INSTITUTE
OF ECONOMICS

Scuola Superiore
Sant'Anna

LEM | Laboratory of Economics and Management

Institute of Economics
Scuola Superiore Sant'Anna

Piazza Martiri della Libertà, 33 - 56127 Pisa, Italy
ph. +39 050 88.33.43
institute.economics@sssup.it

LEM
WORKING PAPER SERIES

On the impact of serial dependence on penalized regression methods

Simone Tonini [a]
Francesca Chiaromonte [a,b]
Alessandro Giovannelli [c]

[a] Institute of Economics & EMbeDS, Sant'Anna School of Advanced Studies, Pisa, Italy.
[b] