	1	1	1	1	1	1	1	1	1	1
FPR	0.08	0.08	0	0.01	0	0	0.01	0	0	0.01	0	0	0.01	0.05
RMSE(β)	1	1.34	0.68	0.69	0.71	0.73	0.74	0.7	0.66	0.61	0.71	0.7	0.65	0.7
TrueModel	0	0	0.94	0.35	0.91	0.82	0.27	0.93	0.98	0.53	0.91	0.92	0.39	0.89
RMFSE	1	1.02	0.98	0.97	0.97	0.98	0.98	0.98	0.98	0.98	0.97	0.97	0.97	0.97
$R^2 = 0.5$														
TPR	1	1	1	1	0.99	1	1	1	1	1	0.99	1	1	1
FPR	0.08	0.08	0	0.01	0	0	0.01	0	0	0	0	0	0	0.15
RMSE(β)	1	1.38	0.65	0.67	0.69	0.76	0.76	0.68	0.62	0.59	0.69	0.64	0.62	0.65
TrueModel	0.01	0.01	0.85	0.44	0.82	0.66	0.27	0.8	0.91	0.6	0.82	0.87	0.51	0.74
RMFSE	1	1.05	0.98	0.98	0.98	0.97	0.98	0.98	0.98	0.98	0.98	0.98	0.98	0.97
$R^2 = 0.3$														
TPR	1	1	0.98	0.99	0.95	0.99	0.99	0.99	0.95	0.96	0.95	0.96	0.97	0.98
FPR	0.07	0.07	0	0.01	0	0	0.01	0.01	0	0	0	0	0	0.48
RMSE(β)	1	1.34	0.76	0.71	0.81	0.8	0.76	0.81	0.71	0.66	0.81	0.7	0.66	0.77
TrueModel	0	0	0.54	0.4	0.52	0.42	0.27	0.49	0.66	0.59	0.52	0.68	0.57	0.34
RMFSE	1	1.01	0.99	0.98	1	0.98	0.99	1	0.99	0.98	1	0.98	0.98	1
T=150, p=200														
$R^2 = 0.7$														
TPR	1	1	0.99	1	0.99	1	1	1	0.99	1	0.99	0.99	1	0.99
FPR	0.08	0.08	0	0	0	0	0.01	0	0	0	0	0	0	0.17
RMSE(β)	1	1.28	0.74	0.54	1	0.72	0.63	0.76	0.76	0.53	1	0.76	0.52	0.83
TrueModel	0.01	0.01	0.91	0.66	0.49	0.83	0.4	0.85	0.88	0.64	0.49	0.85	0.61	0.66
RMFSE	1	1.04	0.95	0.93	0.98	0.96	0.94	0.95	0.94	0.93	0.98	0.95	0.94	0.96
$R^2 = 0.5$														
TPR	1	1	0.94	0.97	0.93	0.97	0.98	0.95	0.93	0.94	0.93	0.93	0.94	0.93
FPR	0.08	0.08	0	0.01	0	0	0.01	0	0	0	0	0	0	0.41
RMSE(β)	1	1.29	0.88	0.72	1.01	0.84	0.72	0.9	0.87	0.72	1.01	0.88	0.73	0.99
TrueModel	0	0	0.54	0.4	0.33	0.51	0.33	0.47	0.53	0.45	0.33	0.54	0.47	0.24
RMFSE	1	1.03	1.02	0.99	1.03	0.99	0.98	1.02	1.03	0.99	1.03	1.03	0.99	1.03
$R^2 = 0.3$														
TPR	0.97	0.97	0.73	0.77	0.79	0.77	0.78	0.77	0.61	0.62	0.79	0.61	0.63	0.77
FPR	0.07	0.07	0	0.01	0.01	0	0.01	0.01	0	0	0.01	0	0	0.55
RMSE(β)	1	1.24	1.13	1.02	1.19	1.09	1.02	1.17	1.16	1.12	1.19	1.17	1.1	1.21
TrueModel	0	0	0.09	0.08	0.06	0.09	0.07	0.06	0.08	0.08	0.06	0.08	0.08	0.02
RMFSE	1	1.05	1.03	1.01	1.04	1.01	1	1.05	1.03	1.03	1.04	1.03	1.02	1.05

Table 1: Simulation results for iid covariates. Method names are given in Table A. The number of simulations is 500. The MSE and the RMSFE are given as ratios from the Lasso method.

T=200, p=200	Lasso	A-Lasso	S-R-BL-t	S-R-BL-t-r	S-R-BL	S-R-CL-t	S-R-CL-t-r	S-R-CL	R-BL-t	R-BL-t-r	R-BL	R-CL-t	R-CL-t-r	R-CL
R^2 = 0.7														
TPR	1	1	0.74	0.99	0.86	0.99	0.99	1	0.65	1	0.78	0.99	0.99	1
FPR	0.04	0.04	0	0	0.06	0	0	0	0	0	0.01	0	0	0
RMSE(β)	1	1.36	2.73	0.82	4.83	0.79	0.91	0.75	5.09	0.74	5.65	0.75	0.79	0.74
TrueModel	0.1	0.1	0.09	0.91	0.01	0.94	0.7	0.98	0.02	0.98	0.14	0.97	0.95	1
RMFSE	1	1.01	1.14	0.97	1.46	0.98	0.97	0.97	1.46	0.97	1.6	0.97	0.97	0.97
R^2 = 0.5														
TPR	1	1	0.55	0.97	0.78	0.79	0.85	0.99	0.6	0.99	0.75	0.8	0.9	1
FPR	0.04	0.04	0	0	0.05	0	0	0	0	0	0	0	0	0.01
RMSE(β)	1	1.35	2.57	0.83	3.5	1.48	1.32	0.89	3.26	0.77	3.5	1.32	1.08	0.76
TrueModel	0.06	0.06	0	0.85	0	0.19	0.35	0.82	0	0.94	0.11	0.24	0.55	0.99
RMFSE	1	1.01	1.14	0.98	1.32	1.02	1	0.98	1.2	0.97	1.33	1.01	0.99	0.97
R^2 = 0.3														
TPR	0.95	0.95	0.49	0.9	0.74	0.52	0.55	0.94	0.5	0.89	0.73	0.49	0.76	1
FPR	0.03	0.03	0	0	0.02	0	0	0	0	0	0	0	0	0.02
RMSE(β)	1	1.34	2.07	0.93	2.48	1.56	1.7	1.09	2.2	0.91	2.4	1.64	1.2	0.81
TrueModel	0.07	0.07	0	0.57	0.09	0	0	0.41	0	0.59	0.13	0.02	0.43	0.98
RMFSE	1	1.03	1.06	0.98	1.17	1.04	1.02	0.99	1.05	0.99	1.15	1.05	1.01	0.99
T=300, p=200														
R^2 = 0.7														
TPR	1	1	0.98	1	1	1	1	1	0.89	1	0.82	1	1	1
FPR	0.03	0.03	0	0.01	0.06	0	0.01	0	0	0	0.11	0	0	0
RMSE(β)	1	1.38	1.09	0.97	3.63	0.81	1.06	0.75	2.77	0.87	14.43	0.74	0.78	0.74
TrueModel	0.08	0.08	0.77	0.56	0	0.93	0.43	0.99	0.52	0.76	0.03	1	0.97	1
RMFSE	1	1	0.99	0.97	1.22	0.98	0.98	0.98	1.12	0.98	1.55	0.98	0.98	0.98
R^2 = 0.5														
TPR	1	1	0.81	0.97	0.98	0.93	0.94	1	0.67	0.99	0.77	0.94	0.96	1
FPR	0.03	0.03	0	0	0.1	0	0	0	0	0	0.02	0	0	0
RMSE(β)	1	1.38	1.82	0.96	3.49	1.24	1.21	0.91	2.82	0.83	5	0.97	0.95	0.75
TrueModel	0.09	0.09	0.1	0.74	0	0.5	0.42	0.82	0.04	0.88	0.07	0.71	0.8	1
RMFSE	1	1.01	1.02	0.99	1.26	1	1.01	0.99	1.14	0.98	1.3	0.98	0.99	0.98
R^2 = 0.3														
TPR	0.99	0.99	0.53	0.94	0.85	0.64	0.67	0.97	0.54	0.95	0.75	0.62	0.88	1
FPR	0.03	0.03	0	0	0.12	0	0	0	0	0	0.01	0	0	0.02
RMSE(β)	1	1.39	2.1	0.94	3.33	1.62	1.74	1.13	2.44	0.9	3.07	1.62	1.09	0.79
TrueModel	0.08	0.09	0	0.67	0	0.01	0.01	0.48	0	0.77	0.14	0.01	0.67	0.98
RMFSE	1	1.01	1.04	0.98	1.17	1.02	1.01	1.01	1.08	0.98	1.19	1.03	1	0.98
T=150, p=200														
R^2 = 0.7														
TPR	1	1	0.56	0.99	0.79	0.96	0.97	1	0.6	1	0.78	0.96	0.98	0.99
FPR	0.04	0.04	0	0	0.04	0	0	0	0	0	0	0	0	0
RMSE(β)	1	1.34	3.73	0.76	4.48	0.91	0.91	0.75	4.22	0.74	4.59	0.91	0.84	0.8
TrueModel	0.08	0.08	0	0.97	0.05	0.79	0.72	0.98	0.01	0.99	0.17	0.81	0.88	0.98
RMFSE	1	1.02	1.37	0.97	1.73	0.99	0.99	0.98	1.45	0.98	1.92	0.99	0.98	0.98
R^2 = 0.5														
TPR	0.99	0.99	0.51	0.97	0.77	0.69	0.78	0.99	0.48	0.97	0.77	0.68	0.84	1
FPR	0.03	0.03	0	0	0.02	0	0	0	0	0	0	0	0	0.01
RMSE(β)	1	1.35	2.76	0.82	3.21	1.65	1.44	0.88	2.81	0.8	3.15	1.59	1.23	0.83
TrueModel	0.09	0.09	0	0.83	0.13	0.05	0.19	0.83	0	0.85	0.19	0.06	0.35	0.98
RMFSE	1	1.03	1.19	0.99	1.45	1.03	1.01	1	1.23	0.99	1.45	1.02	1	0.99
R^2 = 0.3														
TPR	0.91	0.91	0.43	0.87	0.75	0.48	0.51	0.91	0.43	0.83	0.75	0.46	0.7	0.99
FPR	0.03	0.03	0	0	0	0	0	0	0	0	0	0	0	0.04
RMSE(β)	1	1.34	1.98	0.94	2.21	1.55	1.63	1.05	1.97	0.95	2.21	1.79	1.23	1.17
TrueModel	0.06	0.06	0	0.56	0.16	0	0	0.42	0	0.45	0.17	0.01	0.29	0.93
RMFSE	1	1.02	1.03	0.97	1.14	1.03	1.02	0.99	1.05	0.97	1.15	1.08	1	0.99

Table 2: Simulation results for temporally uncorrelated and weakly collinear covariates.

T=200, p=200	Lasso	A-Lasso	S-R-BL-t	S-R-BL-t-r	S-R-BL	S-R-CL-t	S-R-CL-t-r	S-R-CL	R-BL-t	R-BL-t-r	R-BL	R-CL-t	R-CL-t-r	R-CL
$R^2 = 0.7$														
TPR	1	1	0.56	1	0.77	0.9	0.97	1	0.51	0.99	0.45	0.65	0.94	0.98
FPR	0.03	0.03	0	0	0.08	0	0	0.01	0	0	0.02	0	0.01	0.68
RMSE(β)	1	1.31	3.57	0.86	4.57	1.49	0.96	1.04	4.92	0.86	5.05	3.16	1.26	4.21
TrueModel	0.06	0.06	0.01	0.51	0	0.49	0.57	0	0.04	0.6	0	0.07	0.07	0
RMFSE	1	1.01	1.37	0.97	1.45	1.04	0.98	0.99	1.53	0.98	1.53	1.2	0.98	1.07
$R^2 = 0.5$														
TPR	1	0.99	0.48	0.97	0.5	0.66	0.8	1	0.56	0.92	0.38	0.5	0.83	0.99
FPR	0.04	0.04	0	0	0.05	0	0	0.01	0	0	0.01	0	0.01	0.65
RMSE(β)	1	1.37	2.8	0.92	3.42	1.99	1.43	1.06	3.07	1.04	3.36	2.78	1.46	2.87
TrueModel	0.05	0.05	0.03	0.44	0	0.03	0.2	0.06	0.03	0.48	0	0	0.01	0
RMFSE	1	1.02	1.17	0.98	1.31	1.06	0.98	0.98	1.12	0.96	1.25	1.14	0.99	1.02
$R^2 = 0.3$														
TPR	0.92	0.92	0.49	0.85	0.35	0.5	0.52	0.94	0.52	0.72	0.34	0.35	0.65	0.99
FPR	0.03	0.03	0	0	0.02	0	0	0.01	0	0	0.01	0	0.01	0.69
RMSE(β)	1	1.35	2.04	1.07	2.34	1.59	1.62	1.13	2.1	1.26	2.32	2.35	1.49	6.23
TrueModel	0.04	0.04	0	0.25	0	0	0	0.13	0	0.08	0	0	0	0
RMFSE	1	1.02	1.09	1	1.14	1.04	1.05	1.01	1.06	1.01	1.12	1.15	1.02	1.29
T=300, p=200														
$R^2 = 0.7$														
TPR	1	1	0.87	1	1	0.98	0.98	1	0.66	1	0.69	0.77	0.98	1
FPR	0.04	0.04	0.01	0.01	0.07	0	0.01	0.01	0	0	0.15	0	0.01	0.86
RMSE(β)	1	1.33	2.41	0.92	3.51	1.08	1.02	0.97	3.76	0.86	15.36	2.83	1.11	4.65
TrueModel	0.06	0.06	0.11	0.4	0	0.8	0.42	0	0	0.48	0	0.15	0.08	0
RMFSE	1	1.01	1.08	0.98	1.24	0.99	0.98	0.98	1.27	0.99	1.52	1.11	0.99	1.03
$R^2 = 0.5$														
TPR	1	1	0.63	0.99	0.94	0.81	0.89	1	0.55	0.99	0.54	0.58	0.92	0.99
FPR	0.04	0.04	0	0.01	0.11	0	0	0.01	0	0	0.02	0	0.01	0.83
RMSE(β)	1	1.37	2.48	0.96	3.44	1.71	1.28	1.08	3.17	0.92	4.12	2.87	1.34	3.29
TrueModel	0.05	0.05	0.01	0.33	0	0.24	0.31	0.1	0.04	0.45	0	0.01	0.01	0
RMFSE	1	1.03	1.12	0.99	1.23	1.03	1.02	1	1.19	0.99	1.23	1.13	1.01	1.02
$R^2 = 0.3$														
TPR	0.98	0.98	0.47	0.95	0.63	0.59	0.62	0.98	0.56	0.92	0.4	0.37	0.7	1
FPR	0.03	0.03	0	0	0.11	0	0	0.01	0	0	0.02	0	0.01	0.88
RMSE(β)	1	1.38	2.22	1.02	3.11	1.67	1.71	1.1	2.38	1.01	2.87	2.44	1.6	4.62
TrueModel	0.04	0.04	0.02	0.34	0	0.01	0.01	0.2	0.03	0.39	0	0.01	0.01	0
RMFSE	1	1.01	1.1	1	1.16	1.05	1.03	0.99	1.09	1	1.14	1.13	1.05	1.03
T=150, p=200														
$R^2 = 0.7$														
TPR	1	1	0.47	0.99	0.42	0.8	0.93	1	0.56	0.88	0.37	0.57	0.89	0.92
FPR	0.04	0.04	0	0	0.02	0	0	0.01	0	0	0.01	0	0.01	0.38
RMSE(β)	1	1.35	4.06	0.85	4.4	1.92	1.11	1.22	4.04	1.39	4.43	3.26	1.45	4
TrueModel	0.06	0.06	0.01	0.61	0	0.22	0.53	0	0.01	0.43	0	0.04	0.07	0
RMFSE	1	1.03	1.48	0.97	1.51	1.06	0.98	0.99	1.34	0.99	1.52	1.31	0.99	1.2
$R^2 = 0.5$														
TPR	0.99	0.98	0.49	0.89	0.34	0.58	0.74	0.98	0.53	0.7	0.34	0.44	0.76	0.92
FPR	0.04	0.04	0	0	0.01	0	0	0.01	0	0	0.01	0	0.01	0.38
RMSE(β)	1	1.31	2.62	1.05	2.92	2	1.44	1.14	2.62	1.49	2.92	2.63	1.5	2.94
TrueModel	0.06	0.06	0.01	0.41	0	0.01	0.12	0.05	0	0.1	0	0	0.01	0
RMFSE	1	1.04	1.19	0.99	1.23	1.12	1.03	0.99	1.16	1.04	1.2	1.16	0.99	1.07
$R^2 = 0.3$														
TPR	0.89	0.89	0.48	0.76	0.37	0.45	0.46	0.88	0.34	0.4	0.37	0.34	0.6	0.99
FPR	0.04	0.04	0	0	0.01	0	0	0.01	0	0	0.01	0	0.01	0.39
RMSE(β)	1	1.35	1.92	1.11	2.12	1.56	1.58	1.2	1.79	1.51	2.11	2.74	1.46	4.6
TrueModel	0.03	0.03	0	0.09	0	0	0.01	0.11	0	0	0	0	0.01	0
RMFSE	1	1.02	1.02	0.98	1.16	1.03	1.04	1	1.05	1.06	1.15	2.07	1.05	1.22

Table 3: Simulation results for pseudo signal variables.

T=200, p=200	Lasso	A-Lasso	S-R-BL-t	S-R-BL-t-r	S-R-BL	S-R-CL-t	S-R-CL-t-r	S-R-CL	R-BL-t	R-BL-t-r	R-BL	R-CL-t	R-CL-t-r	R-CL
$R^2 = 0.7$														
TPR	1	1	0.82	0.99	0.88	0.97	0.98	1	0.65	0.99	0.74	0.67	0.98	0.69
FPR	0.03	0.03	0	0	0.02	0	0.01	0.02	0	0	0	0	0	0
RMSE(β)	1	1.43	3.2	0.86	3.58	1.67	1.12	2.05	4.47	0.76	4.69	4.39	0.8	4.41
TrueModel	0.08	0.05	0.08	0.7	0.01	0.34	0.26	0.01	0	0.93	0.16	0.07	0.91	0.1
RMFSE	1	0.99	1.18	0.92	1.21	0.93	0.88	0.93	1.46	0.96	1.86	1.42	0.94	1.47
$R^2 = 0.5$														
TPR	1	1	0.64	0.99	0.81	0.77	0.81	1	0.64	0.96	0.73	0.68	0.84	0.82
FPR	0.03	0.03	0	0	0.02	0	0	0.02	0	0	0	0	0	0
RMSE(β)	1	1.43	2.86	0.74	3.17	1.8	1.52	1.94	3.15	0.78	3.3	3.09	1.28	3.22
TrueModel	0.15	0.12	0	0.9	0.08	0.03	0.16	0.01	0.02	0.9	0.16	0.14	0.35	0.44
RMFSE	1	1.01	1.1	0.93	1.21	1.03	1.03	1.03	1.22	0.97	1.37	1.13	0.94	1.2
$R^2 = 0.3$														
TPR	0.96	0.95	0.6	0.98	0.72	0.49	0.48	0.98	0.66	0.95	0.72	0.49	0.59	0.99
FPR	0.02	0.02	0	0	0	0	0	0.01	0	0	0	0	0	0
RMSE(β)	1	1.47	2.27	0.76	2.4	1.51	1.84	1.34	2.36	0.82	2.39	1.88	1.56	1.09
TrueModel	0.1	0.07	0	0.91	0.06	0	0	0.47	0	0.78	0.1	0.07	0.05	0.94
RMFSE	1	1.03	1.24	0.94	1.22	1.01	1.04	0.98	1.18	0.99	1.21	1.1	1	1
T=300, p=200														
$R^2 = 0.7$														
TPR	1	1	0.98	1	0.98	1	1	1	0.67	1	0.74	0.64	1	0.64
FPR	0.02	0.02	0	0	0.04	0	0	0.02	0	0	0.01	0	0	0
RMSE(β)	1	1.37	1.98	0.82	3.04	1.47	1.02	2.3	6.12	0.73	6.38	6.04	0.73	6.11
TrueModel	0.16	0.13	0.43	0.74	0	0.52	0.41	0.02	0	0.94	0.15	0.03	1	0.01
RMFSE	1	0.97	1.08	0.94	1.1	1.01	0.98	1.1	1.62	0.93	1.7	1.4	0.97	1.42
$R^2 = 0.5$														
TPR	1	1	0.84	0.99	0.95	0.9	0.91	1	0.63	0.99	0.73	0.74	0.87	0.79
FPR	0.02	0.02	0	0	0.05	0	0	0.02	0	0	0.01	0	0	0
RMSE(β)	1	1.45	2	0.85	2.97	1.67	1.4	2.16	3.99	0.8	4.17	4.06	1.35	4.1
TrueModel	0.07	0.03	0.24	0.79	0.01	0.24	0.2	0.02	0	0.91	0.09	0.21	0.44	0.36
RMFSE	1	1.02	1.08	0.99	1.24	1.04	1.04	1.02	1.24	0.97	1.34	1.24	0.98	1.29
$R^2 = 0.3$														
TPR	0.99	0.99	0.57	0.97	0.83	0.6	0.61	1	0.63	0.94	0.72	0.6	0.7	1
FPR	0.02	0.02	0	0	0.03	0	0	0.01	0	0	0.03	0	0	0
RMSE(β)	1	1.48	2.34	0.82	2.93	1.56	1.79	1.54	2.66	0.85	2.76	1.72	1.49	0.98
TrueModel	0.07	0.06	0.01	0.84	0.02	0	0	0.34	0	0.79	0.08	0.05	0.13	0.99
RMFSE	1	1.02	1.09	1.01	1.13	0.97	0.98	1.01	1.1	0.99	1.19	1	0.96	1.01
T=150, p=200														
$R^2 = 0.7$														
TPR	1	1	0.7	1	0.77	0.94	0.95	1	0.66	1	0.74	0.68	0.97	0.73
FPR	0.03	0.03	0	0	0	0	0	0.01	0	0	0	0	0	0
RMSE(β)	1	1.45	3.81	0.68	4.11	1.84	1.22	2.09	4.18	0.68	4.39	4.15	0.88	4.17
TrueModel	0.09	0.07	0.01	0.94	0.05	0.21	0.16	0	0.02	1	0.12	0.06	0.84	0.16
RMFSE	1	1.01	1.47	0.97	1.95	1.06	1.03	1.13	1.73	0.99	2.06	1.82	1.01	1.91
$R^2 = 0.5$														
TPR	0.99	0.99	0.64	1	0.77	0.66	0.71	1	0.66	0.97	0.77	0.66	0.78	0.84
FPR	0.02	0.03	0	0	0	0	0	0.01	0	0	0.02	0	0	0.01
RMSE(β)	1	1.45	2.69	0.72	2.92	1.95	1.67	1.86	2.8	0.79	2.94	2.83	1.34	2.82
TrueModel	0.07	0.06	0.01	0.98	0.13	0	0.05	0.02	0.01	0.88	0.17	0.06	0.22	0.47
RMFSE	1	1	1.11	1.02	1.34	1.11	1	1.05	1.13	0.99	1.34	1.13	1.02	1.16
$R^2 = 0.3$														
TPR	0.92	0.92	0.58	0.96	0.73	0.42	0.42	0.98	0.64	0.86	0.74	0.53	0.56	0.98
FPR	0.03	0.03	0	0	0	0	0	0.01	0	0	0.05	0	0	0.01
RMSE(β)	1	1.53	1.88	0.77	1.99	1.36	1.58	1.17	1.91	0.83	1.97	2.03	1.28	1.27
TrueModel	0.02	0.01	0	0.8	0.12	0	0	0.52	0	0.49	0.13	0.12	0.06	0.93
RMFSE	1	1.05	1.08	0.99	1.13	1.01	1.04	0.95	1.08	0.99	1.12	1.13	1.06	1.01

Table 4: Simulation results for strongly collinear noise variables due to a persistent unobserved common factor.

T=200, p=200	Lasso	A-Lasso	S-R-BL-t	S-R-BL-t-r	S-R-BL	S-R-CL-t	S-R-CL-t-r	S-R-CL	R-BL-t	R-BL-t-r	R-BL	R-CL-t	R-CL-t-r	R-CL
$R^2 = 0.7$														
TPR	1	1	0.61	1	0.77	1	1	1	0.66	1	0.78	1	1	1
FPR	0.04	0.04	0	0	0.02	0	0	0	0	0	0	0	0	0
RMSE(β)	1	1.41	3.77	0.74	5.23	0.76	0.81	0.75	5.13	0.74	5.41	0.78	0.76	0.75
TrueModel	0.02	0.02	0	0.99	0.06	0.98	0.83	0.98	0	1	0.19	0.98	0.99	1
RMFSE	1	1	1.18	0.91	1.37	0.91	0.96	0.92	1.35	0.92	1.62	0.91	0.92	0.91
$R^2 = 0.5$														
TPR	1	1	0.52	1	0.71	0.8	0.8	1	0.62	1	0.71	0.8	0.83	1
FPR	0.04	0.03	0	0	0.01	0	0	0	0	0	0	0	0	0
RMSE(β)	1	1.29	2.77	0.75	3.37	1.31	1.45	0.73	3.21	0.73	3.37	1.27	1.33	0.74
TrueModel	0.09	0.06	0	1	0.06	0.22	0.22	0.99	0	1	0.1	0.24	0.3	1
RMFSE	1	1	1.1	0.95	1.46	0.96	1.02	0.94	1.17	0.95	1.45	0.96	0.98	0.93
$R^2 = 0.3$														
TPR	0.96	0.95	0.49	0.96	0.75	0.49	0.51	0.95	0.52	0.96	0.75	0.47	0.66	1
FPR	0.03	0.03	0	0	0	0	0	0	0	0	0	0	0	0.01
RMSE(β)	1	1.34	2.35	0.82	2.53	1.59	1.84	0.84	2.34	0.78	2.53	1.81	1.49	1.04
TrueModel	0.03	0.01	0	0.79	0.15	0	0	0.78	0.01	0.79	0.15	0.03	0.16	0.99
RMFSE	1	1.05	1.06	1	1.28	1.05	0.98	1.02	1.04	1.02	1.28	1.08	1.01	1.03
T=300, p=200														
$R^2 = 0.7$														
TPR	1	1	0.97	1	0.98	1	1	1	0.67	1	0.75	1	1	1
FPR	0.03	0.03	0	0	0.07	0	0	0	0	0	0.01	0	0	0
RMSE(β)	1	1.33	1.41	0.84	4.36	0.72	0.83	0.72	6.42	0.76	6.73	0.72	0.72	0.72
TrueModel	0.09	0.05	0.56	0.69	0	1	0.74	1	0.01	0.88	0.06	1	1	1
RMFSE	1	0.98	0.94	0.98	1.25	0.99	0.96	0.99	1.47	0.99	1.65	0.99	0.98	0.99
$R^2 = 0.5$														
TPR	1	1	0.72	1	0.85	0.94	0.94	1	0.64	1	0.74	0.95	0.93	1
FPR	0.04	0.04	0	0	0.07	0	0	0	0	0	0	0	0	0
RMSE(β)	1	1.42	2.11	0.8	3.85	1.02	1.08	0.73	4.06	0.74	4.27	0.9	1.06	0.72
TrueModel	0.06	0.07	0	0.87	0	0.64	0.55	1	0	0.97	0.1	0.76	0.67	1
RMFSE	1	0.99	1.03	1.03	1.21	1.04	1.04	1.03	1.27	1.03	1.45	1.04	1.03	1.03
$R^2 = 0.3$														
TPR	0.99	0.98	0.5	0.99	0.75	0.64	0.65	0.98	0.62	1	0.71	0.64	0.76	1
FPR	0.03	0.03	0	0	0.04	0	0	0	0	0	0	0	0	0
RMSE(β)	1	1.38	2.25	0.81	3.04	1.56	1.76	0.79	2.8	0.79	2.95	1.51	1.43	0.75
TrueModel	0.05	0.02	0	0.88	0.06	0	0	0.88	0.01	0.95	0.12	0	0.23	1
RMFSE	1	1.02	0.93	0.98	1.18	0.99	1	0.97	1.09	0.95	1.11	1	0.96	0.95
T=150, p=200														
$R^2 = 0.7$														
TPR	1	1	0.56	1	0.78	0.97	0.97	1	0.63	1	0.77	0.93	0.97	1
FPR	0.04	0.04	0	0	0.01	0	0	0	0	0	0	0	0	0.01
RMSE(β)	1	1.35	4.42	0.75	4.85	0.94	0.93	0.81	4.55	0.75	4.89	1.19	0.94	1.08
TrueModel	0.02	0.01	0	1	0.13	0.8	0.79	0.95	0	1	0.2	0.74	0.85	0.98
RMFSE	1	1.01	1.69	1	2.06	1.04	1.02	1.05	1.73	1	2.27	1.05	1.07	1.08
$R^2 = 0.5$														
TPR	0.99	0.99	0.51	0.99	0.75	0.66	0.7	0.99	0.54	0.99	0.75	0.65	0.77	1
FPR	0.04	0.04	0	0	0	0	0	0	0	0	0	0	0	0.03
RMSE(β)	1	1.32	2.82	0.81	3.12	1.6	1.71	0.87	2.84	0.79	3.12	1.6	1.46	0.93
TrueModel	0.06	0.06	0	0.91	0.16	0	0.02	0.9	0	0.94	0.16	0.01	0.17	0.95
RMFSE	1	0.95	1.16	0.99	1.29	1.04	1.09	0.98	1.16	0.98	1.29	1.02	1	0.97
$R^2 = 0.3$														
TPR	0.9	0.89	0.41	0.94	0.74	0.44	0.47	0.85	0.44	0.83	0.74	0.46	0.61	1
FPR	0.03	0.03	0	0	0	0	0	0	0	0	0	0	0	0.02
RMSE(β)	1	1.28	1.91	0.84	2.16	1.5	1.64	0.97	1.94	0.93	2.16	1.98	1.34	1.79
TrueModel	0.07	0.02	0	0.69	0.14	0	0	0.49	0	0.38	0.14	0.05	0.12	0.97
RMFSE	1	1.02	1.11	0.96	1.23	0.97	0.93	0.99	1.12	0.99	1.23	1.06	0.95	1

Table 5: Simulation results for temporally correlated and weakly collinear covariates.

T=200, p=200	Lasso	A-Lasso	S-R-BL-t	S-R-BL-t-r	S-R-BL	S-R-CL-t	S-R-CL-t-r	S-R-CL	R-BL-t	R-BL-t-r	R-BL	R-CL-t	R-CL-t-r	R-CL
$R^2 = 0.7$														
TPR	1	1	0.69	1	0.8	0.99	1	1	0.69	1	0.74	0.81	0.97	0.89
FPR	0.04	0.04	0	0	0.02	0	0	0	0	0	0	0	0	0.45
RMSE(β)	1	1.37	3.25	0.72	4.47	1.26	0.84	1.28	4.49	0.69	4.67	2.83	1.01	2.92
TrueModel	0.01	0	0	0.67	0.02	0.66	0.58	0.48	0.01	0.9	0.07	0.2	0.49	0.19
RMFSE	1	1.06	1.21	1.08	1.62	1.12	1.06	1.12	1.77	1.05	1.78	1.15	1.07	1.18
$R^2 = 0.5$														
TPR	1	1	0.61	1	0.75	0.85	0.86	1	0.6	1	0.75	0.6	0.84	0.83
FPR	0.04	0.04	0	0	0	0	0	0	0	0	0	0	0	0.45
RMSE(β)	1	1.32	2.37	0.71	2.86	1.38	1.29	0.99	2.76	0.66	2.82	2.43	1.33	2.26
TrueModel	0.07	0.03	0	0.74	0.12	0.3	0.29	0.56	0.01	0.95	0.16	0.01	0.2	0.07
RMFSE	1	1.03	1.03	0.97	1.11	0.97	0.98	0.99	1.06	0.95	1.11	1	0.98	1.01
$R^2 = 0.3$														
TPR	0.97	0.96	0.53	0.96	0.73	0.59	0.63	0.96	0.53	0.93	0.73	0.43	0.72	0.93
FPR	0.04	0.04	0	0	0	0	0	0	0	0	0	0	0	0.77
RMSE(β)	1	1.35	1.96	0.77	2.08	1.57	1.55	1.11	2.02	0.8	2.08	2.08	1.3	2.48
TrueModel	0.06	0.02	0	0.67	0.07	0	0	0.47	0	0.64	0.07	0	0.08	0.01
RMFSE	1	1.01	0.99	0.91	0.98	0.98	0.96	0.91	1.02	0.89	0.98	1.02	0.97	0.99
T=300, p=200														
$R^2 = 0.7$														
TPR	1	1	0.93	1	0.99	1	1	1	0.81	1	0.78	0.97	1	1
FPR	0.04	0.04	0	0.01	0.04	0	0.01	0.01	0.01	0	0	0	0	0.86
RMSE(β)	1	1.37	1.71	0.85	4.08	1.11	0.95	1.22	3.51	0.75	5.45	1.89	0.79	2.35
TrueModel	0.07	0.05	0.58	0.28	0	0.71	0.34	0.22	0.03	0.66	0.05	0.58	0.53	0.12
RMFSE	1	0.99	1.04	0.93	1.2	0.97	0.96	0.97	1.23	0.93	1.43	0.98	0.93	0.98
$R^2 = 0.5$														
TPR	1	1	0.77	1	0.93	0.96	0.97	1	0.63	1	0.73	0.77	0.91	0.99
FPR	0.04	0.04	0	0	0.06	0	0	0.01	0	0	0	0	0	0.88
RMSE(β)	1	1.34	2.17	0.82	3.48	1.21	1.09	1.11	2.95	0.76	3.68	2.22	1.33	2.08
TrueModel	0.06	0.02	0.03	0.45	0	0.52	0.37	0.36	0.02	0.72	0.07	0.1	0.24	0.08
RMFSE	1	1.04	1.04	1.02	1.18	1.04	1.03	1.04	1.02	1.01	1.26	1.03	0.94	1.03
$R^2 = 0.3$														
TPR	0.99	0.98	0.59	0.97	0.78	0.72	0.72	0.99	0.52	0.98	0.74	0.57	0.75	1
FPR	0.04	0.04	0	0	0.02	0	0	0.01	0	0	0	0	0	0.99
RMSE(β)	1	1.33	1.91	0.83	2.5	1.53	1.55	1.05	2.19	0.77	2.41	2.16	1.45	1.78
TrueModel	0.02	0	0	0.62	0.06	0	0	0.38	0.01	0.79	0.07	0	0.06	0
RMFSE	1	1.05	1.1	1.11	1.23	1.13	1.09	1.1	1.16	1.09	1.23	1.08	1.05	1.09
T=150, p=200														
$R^2 = 0.7$														
TPR	1	1	0.66	1	0.75	0.97	0.99	1	0.66	1	0.74	0.76	0.93	0.83
FPR	0.05	0.04	0	0	0	0	0	0	0	0	0	0	0	0.21
RMSE(β)	1	1.33	3.27	0.69	3.89	1.28	0.84	1.22	3.78	0.65	3.89	2.76	1.15	2.73
TrueModel	0.01	0.01	0.01	0.84	0.08	0.7	0.71	0.65	0	0.98	0.08	0.15	0.44	0.21
RMFSE	1	1.04	1.32	0.98	1.77	1.03	0.95	1.05	1.69	0.97	1.77	1.23	1.02	1.17
$R^2 = 0.5$														
TPR	1	0.99	0.54	0.99	0.75	0.75	0.79	1	0.52	0.98	0.75	0.56	0.77	0.7
FPR	0.04	0.04	0	0	0	0	0	0	0	0	0	0	0	0.15
RMSE(β)	1	1.28	2.37	0.73	2.51	1.6	1.44	1.09	2.51	0.7	2.51	2.26	1.46	2.29
TrueModel	0.06	0.05	0	0.74	0.14	0	0.08	0.63	0.01	0.9	0.14	0	0.08	0.08
RMFSE	1	1.08	1.2	0.99	1.3	1.02	1.01	1.02	1.13	1	1.3	1.08	1.04	1.09
$R^2 = 0.3$														
TPR	0.92	0.93	0.46	0.92	0.77	0.54	0.55	0.86	0.47	0.8	0.77	0.36	0.61	0.87
FPR	0.03	0.03	0	0	0	0	0	0	0	0	0	0	0	0.26
RMSE(β)	1	1.31	1.78	0.84	1.84	1.56	1.52	1.29	1.79	0.98	1.84	1.92	1.34	3.29
TrueModel	0.05	0	0	0.55	0.14	0	0	0.29	0	0.27	0.14	0	0.03	0.06
RMFSE	1	1.03	1.11	0.98	1.06	1.02	1.05	0.99	1.06	0.99	1.06	1.09	1.04	1.17

Table 6: Simulation results for the case that all variables are collinear with the true signals.

series name	sort explanation
AMDMUOx	unfilled orders for durable goods
INDPRO	IP index
UNRATE	civilian unemployment rate
CPIAUCSL	CPI: all items
HOUSTNE	housing starts, northeast

Table 7: Macroeconomic series used in the out of sample forecasting exercise. These are monthly series included in the FRED-MD dataset.

Out of Sample Forecasting Exercise

	INDPRO			UNRATE			HOUSTNE		
	Sel.Var.	r- RMSFE full sample	r- RMSFE crisis period	Sel.Var.	r- RMSFE full sample	r- RMSFE crisis period	Sel.Var.	r- RMSFE full sample	r- RMSFE crisis period
Lasso	12.46	**0.93**$_{\diamond\diamond}$	0.86	13.34	**0.89**$^{**}_{\diamond\diamond}$	0.87	9.81	$0.89^{**}_{\diamond\diamond}$	0.9
Ad. Lasso	12.02	$0.99_{\diamond}$	0.85	12.94	**0.89**$^{**}_{\diamond\diamond}$	0.79	9.53	**0.87**$^{**}_{\diamond}$	0.87
FA-AR		1.02	0.9		0.92^{*}	**0.74**		$0.9^{**}_{\diamond}$	**0.86**
AR		1	1		1	1		1	1
S-R-BL-t	2.29	**0.94**	**0.81**	2.52	2.64	3.69	2.21	1.48	1.54
S-R-BL-t-r	16.41	1.06	0.85	17.4	0.95	0.89	14.57	**0.88****	0.87
S-R-BL	7.63	$0.97_{\diamond\diamond}$	0.86	6.79	0.96	0.92	15.1	0.97	0.98
S-R-CL-t	3.06	**0.94**$_{\diamond\diamond}$	**0.82**	2.06	1.05	0.92	2.3	9.45	1.2
S-R-CL-t-r	15.93	1	0.89	15	0.93^{*}	0.83	12.79	0.9^{*}	**0.84**
S-R-CL	9.52	$1.14_{\diamond\diamond}$	**0.84**	12.79	$0.92^{*}_{\diamond}$	0.83	17.43	1.58	2.72
R-BL-t	3.23	5.1	3.33	3.78	3.67	3.62	6.82	3.03	3.4
R-BL-t-r	17.3	1.08	1	17.08	**0.91**$^{*}_{\diamond}$	**0.78**	14.88	**0.87****	**0.86**
R-BL	128	2.78	1.66	128	2.6	4.24	77.28	2	2.78
R-CL-t	8.85	7.11	2.96	10.95	10.15	18.58	6.9	78.25	6.13
R-CL-t-r	16.86	0.98	0.84	16.49	0.96	**0.77**	14.86	0.89^{**}	0.9
R-CL	121.06	6.71	2.42	119.29	4.93	7.56	46.9	3.81	7.23

	AMDMNOx			CPIAUCSL		
	Sel.Var.	r- RMSFE full sample	r- RMSFE crisis period	Sel.Var.	r- RMSFE full sample	r- RMSFE crisis period
Lasso	20.94	**0.9**$_{\diamond\diamond}$	0.67	18.21	**0.97**	1.03
Ad. Lasso	19.89	**0.91**	**0.66**	17.93	**0.99**	1.02
FA-AR		$0.97_{\diamond\diamond}$	0.76		1.06	1.02
AR		1	1		1	1
S-R-BL-t	2.91	2.33	2.07	2.79	15.85	**0.94**
S-R-BL-t-r	16.77	$0.93_{\diamond\diamond}$	0.77	13.07	1.1	1.22
S-R-BL	7.26	1.06	0.74	7.37	1.02	**0.97**
S-R-CL-t	3.86	$6.74_{\diamond\diamond}$	0.7	4.43	1.08	1.27
S-R-CL-t-r	13.85	**0.91**$_{\diamond\diamond}$	**0.56**	11.42	**1**	0.99
S-R-CL	10.15	$1.18_{\diamond\diamond}$	0.72	11.68	1.06	**0.96**
R-BL-t	2.43	4.29	3.68	3.32	7.56	5.52
R-BL-t-r	16.48	$0.98_{\diamond\diamond}$	**0.62**	16.72	1.18	1.28
R-BL	128	2.27	3.27	127.61	2.72	2.65
R-CL-t	6.87	13.4	4.53	2.52	1.9	1.02
R-CL-t-r	16.32	$0.92_{\diamond\diamond}$	0.68	15.8	1.17	1.23
R-CL	92.38	9.4	11.11	103.71	3.58	2.11

Table 8 : Out of sample forecasting evaluation for 5 key macroeconomic series, using the FRED-MD, 124 monthly, macroeconomic and financial series. The evaluation is performed over the last 136 observations, that is, from 01-Sep-2004, until the 01-Dec-2015. The crisis period corresponds to observations from 01-Sept-2007 until the 01-Dec-2008. The RMSE are relative to the AR(p) model, with p selected by the BIC. The table presents in blue the 3-best methods for each series, and in red the best performning one. The Sel.Var. corresponds to the average number of selected variables over the forecasting interval. The one star (*) (two stars (**)) denotes statistically different forecasts from the AR(1) model at the 10% (5%) significance level, according to the Diebold and Mariano test. The one rhombus (◇) (two rhombus (◇◇)) denotes statistically different forecasts from the AR(1) model at the 10 (5%) significance level, according to the forecast fluctuation test.

11 Appendix

11.1 Auxiliary Lemma

The following Lemma provides supporting results for the main Theorems of the paper.

Lemma 1 *Let $\boldsymbol{A}_T = \left(a_{ijT}\right)$ be a symmetric $p \times p$ matrix with eigenvalues $\mu_1 \leq \mu_2 \leq ... \leq \mu_p$ such that for $\varepsilon > 0$ independent of p, $\varepsilon \leq \mu_1 \leq \mu_p \leq 1/\varepsilon < \infty$.Then $||\boldsymbol{A}_T||_2 = O\left(1\right)$ and $||\boldsymbol{A}_T^{-1}||_2 = O\left(1\right)$.*

Proof.

$$||\boldsymbol{A}_T||_2^2 \;=\; \max_{1 \leq j \leq p} \left\{\mu_j\right\} = \mu_p = O\left(1\right), \tag{29}$$

and

$$||\boldsymbol{A}_T||_2 \;=\; O\left(1\right). \tag{30}$$

Also,

$$||\boldsymbol{A}_T^{-1}||_2^2 = \max_{1 \leq j \leq p} \left\{\frac{1}{\mu_j}\right\} = 1/\mu_1 = O\left(1\right), \text{and } ||\boldsymbol{A}_T^{-1}||_2 = O\left(1\right). \tag{31}$$

■

11.2 Proof of Theorem 1

11.2.1 Parameter consistency

Consider the true parameter vector of model (1) given by $\boldsymbol{\beta} = \boldsymbol{\Sigma}_x^{-1}\boldsymbol{\Sigma}_{xy}$. We then have that:

$$\begin{aligned}\hat{\boldsymbol{\beta}} - \boldsymbol{\beta} = T\left(\hat{\boldsymbol{\Sigma}}_x\right)^{-1} T(\hat{\boldsymbol{\Sigma}}_{xy}) - \Sigma_x^{-1}\boldsymbol{\Sigma}_{xy} = T\left(\hat{\boldsymbol{\Sigma}}_x\right)^{-1} T(\hat{\boldsymbol{\Sigma}}_{xy}) - \\ T\left(\hat{\boldsymbol{\Sigma}}_x\right)^{-1} \boldsymbol{\Sigma}_{xy} + T\left(\hat{\boldsymbol{\Sigma}}_x\right)^{-1} \boldsymbol{\Sigma}_{xy} - \Sigma_x^{-1}\boldsymbol{\Sigma}_{xy} \\ = T\left(\hat{\boldsymbol{\Sigma}}_x\right)^{-1}\left(T(\hat{\boldsymbol{\Sigma}}_{xy}) - \boldsymbol{\Sigma}_{xy}\right) + \left(T\left(\hat{\boldsymbol{\Sigma}}_x\right)^{-1} - \boldsymbol{\Sigma}_x^{-1}\right)\boldsymbol{\Sigma}_{xy}.\end{aligned} \tag{32}$$

So,

$$\begin{aligned}\left\|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right\| \leq \left\|T\left(\hat{\boldsymbol{\Sigma}}_x\right)^{-1}\right\| \left\|T(\hat{\boldsymbol{\Sigma}}_{xy}) - \boldsymbol{\Sigma}_{xy}\right\| + \left\|T\left(\hat{\boldsymbol{\Sigma}}_x\right)^{-1} - \boldsymbol{\Sigma}_x^{-1}\right\| \left\|\boldsymbol{\Sigma}_{xy}\right\| \\ = \left(\left\|T\left(\hat{\boldsymbol{\Sigma}}_x\right)^{-1} - \boldsymbol{\Sigma}_x^{-1}\right\| + \left\|\boldsymbol{\Sigma}_x^{-1}\right\|\right)\left\|T(\hat{\boldsymbol{\Sigma}}_{xy}) - \boldsymbol{\Sigma}_{xy}\right\| + \left\|T\left(\hat{\boldsymbol{\Sigma}}_x\right)^{-1} - \boldsymbol{\Sigma}_x^{-1}\right\| \left\|\boldsymbol{\Sigma}_{xy}\right\|.\end{aligned} \tag{33}$$

From Theorem 1, in Dendramis et al. (2018), for $\lambda_{p_z} = \kappa\sqrt{\frac{\log p_z}{T}}$, where n_{p_z} is the sparsity parameter of $\mathbf{\Sigma}_z$, κ is a fixed constant, and $p_z = p+1$ we have that

$$\left\|T\left(\hat{\mathbf{\Sigma}}_x\right)^{-1} - \mathbf{\Sigma}_x^{-1}\right\| \leq O_p\left(n_{p_z}\lambda_{p_z}\right), \tag{34}$$

and

$$\left\|T(\hat{\mathbf{\Sigma}}_{xy}) - \mathbf{\Sigma}_{xy}\right\| \leq O_p\left(n_{p_z}\lambda_{p_z}\right). \tag{35}$$

Also from Lemma 1 we have that $\left\|\mathbf{\Sigma}_x^{-1}\right\| = O\left(1\right)$. For the term $\left\|\mathbf{\Sigma}_{xy}\right\|$, first notice that

$$\|\mathbf{\Sigma}_z\|_F^2 = \frac{1}{T^2}\sum_{i=1}^{p}\sum_{j=1}^{p}\left(E\left(\sum_{t=1}^{T}\left(z_{it} - E\left(z_{it}\right)\right)\left(z_{jt} - E\left(z_{jt}\right)\right)\right)\right)^2 \tag{36}$$

$$= \frac{1}{T^2}\sum_{i=1}^{p}\sum_{j=1}^{p}\sum_{t=1}^{T}\sum_{t'=1}^{T} E\left[\left(z_{it} - E\left(z_{it}\right)\right)\left(z_{jt} - E\left(z_{jt}\right)\right)\right]\cdot E\left[\left(z_{it'} - E\left(z_{it'}\right)\right)\left(z_{jt'} - E\left(z_{jt'}\right)\right)\right] \tag{37}$$

$$= O\left(n_{p_z}^2\right), \text{and } \|\mathbf{\Sigma}_z\|_2^2 \leq \|\mathbf{\Sigma}_z\|_F^2. \tag{38}$$

Notice that $\mathbf{\Sigma}_z$ has at most n_{p_z} non zero row elements and $E\left[\left(z_{it} - E\left(z_{it}\right)\right)\left(z_{jt} - E\left(z_{jt}\right)\right)\right]$ is bounded $\forall i, j, t$ which is satisfied when it is a covariance stationary a-mixing process. It then follows that

$$\left\|\mathbf{\Sigma}_{xy}\right\| = \left(\frac{1}{T^2}\sum_{i=1}^{p}\left(E\left(\sum_{t=1}^{T}\left(x_{it} - E\left(x_{it}\right)\right)\left(y_t - E\left(y_t\right)\right)\right)\right)^2\right)^{1/2}$$

$$= \left(\frac{1}{T^2}\sum_{i=1}^{p}\sum_{t=1}^{T}\sum_{t'=1}^{T} E\left[\left(x_{it} - E\left(x_{it}\right)\right)\left(y_{jt} - E\left(y_{jt}\right)\right)\right]\cdot E\left[\left(x_{it'} - E\left(x_{it'}\right)\right)\left(y_{jt'} - E\left(y_{jt'}\right)\right)\right]\right)^{1/2}$$

$$= O\left(\sqrt{n_{p_z}}\right). \tag{39}$$

We conclude that

$$\left\|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right\| = \left\{O_p\left(n_{p_z}\lambda_{p_z}\right) + O\left(1\right)\right\} O_p\left(n_{p_z}\lambda_{p_z}\right) + O_p\left(n_{p_z}\lambda_{p_z}\right) O\left(\sqrt{n_{p_z}}\right), \tag{40}$$

or equivalently

$$\left\|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right\| = O_p\left(n_{p_z}^2\lambda_{p_z}^2\right) + O_p\left(n_{p_z}^{3/2}\lambda_{p_z}\right) \tag{41}$$

11.2.2 Probability Bound

To find the probability bound for the term $\left\|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right\|$ we work as follows

$$\Pr\left(\left\|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right\| > q\right) \leq \tag{42}$$
$$\leq \Pr\left(\left(\left\|T\left(\hat{\boldsymbol{\Sigma}}_x\right)^{-1} - \boldsymbol{\Sigma}_x^{-1}\right\| + \left\|\boldsymbol{\Sigma}_x^{-1}\right\|\right)\left\|T(\hat{\boldsymbol{\Sigma}}_{xy}) - \boldsymbol{\Sigma}_{xy}\right\| + \left\|T\left(\hat{\boldsymbol{\Sigma}}_x\right)^{-1} - \boldsymbol{\Sigma}_x^{-1}\right\|\left\|\boldsymbol{\Sigma}_{xy}\right\| > q\right)$$

$$\begin{aligned}\leq \Pr\left(\left(\left\|T\left(\hat{\boldsymbol{\Sigma}}_x\right)^{-1} - \boldsymbol{\Sigma}_x^{-1}\right\| + \left\|\boldsymbol{\Sigma}_x^{-1}\right\|\right)\left\|T(\hat{\boldsymbol{\Sigma}}_{xy}) - \boldsymbol{\Sigma}_{xy}\right\| > \frac{q}{2}\right) + \\ + \Pr\left(\left\|T\left(\hat{\boldsymbol{\Sigma}}_x\right)^{-1} - \boldsymbol{\Sigma}_x^{-1}\right\|\left\|\boldsymbol{\Sigma}_{xy}\right\| > \frac{q}{2}\right).\end{aligned} \tag{43}$$

$$\begin{aligned}\leq \Pr\left(\left(\left\|T\left(\hat{\boldsymbol{\Sigma}}_x\right)^{-1} - \boldsymbol{\Sigma}_x^{-1}\right\| + \left\|\boldsymbol{\Sigma}_x^{-1}\right\|\right) > \frac{q}{2C_1}\right) + \Pr\left(\left\|T(\hat{\boldsymbol{\Sigma}}_{xy}) - \boldsymbol{\Sigma}_{xy}\right\| > C_1\right) + \\ + \Pr\left(\left\|T\left(\hat{\boldsymbol{\Sigma}}_x\right)^{-1} - \boldsymbol{\Sigma}_x^{-1}\right\| > \frac{q}{2C_2}\right) + \Pr\left(\left\|\boldsymbol{\Sigma}_{xy}\right\| > C_2\right)\end{aligned} \tag{44}$$

$$\begin{aligned}\leq \Pr\left(\left\|T\left(\hat{\boldsymbol{\Sigma}}_x\right)^{-1} - \boldsymbol{\Sigma}_x^{-1}\right\| > \frac{q}{4C_1}\right) + \Pr\left(\left\|\boldsymbol{\Sigma}_x^{-1}\right\| > \frac{q}{4C_1}\right) + \Pr\left(\left\|T(\hat{\boldsymbol{\Sigma}}_{xy}) - \boldsymbol{\Sigma}_{xy}\right\| > C_1\right) + \\ + \Pr\left(\left\|T\left(\hat{\boldsymbol{\Sigma}}_x\right)^{-1} - \boldsymbol{\Sigma}_x^{-1}\right\| > \frac{q}{2C_2}\right) + \Pr\left(\left\|\boldsymbol{\Sigma}_{xy}\right\| > C_2\right).\end{aligned} \tag{45}$$

Omitting the constants, and assuming $C_1 \geq n_{p_z}\lambda_{p_z}$, we have that

$$\Pr\left(\left\|T\left(\hat{\boldsymbol{\Sigma}}_{xy}\right) - \boldsymbol{\Sigma}_{xy}\right\| > C_1\right) < p_z^2 D_1 \exp\left(-D_2 T \lambda_{p_z}^2\right) + p_z D_3 \exp\left(-D_4 T \lambda_{p_z}\right), \tag{46}$$

for D_1, D_2, D_3, D_4 sufficiently large constants. When C_2 is large enough, such that $C_2 \geq \sqrt{n_{p_z}}$, we have that

$$\Pr\left(\left\|\boldsymbol{\Sigma}_{xy}\right\| > C_2\right) = 0\,. \tag{47}$$

For

$$\frac{q}{2C_2} \geq n_{p_z}\lambda_{p_z} \Rightarrow q \geq 2C_2 n_{p_z}\lambda_{p_z} \Rightarrow q \geq 2n_{p_z}^{3/2}\lambda_{p_z}, \tag{48}$$

we have

$$\Pr\left(\left\|T\left(\hat{\boldsymbol{\Sigma}}_x\right)^{-1} - \boldsymbol{\Sigma}_{xx}^{-1}\right\| > \frac{q}{2C_2}\right) < p_z^2 D_1 \exp\left(-D_2 T \lambda_{p_z}^2\right) + p_z D_3 \exp\left(-D_4 T \lambda_{p_z}\right). \tag{49}$$

When C_1 is a sequence such that

$$\frac{q}{4C_1} \geq 1 \Rightarrow C_1 \leq \frac{q}{4}, \tag{50}$$

it follows that

$$\Pr\left(\left\|\boldsymbol{\Sigma}_x^{-1}\right\| > \frac{q}{4C_1}\right) = 0. \tag{51}$$

When

$$\frac{q}{4C_1} \geq n_{p_z}\lambda_{p_z} \Rightarrow C_1 \leq \frac{q}{4n_{p_z}\lambda_{p_z}}, \tag{52}$$

we have that

$$\Pr\left(\left\|T\left(\hat{\boldsymbol{\Sigma}}_x\right)^{-1} - \boldsymbol{\Sigma}_x^{-1}\right\| > \frac{q}{4C_1}\right) < p_z^2 D_1 \exp\left(-D_2 T\lambda_{p_z}^2\right) + p_z D_3 \exp\left(-D_4 T\lambda_{p_z}\right). \tag{53}$$

For $q = n_p^{3/2}\lambda_p$, all inequalities hold for a proper choice of constants C_1, C_2, and overall we have

$$\Pr\left(\left\|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right\| > n_p^{3/2}\lambda_p\right) \leq p_z^2 D_1 \exp\left(-D_2 T\lambda_{p_z}^2\right) + p_z D_3 \exp\left(-D_4 T\lambda_{p_z}\right). \tag{54}$$

11.3 Proof of Theorem 2

11.3.1 Proof for TPR

Notice that

$$E\left[TPR_{p,T}\right] = k^{-1}\sum_{i=1}^{k}\Pr\left[\left|\widehat{t}_i\right| > c_T \mid \beta_i \neq 0\right]. \tag{55}$$

When $\beta_i \neq 0$ we have that

$$\Pr\left[\left|\widehat{t}_i\right| > c_T \left|\beta_i \neq 0\right.\right] = \Pr\left[\left|\frac{\widehat{\beta}_i}{\widehat{se}\left(\widehat{\beta}_i\right)} - \frac{\widehat{\beta}_i}{se\left(\widehat{\beta}_i\right)} + \frac{\widehat{\beta}_i}{se\left(\widehat{\beta}_i\right)}\right| > c_T\right]. \tag{56}$$

But,

$$\left|\frac{\widehat{\beta}_i}{\widehat{se}\left(\widehat{\beta}_i\right)} - \frac{\widehat{\beta}_i}{se\left(\widehat{\beta}_i\right)} + \frac{\widehat{\beta}_i}{se\left(\widehat{\beta}_i\right)}\right| > c_T,$$

will occur when

$$\left|\frac{\widehat{\beta}_i}{se\left(\widehat{\beta}_i\right)}\right| > 2c_T, \text{ and, } \left|\frac{\widehat{\beta}_i}{\widehat{se}\left(\widehat{\beta}_i\right)} - \frac{\widehat{\beta}_i}{se\left(\widehat{\beta}_i\right)}\right| < c_T.$$

Also notice

$$eq.(56) \geq \Pr\left[\left\{\left|\frac{\widehat{\beta}_i}{se\left(\widehat{\beta}_i\right)}\right| > 2c_T\right\} \cap \left\{\left|\frac{\widehat{\beta}_i}{\widehat{se}\left(\widehat{\beta}_i\right)} - \frac{\widehat{\beta}_i}{se\left(\widehat{\beta}_i\right)}\right| < c_T\right\} \; |\beta_i \neq 0\right] \tag{57}$$

$$= 1 - \Pr\left[\left\{\left|\frac{\widehat{\beta}_i}{se\left(\widehat{\beta}_i\right)}\right| < 2c_T\right\} \cup \left\{\left|\frac{\widehat{\beta}_i}{\widehat{se}\left(\widehat{\beta}_i\right)} - \frac{\widehat{\beta}_i}{se\left(\widehat{\beta}_i\right)}\right| > c_T\right\} \; |\beta_i \neq 0\right] \tag{58}$$

$$\geq 1 - \Pr\left[\left|\frac{\widehat{\beta}_i}{se\left(\widehat{\beta}_i\right)}\right| < 2c_T\right] - \Pr\left[\left|\frac{\widehat{\beta}_i}{\widehat{se}\left(\widehat{\beta}_i\right)} - \frac{\widehat{\beta}_i}{se\left(\widehat{\beta}_i\right)}\right| > c_T\right]. \tag{59}$$

For the first probability term in (59) we have

$$\Pr\left[\left|\frac{\widehat{\beta}_i}{se\left(\widehat{\beta}_i\right)}\right| < 2c_T\right] = \Pr\left[\;\left|\widehat{\beta}_i\right| < 2c_T\sqrt{\left[\sigma_u^2\frac{1}{T}\boldsymbol{\Sigma}_x^{-1}\right]_{ii}}\;\right] \tag{60}$$

$$= \Pr\left[\;\left|\widehat{\beta}_i - \beta_i + \beta_i\right| < 2c_T\sqrt{\left[\sigma_u^2\frac{1}{T}\boldsymbol{\Sigma}_x^{-1}\right]_{ii}}\;\right]. \tag{61}$$

It is true that when

$$|\beta_i| - 2c_T\sqrt{\left[\sigma_u^2\frac{1}{T}\boldsymbol{\Sigma}_x^{-1}\right]_{ii}} > 0,$$

we have

$$\Pr\left[\left|\widehat{\beta}_i - \beta_i + \beta_i\right| < 2c_T\sqrt{\left[\sigma_u^2\frac{1}{T}\boldsymbol{\Sigma}_x^{-1}\right]_{ii}}\right] \leq \tag{62}$$

$$\leq \Pr\left[\left|\widehat{\beta}_i - \beta_i\right| > \; |\beta_i| - 2c_T\sqrt{\left[\sigma_u^2\frac{1}{T}\boldsymbol{\Sigma}_x^{-1}\right]_{ii}}\right] \leq p_z^2 D_1 \exp\left(-D_2 T \lambda_{p_z}^2\right) + p_z D_3 \exp\left(-D_4 T \lambda_{p_z}\right), \tag{63}$$

with further assuming that

$$|\beta_i| - 2\frac{c_T}{\sqrt{T}}\sqrt{\left[\sigma_u^2\boldsymbol{\Sigma}_x^{-1}\right]_{ii}} \geq n_p^{3/2}\lambda_p (\geq 0). \tag{64}$$

For the second probability term in (59) we have

$$\Pr\left[\left|\frac{\widehat{\beta}_i}{\widehat{se}\left(\widehat{\beta}_i\right)} - \frac{\widehat{\beta}_i}{se\left(\widehat{\beta}_i\right)}\right| > c_T\right] = \Pr\left[\left|\frac{\widehat{\beta}_i}{se\left(\widehat{\beta}_i\right)}\right|\left|\frac{se\left(\widehat{\beta}_i\right)}{\widehat{se}\left(\widehat{\beta}_i\right)} - 1\right| > c_T\right] \tag{65}$$

$$= \Pr\left[\left|\widehat{\beta}_i\right|\left|\frac{se\left(\widehat{\beta}_i\right)}{\widehat{se}\left(\widehat{\beta}_i\right)} - 1\right| > c_T se\left(\widehat{\beta}_i\right)\right] \tag{66}$$

$$\leq \Pr\left[\left|\frac{se\left(\widehat{\beta}_i\right)}{\widehat{se}\left(\widehat{\beta}_i\right)} - 1\right| > \frac{c_T\, se\left(\widehat{\beta}_i\right)}{C_1}\right] + \Pr\left[\left|\widehat{\beta}_i - \beta_i + \beta_i\right| > C_1\right] \tag{67}$$

$$\leq \Pr\left[\left|\frac{se\left(\widehat{\beta}_i\right)}{\widehat{se}\left(\widehat{\beta}_i\right)} - 1\right| > \frac{c_T\, se\left(\widehat{\beta}_i\right)}{C_1}\right] + \Pr\left[\left|\widehat{\beta}_i - \beta_i\right| + |\beta_i| > C_1\right] \tag{68}$$

$$\leq \Pr\left[\left|\frac{se\left(\widehat{\beta}_i\right)}{\widehat{se}\left(\widehat{\beta}_i\right)} - 1\right| > \frac{c_T\, se\left(\widehat{\beta}_i\right)}{C_1}\right] + \Pr\left[\left|\widehat{\beta}_i - \beta_i\right| > \frac{1}{2}C_1\right] + \Pr\left[|\beta_i| > \frac{1}{2}C_1\right], \tag{69}$$

where $se\left(\widehat{\beta}_i\right) = \sqrt{\left[\sigma_u^2 \frac{1}{T}\mathbf{\Sigma}_x^{-1}\right]_{ii}}$, and $C_1 > 0$. The second probability term in (69) is

$$\Pr\left[\left|\widehat{\beta}_i - \beta_i\right| > \frac{1}{2}C_1\right] \leq p_z^2 D_1 \exp\left(-D_2 T \lambda_{p_z}^2\right) + p_z D_3 \exp\left(-D_4 T \lambda_{p_z}\right), \tag{70}$$

when

$$\frac{1}{2}C_1 \geq n_p^{3/2}\lambda_p (\geq 0). \tag{71}$$

The third probability term in (69) (since (64)) is zero, when $C_1 > 2|\beta_i| > 4c_T se\left(\widehat{\beta}_i\right) + 2n_p^{3/2}\lambda_p$. For the first probability term in (69), first notice

$$\frac{se\left(\widehat{\beta}_i\right)}{\widehat{se}\left(\widehat{\beta}_i\right)} - 1 = \frac{\sqrt{\left[\sigma_u^2 \frac{1}{T}\mathbf{\Sigma}_x^{-1}\right]_{ii}}}{\sqrt{\left[\widehat{\sigma}_u^2 \frac{1}{T} T_\lambda\left(\widehat{\mathbf{\Sigma}}_x\right)^{-1}\right]_{ii}}} - 1 \tag{72}$$

$$= \frac{\frac{\left[\sigma_u^2 \frac{1}{T}\mathbf{\Sigma}_x^{-1}\right]_{ii}}{\left[\widehat{\sigma}_u^2 \frac{1}{T} T_\lambda\left(\widehat{\mathbf{\Sigma}}_x\right)^{-1}\right]_{ii}} - 1}{\frac{\sqrt{\left[\sigma_u^2 \frac{1}{T}\mathbf{\Sigma}_x^{-1}\right]_{ii}}}{\sqrt{\left[\widehat{\sigma}_u^2 \frac{1}{T} T_\lambda\left(\widehat{\mathbf{\Sigma}}_x\right)^{-1}\right]_{ii}}} + 1} \leq \frac{\left[\sigma_u^2 \mathbf{\Sigma}_x^{-1}\right]_{ii}}{\left[\widehat{\sigma}_u^2 T_\lambda\left(\widehat{\mathbf{\Sigma}}_x\right)^{-1}\right]_{ii}} - 1, \tag{73}$$

since $\frac{\sqrt{\left[\sigma_u^2 \frac{1}{T}\mathbf{\Sigma}_x^{-1}\right]_{ii}}}{\sqrt{\left[\widehat{\sigma}_u^2 \frac{1}{T} T_\lambda\left(\widehat{\mathbf{\Sigma}}_x\right)^{-1}\right]_{ii}}} + 1 > 1$. Then,

$$\Pr\left[\left|\frac{se\left(\widehat{\beta}_i\right)}{\widehat{se}\left(\widehat{\beta}_i\right)} - 1\right| > \frac{c_T\, se\left(\widehat{\beta}_i\right)}{C_1}\right] \leq \Pr\left[\left|\frac{\left[\sigma_u^2 \mathbf{\Sigma}_x^{-1}\right]_{ii}}{\left[\widehat{\sigma}_u^2 T_\lambda\left(\widehat{\mathbf{\Sigma}}_x\right)^{-1}\right]_{ii}} - \frac{\left[\widehat{\sigma}_u^2 T_\lambda\left(\widehat{\mathbf{\Sigma}}_x\right)^{-1}\right]_{ii}}{\left[\widehat{\sigma}_u^2 T_\lambda\left(\widehat{\mathbf{\Sigma}}_x\right)^{-1}\right]_{ii}}\right| > \frac{c_T\, se\left(\widehat{\beta}_i\right)}{C_1}\right]$$

$$= \Pr\left[\left|\frac{1}{\left[\widehat{\sigma}_u^2 T_\lambda\left(\widehat{\mathbf{\Sigma}}_x\right)^{-1}\right]_{ii}}\right| \left|\left[\sigma_u^2 \mathbf{\Sigma}_x^{-1}\right]_{ii} - \left[\widehat{\sigma}_u^2 T_\lambda\left(\widehat{\mathbf{\Sigma}}_x\right)^{-1}\right]_{ii}\right| > \frac{c_T\, se\left(\widehat{\beta}_i\right)}{C_1}\right] \tag{74}$$

$$\leq \Pr\left[\left|\frac{1}{\left[\widehat{\sigma}_u^2 T_\lambda\left(\widehat{\mathbf{\Sigma}}_x\right)^{-1}\right]_{ii}}\right| > C_2\right] + \Pr\left[\left|\left[\sigma_u^2 \mathbf{\Sigma}_x^{-1}\right]_{ii} - \left[\widehat{\sigma}_u^2 T_\lambda\left(\widehat{\mathbf{\Sigma}}_x\right)^{-1}\right]_{ii}\right| > \frac{c_T\, se\left(\widehat{\beta}_i\right)}{C_1 C_2}\right]$$

$$\leq \Pr\left[\left|\left[\widehat{\sigma}_u^2 T_\lambda\left(\widehat{\mathbf{\Sigma}}_x\right)^{-1}\right]_{ii}\right| > 1/C_2\right]$$

$$+ \Pr\left[\left|\left[\sigma_u^2 \mathbf{\Sigma}_x^{-1}\right]_{ii} - \left[\widehat{\sigma}_u^2 \mathbf{\Sigma}_x^{-1}\right]_{ii} + \left[\widehat{\sigma}_u^2 \mathbf{\Sigma}_x^{-1}\right]_{ii} - \left[\widehat{\sigma}_u^2 T_\lambda\left(\widehat{\mathbf{\Sigma}}_x\right)^{-1}\right]_{ii}\right| > \frac{c_T\, se\left(\widehat{\beta}_i\right)}{C_1 C_2}\right] \tag{75}$$

$$
\begin{aligned}
&\leq \Pr\left[\left|\left[\widehat{\sigma}_u^2 T_\lambda\left(\widehat{\boldsymbol{\Sigma}}_x\right)^{-1}\right]_{ii} - \left[\sigma_u^2 T_\lambda\left(\widehat{\boldsymbol{\Sigma}}_x\right)^{-1}\right]_{ii} + \left[\sigma_u^2 T_\lambda\left(\widehat{\boldsymbol{\Sigma}}_x\right)^{-1}\right]_{ii}\right| > 1/C_2\right] \\
&\quad + \Pr\left[\left|\left[\sigma_u^2 \boldsymbol{\Sigma}_x^{-1}\right]_{ii} - \left[\widehat{\sigma}_u^2 \boldsymbol{\Sigma}_x^{-1}\right]_{ii}\right| > \frac{1}{2}\frac{c_T\, se\left(\widehat{\beta}_i\right)}{C_1 C_2}\right] \\
&\quad + \Pr\left[\left|\left[\widehat{\sigma}_u^2 \boldsymbol{\Sigma}_x^{-1}\right]_{ii} - \left[\widehat{\sigma}_u^2 T_\lambda\left(\widehat{\boldsymbol{\Sigma}}_x\right)^{-1}\right]_{ii}\right| > \frac{1}{2}\frac{c_T\, se\left(\widehat{\beta}_i\right)}{C_1 C_2}\right]
\end{aligned} \tag{76}
$$

$$
\begin{aligned}
&\leq \Pr\left[\left|\left[T_\lambda\left(\widehat{\boldsymbol{\Sigma}}_x\right)^{-1}\right]_{ii}\left(\widehat{\sigma}_u^2 - \sigma_u^2\right)\right| > \frac{1}{2C_2}\right] + \Pr\left[\left|\left[\sigma_u^2 T_\lambda\left(\widehat{\boldsymbol{\Sigma}}_x\right)^{-1}\right]_{ii}\right| > \frac{1}{2C_2}\right] + \\
&\quad + \Pr\left[\left|\sigma_u^2 - \widehat{\sigma}_u^2\right| > \frac{1}{2}\frac{c_T\, se\left(\widehat{\beta}_i\right)}{C_1 C_2}\frac{1}{\left[\boldsymbol{\Sigma}_x^{-1}\right]_{ii}}\right] \\
&\quad + \Pr\left[\left|\widehat{\sigma}_u^2\right|\left|\left[\boldsymbol{\Sigma}_x^{-1}\right]_{ii} - \left[T_\lambda\left(\widehat{\boldsymbol{\Sigma}}_x\right)^{-1}\right]_{ii}\right| > \frac{1}{2}\frac{c_T\, se\left(\widehat{\beta}_i\right)}{C_1 C_2}\right] \leq
\end{aligned} \tag{77}
$$

$$
\begin{aligned}
&\leq \Pr\left[\left|\left[T_\lambda\left(\widehat{\boldsymbol{\Sigma}}_x\right)^{-1}\right]_{ii}\right| > \sqrt{\frac{1}{2C_2}}\right] + \Pr\left[\left|\widehat{\sigma}_u^2 - \sigma_u^2\right| > \sqrt{\frac{1}{2C_2}}\right] \\
&\quad + \Pr\left[\left|\left[T_\lambda\left(\widehat{\boldsymbol{\Sigma}}_x\right)^{-1}\right]_{ii}\right| > \frac{1}{2C_2}\frac{1}{\sigma_u^2}\right] + \\
&\Pr\left[\left|\sigma_u^2 - \widehat{\sigma}_u^2\right| > \frac{1}{2}\frac{c_T\, se\left(\widehat{\beta}_i\right)}{C_1 C_2}\frac{1}{\left[\boldsymbol{\Sigma}_x^{-1}\right]_{ii}}\right] + \Pr\left[\left|\widehat{\sigma}_u^2\right| > \sqrt{\frac{1}{2}\frac{c_T\, se\left(\widehat{\beta}_i\right)}{C_1 C_2}}\right] \\
&\quad + \Pr\left[\left|\left[\boldsymbol{\Sigma}_x^{-1}\right]_{ii} - \left[T_\lambda\left(\widehat{\boldsymbol{\Sigma}}_x\right)^{-1}\right]_{ii}\right| > \sqrt{\frac{1}{2}\frac{c_T\, se\left(\widehat{\beta}_i\right)}{C_1 C_2}}\right].
\end{aligned} \tag{78}
$$

Set

$$
\omega_1 = \min\left(\sqrt{\frac{1}{2C_2}}, \frac{1}{2C_2}\frac{1}{\sigma_u^2}\right) = \frac{1}{2C_2}\frac{1}{\sigma_u^2}, \tag{79}
$$

when σ_u^2 is a constant. Also,

$$\omega_2 = \min\left(\frac{1}{2}\frac{c_T\, se\left(\widehat{\beta}_i\right)}{C_1 C_2}\frac{1}{\left[\boldsymbol{\Sigma}_x^{-1}\right]_{ii}}, \sqrt{\frac{1}{2C_2}}\right) \tag{80}$$

$$= \frac{1}{2}\frac{c_T\, se\left(\widehat{\beta}_i\right)}{C_1 C_2}\frac{1}{\left[\boldsymbol{\Sigma}_x^{-1}\right]_{ii}}. \tag{81}$$

For $c_T se\left(\widehat{\beta}_i\right) = O\left(\frac{(\pi \log(p))^{1/2}}{\sqrt{T}}\right)$, we have that

$$\begin{aligned} eq.(78) \leq 2\Pr\left[\left|\left[T_\lambda\left(\widehat{\boldsymbol{\Sigma}}_x\right)^{-1}\right]_{ii}\right| > \omega_1\right] + 2\Pr\left[\left|\sigma_u^2 - \widehat{\sigma}_u^2\right| > \omega_2\right] \\ + \Pr\left[\left|\widehat{\sigma}_u^2\right| > \sqrt{\frac{1}{2}\frac{c_T\, se\left(\widehat{\beta}_i\right)}{C_1C_2}}\right] + \\ + \Pr\left[\left|\left[\boldsymbol{\Sigma}_x^{-1}\right]_{ii} - \left[T_\lambda\left(\widehat{\boldsymbol{\Sigma}}_x\right)^{-1}\right]_{ii}\right| > \sqrt{\frac{1}{2}\frac{c_T\, se\left(\widehat{\beta}_i\right)}{C_1C_2}}\right]. \end{aligned} \tag{82}$$

For the second probability term in (82) we have that

$$\Pr\left[\left|\sigma_u^2 - \widehat{\sigma}_u^2\right| > \omega_2\right] \leq \exp\left(-D_5 T^{D_6}\right), \tag{83}$$

for $C_2 = O\left(\frac{1}{T^v}\right)$ and $v > 0$.

For the first probability term in (82) we have that

$$\Pr\left[\left|\left[T_\lambda\left(\widehat{\boldsymbol{\Sigma}}_x\right)^{-1}\right]_{ii}\right| > \omega_1\right] \leq \Pr\left[\left|\left[T_\lambda\left(\widehat{\boldsymbol{\Sigma}}_x\right)^{-1}\right]_{ii} - \left[\boldsymbol{\Sigma}_x^{-1}\right]_{ii} + \left[\boldsymbol{\Sigma}_x^{-1}\right]_{ii}\right| > \omega_1\right] \leq \tag{84}$$

$$\leq \Pr\left[\left|\left[T_\lambda\left(\widehat{\boldsymbol{\Sigma}}_x\right)^{-1}\right]_{ii} - \left[\boldsymbol{\Sigma}_x^{-1}\right]_{ii}\right| > \frac{\omega_1}{2}\right] + \Pr\left[\left|\left[\boldsymbol{\Sigma}_x^{-1}\right]_{ii}\right| > \frac{\omega_1}{2}\right] \tag{85}$$

$$\leq p_z^2 D_1 \exp\left(-D_2 T \lambda_{p_z}^2\right) + p_z D_3 \exp\left(-D_4 T \lambda_{p_z}\right). \tag{86}$$

For the third probability term in (82) we have

$$\begin{aligned}
\Pr\left[\left|\widehat{\sigma}_u^2\right| > \sqrt{\frac{1}{2}\frac{c_T\ se\left(\widehat{\beta}_i\right)}{C_1C_2}}\right] &\leq \Pr\left[\left|\widehat{\sigma}_u^2 - \sigma_u^2\right| > \frac{1}{2}\sqrt{\frac{1}{2}\frac{c_T\ se\left(\widehat{\beta}_i\right)}{C_1C_2}}\right] \\
&+ \Pr\left[\left|\sigma_u^2\right| > \frac{1}{2}\sqrt{\frac{1}{2}\frac{c_T\ se\left(\widehat{\beta}_i\right)}{C_1C_2}}\right] \leq \exp\left(-D_5T^{D_6}\right),
\end{aligned} \tag{87}$$

when $|\sigma_u^2| < \frac{1}{2}\sqrt{\frac{1}{2}\frac{c_T\ se\left(\widehat{\beta}_i\right)}{C_1C_2}} (\to \infty)$, with C_1 constant, $C_2 = O\left(\frac{1}{T^v}\right)$ and $v > 0$.

For the fourth probability term in (82) we have that

$$\Pr\left[\left|\left[\mathbf{\Sigma}_x^{-1}\right]_{ii} - \left[T_\lambda\left(\widehat{\mathbf{\Sigma}}_x\right)^{-1}\right]_{ii}\right| > \sqrt{\frac{1}{2}\frac{c_T\ se\left(\widehat{\beta}_i\right)}{C_1C_2}}\right] \leq p_z^2D_1\exp\left(-D_2T\lambda_{p_z}^2\right) + p_zD_3\exp\left(-D_4T\lambda_{p_z}\right) \tag{88}$$

since

$$\sqrt{\frac{1}{2}\frac{c_T\ se\left(\widehat{\beta}_i\right)}{C_1C_2}} \to \infty. \tag{89}$$

So overall we have that

$$\begin{aligned}
&\Pr\left[|\widehat{t}_i| > c_T\,|\beta_i \neq 0\right] \\
&\geq 1 - 4p_z^2D_1\exp\left(-D_2T\lambda_{p_z}^2\right) + p_zD_3\exp\left(-D_4T\lambda_{p_z}\right) - \exp\left(-D_5T^{D_6}\right),
\end{aligned} \tag{90}$$

when $c_T = o\left(T^{c_0}\right)$, $\forall c_0 > 0$, and $se\left(\widehat{\beta}_i\right) = O\left(T^{-\theta}\right)$, for $\theta > c_0$.

Actually, we need $c_Tse\left(\widehat{\beta}\right) = \frac{c_T}{\sqrt{T}}\sqrt{\left[\sigma_u^2\mathbf{\Sigma}_x^{-1}\right]_{ii}} \to 0$. Since, $\left[\sigma_u^2\mathbf{\Sigma}_x^{-1}\right]_{ii}$ is a constant and $c_T = O\left(\log p\right)$, then $c_Tse\left(\widehat{\beta}\right) = O\left(\frac{\log p}{\sqrt{T}}\right)$. We conclude that

$$\begin{aligned}
E\left[TPR_{p,T}\right] &= k^{-1}\sum_{i=1}^{k}\Pr\left[|\widehat{t}_i| > c_T \mid \beta_i \neq 0\right] \\
&\geq 1 - O\left(p_z^2D_1\exp\left(-D_2T\lambda_{p_z}^2\right) + p_zD_3\exp\left(-D_4T\lambda_{p_z}\right)\right).
\end{aligned} \tag{91}$$

11.3.2 Proof for FPR

Notice that

$$E\left[FPR_{l_T,T}\right] = (p-k)^{-1} \sum_{i=k+1}^{k} \Pr\left[|\widehat{t}_i| > c_T \mid \beta_i = 0\right] \tag{92}$$

Also,

$$\Pr\left[|\widehat{t}_i| > c_T \,|\beta_i = 0\right] = \Pr\left[\left|\frac{\widehat{\beta}_i - \beta_i + \beta_i}{se\left(\widehat{\beta}_i\right) - \left(se\left(\widehat{\beta}_i\right) - \widehat{se}\left(\widehat{\beta}_i\right)\right)}\right| > c_T\right] \tag{93}$$

$$\leq \Pr\left[\left|\frac{1}{se\left(\widehat{\beta}_i\right) - \left(se\left(\widehat{\beta}_i\right) - \widehat{se}\left(\widehat{\beta}_i\right)\right)}\right| > \sqrt{c_T}\right] + \Pr\left[\left|\widehat{\beta}_i - \beta_i + \beta_i\right| > \sqrt{c_T}\right] \tag{94}$$

$$\leq C_9 \Pr\left[\left|\frac{\widehat{se}\left(\widehat{\beta}_i\right) - se\left(\widehat{\beta}_i\right)}{se\left(\widehat{\beta}_i\right)}\right| > \sqrt{c_T}\middle|\ \beta_i = 0\right] + \Pr\left[\left|\widehat{\beta}_i - \beta_i + \beta_i\right| > \sqrt{c_T}\right], \tag{95}$$

for some finite constant C_9. The second probability term in (95) is

$$\Pr\left[\left|\widehat{\beta}_i - \beta_i + \beta_i\right| > \sqrt{c_T}\right|\right] \leq \Pr\left[\left|\widehat{\beta}_i - \beta_i\right| > \sqrt{c_T}\middle|\ \beta_i = 0\right] + \Pr\left[|\beta_i| > \sqrt{c_T}|\ \beta_i = 0\right], \tag{96}$$

and

$$\Pr\left[\left|\widehat{\beta}_i - \beta_i\right| > \sqrt{c_T}\middle|\ \beta_i = 0\right] \leq p_z^2 D_1 \exp\left(-D_2 T \lambda_{p_z}^2\right) + p_z D_3 \exp\left(-D_4 T \lambda_{p_z}\right), \text{when } \sqrt{c_T} \geq n_{p_z}^{3/2}\lambda_p.$$

So, overall we have that

$$\Pr\left[\left|\widehat{\beta}_i - \beta_i + \beta_i\right| > \sqrt{c_T}\middle|\ \beta_i = 0\right] \leq p_z^2 D_1 \exp\left(-D_2 T \lambda_{p_z}^2\right) + p_z D_3 \exp\left(-D_4 T \lambda_{p_z}\right).$$

For the first probability term in (95) we have that

$$\Pr\left[\left|\frac{\widehat{se}\left(\widehat{\beta}_i\right)-se\left(\widehat{\beta}_i\right)}{se\left(\widehat{\beta}_i\right)}\right|>\sqrt{c_T}\,\middle|\,\beta_i=0\right] \tag{97}$$

$$\leq \Pr\left[\left|\frac{\left[\widehat{\sigma}_u^2 T\cdot T_\lambda\left(\widehat{\boldsymbol{\Sigma}}_x\right)^{-1}\right]_{ii}}{\left[\sigma_u^2 T\cdot \boldsymbol{\Sigma}_x^{-1}\right]_{ii}}-1\right|>\sqrt{c_T}\,\middle|\,\beta_i=0\right] \tag{98}$$

$$=\Pr\left[\left|\frac{\frac{\widehat{\sigma}_u^2}{\sigma_u^2}\left[T_\lambda\left(\widehat{\boldsymbol{\Sigma}}_x\right)^{-1}\right]_{ii}-\left[\boldsymbol{\Sigma}_x^{-1}\right]_{ii}}{\left[\boldsymbol{\Sigma}_x^{-1}\right]_{ii}}\right|>\sqrt{c_T}\right] \tag{99}$$

$$=\Pr\left[\left|\frac{\widehat{\sigma}_u^2}{\sigma_u^2}\left[T_\lambda\left(\widehat{\boldsymbol{\Sigma}}_x\right)^{-1}\right]_{ii}-\left[\boldsymbol{\Sigma}_x^{-1}\right]_{ii}\right|>\sqrt{c_T}\left[\boldsymbol{\Sigma}_x^{-1}\right]_{ii}\right]. \tag{100}$$

Notice that $\left[\boldsymbol{\Sigma}_x^{-1}\right]_{ii}$ is the *i,i* element of $\boldsymbol{\Sigma}_x^{-1}$ which is constant. Then,

$$\begin{aligned} eq.(100)=\Pr&\left[\left|\frac{\widehat{\sigma}_u^2}{\sigma_u^2}\left[T_\lambda\left(\widehat{\boldsymbol{\Sigma}}_x\right)^{-1}\right]_{ii}-\left[T_\lambda\left(\widehat{\boldsymbol{\Sigma}}_x\right)^{-1}\right]_{ii}+\left[T_\lambda\left(\widehat{\boldsymbol{\Sigma}}_x\right)^{-1}\right]_{ii}-\left[\boldsymbol{\Sigma}_x^{-1}\right]_{ii}\right|>\sqrt{c_T}\left[\boldsymbol{\Sigma}_x^{-1}\right]_{ii}\right]\\ \leq \Pr&\left[\left|\frac{\widehat{\sigma}_u^2}{\sigma_u^2}\left[T_\lambda\left(\widehat{\boldsymbol{\Sigma}}_x\right)^{-1}\right]_{ii}-\left[T_\lambda\left(\widehat{\boldsymbol{\Sigma}}_x\right)^{-1}\right]_{ii}\right|>\frac{\sqrt{c_T}}{2}\left[\boldsymbol{\Sigma}_x^{-1}\right]_{ii}\right]\\ +\Pr&\left[\left|\left[T_\lambda\left(\widehat{\boldsymbol{\Sigma}}_x\right)^{-1}\right]_{ii}-\left[\boldsymbol{\Sigma}_x^{-1}\right]_{ii}\right|>\frac{\sqrt{c_T}}{2}\left[\boldsymbol{\Sigma}_x^{-1}\right]_{ii}\right] \end{aligned} \tag{101}$$

$$\begin{aligned} \leq \Pr&\left[\left|\left[T_\lambda\left(\widehat{\boldsymbol{\Sigma}}_x\right)^{-1}\right]_{ii}\right|>\frac{\sqrt{c_T}}{2C_3}\left[\boldsymbol{\Sigma}_x^{-1}\right]_{ii}\right]+\Pr\left[\left|\frac{\widehat{\sigma}_u^2}{\sigma_u^2}-1\right|>C_3\right]\\ +\Pr&\left[\left|\left[T_\lambda\left(\widehat{\boldsymbol{\Sigma}}_x\right)^{-1}\right]_{ii}-\left[\boldsymbol{\Sigma}_x^{-1}\right]_{ii}\right|>\frac{\sqrt{c_T}}{2}\left[\boldsymbol{\Sigma}_x^{-1}\right]_{ii}\right] \end{aligned} \tag{102}$$

$$\begin{aligned} =\Pr&\left[\left|\left[T_\lambda\left(\widehat{\boldsymbol{\Sigma}}_x\right)^{-1}\right]_{ii}-\left[\boldsymbol{\Sigma}_x^{-1}\right]_{ii}+\left[\boldsymbol{\Sigma}_x^{-1}\right]_{ii}\right|>\frac{\sqrt{c_T}}{2C_3}\left[\boldsymbol{\Sigma}_x^{-1}\right]_{ii}\right]\\ +\Pr\left[\left|\frac{\widehat{\sigma}_u^2}{\sigma_u^2}-1\right|>C_3\right]+\Pr&\left[\left|\left[T_\lambda\left(\widehat{\boldsymbol{\Sigma}}_x\right)^{-1}\right]_{ii}-\left[\boldsymbol{\Sigma}_x^{-1}\right]_{ii}\right|>\frac{\sqrt{c_T}}{2}\left[\boldsymbol{\Sigma}_x^{-1}\right]_{ii}\right] \end{aligned} \tag{103}$$

$$\begin{aligned}\leq\ & \Pr\left[\left|\left[T_\lambda\left(\widehat{\mathbf{\Sigma}}_x\right)^{-1}\right]_{ii}-\left[\mathbf{\Sigma}_x^{-1}\right]_{ii}\right|>\frac{\sqrt{c_T}}{4C_3}\left[\mathbf{\Sigma}_x^{-1}\right]_{ii}\right] \\ & +\Pr\left[\left|\left[\mathbf{\Sigma}_x^{-1}\right]_{ii}\right|>\frac{\sqrt{c_T}}{4C_3}\left[\mathbf{\Sigma}_x^{-1}\right]_{ii}\right]+\Pr\left[\left|\frac{\widehat{\sigma}_u^2}{\sigma_u^2}-1\right|>C_3\right] \\ & +\Pr\left[\left|\left[T_\lambda\left(\widehat{\mathbf{\Sigma}}_x\right)^{-1}\right]_{ii}-\left[\mathbf{\Sigma}_x^{-1}\right]_{ii}\right|>\frac{\sqrt{c_T}}{2}\left[\mathbf{\Sigma}_x^{-1}\right]_{ii}\right].\end{aligned} \tag{104}$$

The first probability term in (104) is

$$\begin{aligned}&\Pr\left[\left|\left[T_\lambda\left(\widehat{\mathbf{\Sigma}}_x\right)^{-1}\right]_{ii}-\left[\mathbf{\Sigma}_x^{-1}\right]_{ii}\right|>\frac{\sqrt{c_T}}{4C_3}\left[\mathbf{\Sigma}_x^{-1}\right]_{ii}\right] \\ &\leq p_z^2D_1\exp\left(-D_2T\lambda_{p_z}^2\right)+p_zD_3\exp\left(-D_4T\lambda_{p_z}\right)\ .\end{aligned} \tag{105}$$

When $\frac{\sqrt{c_T}}{4C_3}\left[\mathbf{\Sigma}_x^{-1}\right]_{ii}>n_{p_z}\lambda_{p_z}$, and C_3 is a large constant, the second probability term in (104) is zero.

The third probability term in (104) is

$$\Pr\left[\left|\frac{\widehat{\sigma}_u^2}{\sigma_u^2}-1\right|>C_3\right|\beta_i=0\right]\leq\exp\left(-D_{10}T^{D_{11}}\right) \tag{106}$$

for a constant C_3.

The fourth probability term in (104) is

$$\begin{aligned}&\Pr\left[\left|\left[T_\lambda\left(\widehat{\mathbf{\Sigma}}_x\right)^{-1}\right]_{ii}-\left[\mathbf{\Sigma}_x^{-1}\right]_{ii}\right|>\frac{\sqrt{c_T}}{2}\left[\mathbf{\Sigma}_x^{-1}\right]_{ii}\right] \\ &\leq p_z^2D_1\exp\left(-D_2T\lambda_{p_z}^2\right)+p_zD_3\exp\left(-D_4T\lambda_{p_z}\right),\end{aligned} \tag{107}$$

when $\frac{\sqrt{c_T}}{2}\left[\mathbf{\Sigma}_x^{-1}\right]_{ii}>n_{p_z}\lambda_{p_z}$.

So overall we have that

$$\begin{aligned}\Pr\left[\left|\widehat{t}_i\right|>c_T\left|\beta_i=0\right.\right]\leq C_9\Pr&\left[\left|\frac{\widehat{se}\left(\widehat{\beta}_i\right)-se\left(\widehat{\beta}_i\right)}{se\left(\widehat{\beta}_i\right)}\right|>\sqrt{c_T}\right|\beta_i=0\right] \\ &+\Pr\left[\left|\widehat{\beta}_i-\beta_i+\beta_i\right|>\sqrt{c_T}\right],\end{aligned} \tag{108}$$

that is,

$$\Pr\left[\left|\widehat{t}_i\right|>c_T\left|\beta_i=0\right.\right]\leq p_z^2D_1\exp\left(-D_2T\lambda_{p_z}^2\right)+p_zD_3\exp\left(-D_4T\lambda_{p_z}\right). \tag{109}$$

We conclude that

$$\begin{aligned} E\left[FPR_{l_T,T}\right] &\leq (p-k)^{-1} O\left(p_z^2 D_1 \exp\left(-D_2 T \lambda_{p_z}^2\right) + p_z D_3 \exp\left(-D_4 T \lambda_{p_z}\right)\right) \\ &= O\left(p_z D_1 \exp\left(-D_2 T \lambda_{p_z}^2\right) + D_3 \exp\left(-D_4 T \lambda_{p_z}\right)\right), \end{aligned} \tag{110}$$

with $p_z = p+1$.

11.4 Proof of Theorem 3

We first define the event of correctly identifying the true model as

$$A_0 = A_0 = \left\{ \sum_{i=1}^{k} I(\widehat{\beta_i \neq 0}) = k \right\} \cap \left\{ \sum_{i=k+1}^{p} I(\widehat{\beta_i \neq 0}) = 0 \right\}. \tag{111}$$

Then,

$$\Pr\left(A_0^c\right) \leq \Pr\left(\sum_{i=1}^{k} I(\widehat{\beta_i \neq 0}) < k \right) + \Pr\left(\sum_{i=k+1}^{p} I(\widehat{\beta_i \neq 0}) > 0 \right). \tag{112}$$

For the second probability term in (112) we have

$$\Pr\left(\sum_{i=k+1}^{p} I(\widehat{\beta_i \neq 0}) > 0 \right) \leq \sum_{i=k+1}^{p} E\left[I(\widehat{\beta_i \neq 0}) > 0 \middle| \beta_i = 0 \right] \tag{113}$$

$$= \sum_{i=k+1}^{p} \Pr\left(|\widehat{t}_i| > c_T \middle| \beta_i = 0 \right) \leq (p-k)\, O\left(p_z D_1 \exp\left(-D_2 T \lambda_{p_z}^2\right) + D_3 \exp\left(-D_4 T \lambda_{p_z}\right)\right) \tag{114}$$

$$\leq O\left(p_z^2 D_1 \exp\left(-D_2 T \lambda_{p_z}^2\right) + p_z D_3 \exp\left(-D_4 T \lambda_{p_z}\right)\right). \tag{115}$$

For the first probability term we have

$$\begin{aligned} &\Pr\left(\sum_{i=1}^{k} I(\widehat{\beta_i \neq 0}) < k \right) = 1 - \Pr\left(\sum_{i=1}^{k} I(\widehat{\beta_i \neq 0}) = k \right) \\ &\leq 1 - 1 + O\left(p_z^2 D_1 \exp\left(-D_2 T \lambda_{p_z}^2\right) + p_z D_3 \exp\left(-D_4 T \lambda_{p_z}\right)\right). \end{aligned} \tag{116}$$

That is,

$$\Pr\left(A_0^c\right) \leq O\left(p_z^2 D_1 \exp\left(-D_2 T \lambda_{p_z}^2\right) + p_z D_3 \exp\left(-D_4 T \lambda_{p_z}\right)\right). \tag{117}$$

Also, for any $\varepsilon > 0$ there exists some $B_\varepsilon < \infty$ such that

$$\begin{aligned} \Pr\left(\sqrt{T}\left\|\widehat{\beta} - \beta\right\| > B_\varepsilon\right) &= \Pr\left(\sqrt{T}\left\|\widehat{\beta} - \beta\right\| > B_\varepsilon \middle| A_0\right)\Pr\left(A_0\right) \\ &+ \Pr\left(\sqrt{T}\left\|\widehat{\beta} - \beta\right\| > B_\varepsilon \middle| A_0^c\right)\Pr\left(A_0^c\right) \\ &\leq \Pr\left(\sqrt{T}\left\|\widehat{\beta} - \beta\right\| > B_\varepsilon \middle| A_0\right) + \Pr\left(A_0^c\right) < \varepsilon. \end{aligned} \tag{118}$$

When we define $F_u = T^{-1}\left\|\hat{u}'\hat{u}\right\|$, for the error term of the second step regression we have that for any $\varepsilon > 0$ there exists some $C_\varepsilon < \infty$ such

$$\begin{aligned} \Pr\left(\sqrt{T}\left|F_u - \sigma_u^2\right| > C_\varepsilon\right) &= \Pr\left(\sqrt{T}\left|F_u - \sigma_u^2\right| > C_\varepsilon \middle| A_0\right)\Pr\left(A_0\right) \\ &+ \Pr\left(\sqrt{T}\left|F_u - \sigma_u^2\right| > C_\varepsilon \middle| A_0^c\right)\Pr\left(A_0^c\right) \\ &\leq \Pr\left(\sqrt{T}\left|F_u - \sigma_u^2\right| > C_\varepsilon \middle| A_0\right) + \Pr\left(A_0^c\right) < \varepsilon. \end{aligned} \tag{119}$$

References

K. M. Abadir, W. Distaso, and F. Zikes. Design-free estimation of variance matrices. *Journal of Econometrics*, 181:165–180, 2014.

A. Antoniadis and J. Fan. Regularization of wavelets approximations. *Journal of the American Statistical Association*, 96:939–967, 2001.

J. Bai and S. Ng. Determining the number of factors in approximate factor models. *Econometrica*, pages 191–221, 2002.

Y. Benjamini and Y. Hochberg. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the American Statistical Association*, 57:289–300, 1995.

J. P. Bickel, Y. Ritov, and A. Tsybakov. Simultaneous analysis of lasso and dantzig selector. *Annals of Statistics*, 37:1705–1732, 2009.

M. Bickel, Y. Wang, and H. Zhou. Optimal sparse volatility matrix estimation for high-dimensional ito processes with measurement errors. *The Annals of Statistics*, 41: 1816–1864, 2013.

P. Bickel and E. Levina. Covariance regularization by thresholding. *The Annals of Statistics*, 36:2577–2604, 2008.

P. Buhlmann. Boosting for high-dimensional linear models," the annals of statistics. *Annals of Statistics*, 34(2):599–583, 2006.

P. Buhlmann and S. van de Geer. *Statistics for High-Dimensional Data: Methods, Theory and Applications*. Springer, 2011.

P. Bühlmann and T. Hothorn. Boosting algorithms: regularization, prediction and model fitting. *Statistical Science*, 22:477–522, 2007.

T. Cai and W. Liu. Adaptive thresholding for sparse covariance matrix estimation. *Journal of the American Statistical Association*, 106:672–684, 2011.

E. Candes and T. Tao. The dantzig selector: Statistical estimation when p is much larger than n. *Annals of Statistics*, 35:2313–2404, 2007.

X. Chen, M. Xu, and W. B. Wu. Covariance and precision matrix estimation for high-dimensional time series. *The Annals of Statistics*, 41:2994–3021, 2013.

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

Y. Dendramis, L. Giraitis, and G. Kapetanios. Estimation of random coefficient time varying covariance matrices for large datasets. *Mimeo Queen Mary University of London*, 2018.

F. X. Diebold and R. S. Mariano. Comparing predictive accuracy. *Journal of Business and Economic Statistics*, 13:253–263, 1995.

D. Donoho and M. Elad. Optimally sparse representation in general (nonorthogonal) dictionaries via l1 minimization. *Proceedings of the National Academy of Sciences*, 100: 2197–2202, 2003.

B. Efron, T. Hastie, I. Johnstone, and R. Tibshirani. Least angle regression. *Annals of Statistics*, 32:407–499, 2004.

J. Fan and X. Han. Estimation of the false discovery proportion with unknown dependence. *Journal of the Royal Statistical Society, Series B*, 79:1143–1164, 2017.

J. Fan and R. Li. Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American Statistical Association*, 96:1348–1360, 2001.

J. Fan and J. Lv. Sure independence screening for ultra-high dimensional feature space. *Journal of Royal Statistical Society Series B*, 70:849–911, 2008.

J. Fan and J. Lv. Asymptotic equivalence of regularization methods in thresholded parameter space. *Journal of the American Statistical Association*, 108:1044–1061, 2013.

J. Fan and R. Song. Sure independence screening in generalized linear models with np-dimensionality. *Annals of Statistics*, 38:3567–3604, 2010.

J. Fan and C. Tang. Tuning parameter selection in high dimensional penalized likelihood. *Journal of the Royal Statistical Society Series B*, 75:531–552, 2013.

J. Fan, R. Samworth, and Y. Wu. Ultra high dimensional variable selection: Beyond the linear model. *Journal of Machine Learning Research*, 10:1829–1853, 2009.

J. Fan, Y. Feng, and R. Song. Nonparametric independence screening in sparse ultra-high-dimensional additive models. *Journal of the American Statistical Association*, 106:544–557, 2011.

J. Fan, X. Han, and W. Gu. Estimating false discovery proportion under arbitrary covariance dependence. *Journal of the American Statistical Association*, 107(499):1019–1035, 2012.

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

J. Fan, Y. Fan, and E. Barut. Adaptive robust variable selection. *Annals of Statistics*, 42: 324–351, 2014.

J. Fan, Y. Liao, and H. Liu. An overview of the estimation of large covariance and precision matrices. *Econometrics Journal*, 19:C1–C32, 2016.

W. Fithian, D. Sun, and J. Taylor. Optimal inference after model selection. *arXiv:1410.2597v4*, 2014.

W. Fithian, J. Taylor, R. J. Tibshirani, and R. Tibshirani. Selective sequential model selection. *arXiv:1512.02565*, 2015.

J. Friedman. Greedy function approximation: A gradient boosting machine. *Annals of Statistics*, 29:1189–1232, 2001.

J. Friedman, T. Hastie, and R. Tibshirani. Additive logistic regression: A statistical view of boosting. *Annals of Statistics*, 28:337–374, 2000.

Y. Gavrilov, Y. Benjamini, and S. K. Sarkar. An adaptive step-down procedure with proven fdr control under independence. *Annals of Statistics*, 37:619–629, 2009.

S. Geisser. The predictive sample reuse method with applications. *J. Amer. Statist. Assoc.*, 70:320–328, 1975.

R. Giacomini and B. Rossi. Forecast comparisons in unstable environments. *Journal of Applied Econometrics*, 25:595–620, 2010.

Zijian Guo, Claude Renaux, Peter Bühlmann, and T. Tony Cai. Group inference in high dimensions with applications to hierarchical testing, 2019.

P. Hall and H. Miller. Using generalized correlation to effect variable selection in very high dimensional problems. *Journal of Computational and Graphical Statistics*, 18:533–550, 2009.

P. Hall, D. M. Titterington, and J. H. Xue. Tilting methods for assessing the influence of components in a classifier. *Journal of the Royal Statistical Society, Series B*, 71:783–803, 2009.

F. Han and H. Liu. Statistical analysis of latent generalized correlation matrix estimation in transelliptical distribution. *Bernoulli*, 23:23–57, 2017.

D. Harvey, S. Leybourne, and P. Newbold. Testing the equality of prediction mean squared errors. *International Journal of Forecasting*, 13:281—291, 1997.

S. Holm. A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, 6:65–70, 1979.

T. Hothorn, P. Bühlmann, T. Kneib, M. Schmid, and B. Hofner. Model-based boosting 2.0. *Journal of Machine Learning Research*, 11:2109–2113, 2010.

J. Huang, J. Horowitz, and S. Ma. Asymptotic properties of bridge estimators in sparse high-dimensional regression models. *Annals of Statistics*, 36:587–613, 2008.

D. R. Hunter and R. Li. Variable selection using mm algorithms. *Annals of Statistics*, 33: 1617– 1642, 2005.

M. Kolar and E. Xing. On time varying undirected graphs. *In Proceedings of the 14th International Conference on Artificial Intelligence and Statistics (AISTATS) 2011 (JMLR)*, 15:407–415, 2011.

O. Ledoit and M. Wolf. A well-conditioned estimator for large-dimensional covariance matrices. *Journal of Multivariate Analysis*, 88:365–411, 2004.

G. Li, H. Peng, J. Zhang, and L. Zhu. Robust rank correlation based screening. *The Annals of Statistics*, 40:1846–1877, 2012a.

R. Li, W. Zhong, and L. Zhu. Feature screening via distance correlation learning. *Journal of the American Statistical Association*, 107:1129–1139, 2012b.

J. Lv and Y. Fan. A unified approach to model selection and sparse recovery using regularized least squares. *Annals of Statistics*, 37:3498–3528, 2009.

N. Meinshausen and P. Bühlmann. High-dimensional graphs and variable selection with the lasso. *Annals of Statistics*, 34:1436–1462, 2006.

J. Opsomer, Y. Wang, and Y. Yang. Nonparametric regression with correlated errors. *Statistical Science*, 16:134–153, 2001.

M. Pourahmadi. *High-Dimensional Covariance Estimation*. Hoboken: John Wiley and Sons, 2013.

M. Stone. Asymptotics for and against cross-validation. *Biometrika*, 64:29–35, 1977.

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

R. J. Tibshirani, J. Taylor, R. Lockhart, and R. Tibshirani. Exact postselection inference for sequential regression procedures. *arXiv:1401.3889*, 2014.

Sara A. van de Geer and Benjamin Stucky. $\$\chi^2\$ - confidencesetsinhigh - dimensionalregression$. 2015.

L. Wasserman and K. Roeder. High-dimensional variable selection. *Annals of Statistics*, 37: 2178–2201, 2009.

C. Zhang and J. Huang. The sparsity and bias of the lasso selection in high-dimensional linear regression. *Annals of Statistics*, 36:1567–1594, 2008.

C. H. Zhang. Nearly unbiased variable selection under minimax concave penalty. *Annals of Statistics*, 38:894–942, 2010.

P. Zhao and B. Yu. On model selection consistency of lasso. *Journal of the Machine Learning Research*, 7:2541–2563, 2006.

Z. Zheng, Y. Fan, and J. Lv. High dimensional thresholded regression and shrinkage effect. *Journal of the Royal Statistical Society B*, 76:627–649, 2014.

H. Zhou and T. Hastie. Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society Series B*, 67:301–320, 2005.

S. Zhou, J. Lafferty, and L. Wasserman. Time varying undirected graphs. *Machine Learning*, 80:295–319, 2010.

H. Zou. The adaptive lasso and its oracle properties. *Journal of the American Statistical Association*, 101:1418–1429, 2006.

H. Zou and H. H. Zhang. On the adaptive elastic-net with a diverging number of parameters. *Annals of Statistics*, 37:1733–1751, 2009.

ΟΙΚΟΝΟΜΙΚΟ
ΠΑΝΕΠΙΣΤΗΜΙΟ
ΑΘΗΝΩΝ

ATHENS UNIVERSITY
OF ECONOMICS
AND BUSINESS

ΣΧΟΛΗ
ΟΙΚΟΝΟΜΙΚΩΝ
ΕΠΙΣΤΗΜΩΝ
SCHOOL OF
ECONOMIC
SCIENCES

ΤΜΗΜΑ
ΟΙΚΟΝΟΜΙΚΗΣ
ΕΠΙΣΤΗΜΗΣ
DEPARTMENT OF
ECONOMICS

Department of Economics
Athens University of Economics and Business

List of Recent Working Papers

2020

01-20 Incubated Early Stage Startuppers in Athens and their Initiatives during the Crisis (2010-2016), Ioannis Besis and Ioanna Sapfo Pepelasis
02-20 Alternative Monetary Policy Rules in an Imperfectly Competitive Dynamic Stochastic General Equilibrium Model, George Alogoskoufis and Stelios Giannoulakis
03-20 Quantifying Qualitative Survey Data: New Insights on the (Ir)Rationality of Firms' Forecasts, Alexandros Botsis, Christoph Görtz and Plutarchos Sakellaris
04-20 Can Country-Specific Interest Rate Factors Explain the Forward Premium Anomaly?, Efthymios Argyropoulos, Nikolaos Elias, Dimitris Smyrnakis and Elias Tzavalis
05-20 Signaling product quality through prices in an oligopoly model with costly consumer search, Minghua Cheny, Konstantinos Serfes and Eleftherios Zacharias
06-20 Thinking ahead of the next big crash: Clues from Athens in classical times, George C. Bitros
07-20 Financial crises, firm-level shocks, and large downturns: Evidence from Greece, Stelios Giannoulakis and Plutarchos Sakellaris
08-20 Notes on the Demand Side Secular Stagnation, George D. Demopoulos and Nicholas A. Yannacopoulos
09-20 Institutions and Macroeconomic Performance: Core vs Periphery Countries in the Eurozone, Tryfonas Christou, Apostolis Philippopoulos and Vanghelis Vassilatos
10-20 Macroeconomic Policy Lessons for Greece, George Economides, Dimitris Papageorgiou and Apostolis Philippopoulos
11-20 Energy Transitions and the role of the EU ETS: The case of Greece, Andriana Vlachou and Georgios Pantelias
12-20 Measuring the Default Risk of Small Business Loans: Improved Credit Risk Prediction using Deep Learning, Yiannis Dendramis, Elias Tzavalis and Aikaterini Cheimarioti
14-20 Saving Democracy by Switching to Corporate-like Governance, George C. Bitros
15-20 The impact of the lockdown on the Greek economy and the role of the Recovery Fund, George Economides, Apostolis Philippopoulos and Vanghelis Vassilatos

2021

01-21 Historical Cycles of the Economy of Modern Greece From 1821 to the Present, George Alogoskoufis
02-21 Greece Before and After the Euro: Macroeconomics, Politics and the Quest for Reforms, George Alogoskoufis
03-21 Commodity money and the price level, George C. Bitros
04-21 Destabilizing asymmetries in central banking: With some enlightenment from money in classical Athens, George C. Bitros
05-21 Exploring the Long-Term Impact of Maximum Markup Deregulation, Athanasios Dimas and Christos Genakos

ΟΙΚΟΝΟΜΙΚΟ
ΠΑΝΕΠΙΣΤΗΜΙΟ
ΑΘΗΝΩΝ

ATHENS UNIVERSITY
OF ECONOMICS
AND BUSINESS

ΣΧΟΛΗ
ΟΙΚΟΝΟΜΙΚΩΝ
ΕΠΙΣΤΗΜΩΝ
SCHOOL OF
ECONOMIC
SCIENCES

ΤΜΗΜΑ
ΟΙΚΟΝΟΜΙΚΗΣ
ΕΠΙΣΤΗΜΗΣ
DEPARTMENT OF
ECONOMICS

Department of Economics
Athens University of Economics and Business

The Department is the oldest Department of Economics in Greece with a pioneering role in organising postgraduate studies in Economics since 1978. Its priority has always been to bring together highly qualified academics and top quality students. Faculty members specialize in a wide range of topics in economics, with teaching and research experience in world-class universities and publications in top academic journals.

The Department constantly strives to maintain its high level of research and teaching standards. It covers a wide range of economic studies in micro-and macroeconomic analysis, banking and finance, public and monetary economics, international and rural economics, labour economics, industrial organization and strategy, economics of the environment and natural resources, economic history and relevant quantitative tools of mathematics, statistics and econometrics.

Its undergraduate program attracts high quality students who, after successful completion of their studies, have excellent prospects for employment in the private and public sector, including areas such as business, banking, finance and advisory services. Also, graduates of the program have solid foundations in economics and related tools and are regularly admitted to top graduate programs internationally. Three specializations are offered:1. Economic Theory and Policy, 2. Business Economics and Finance and 3. International and European Economics. The postgraduate programs of the Department (M.Sc and Ph.D) are highly regarded and attract a large number of quality candidates every year.

For more information:

https://www.dept.aueb.gr/en/econ/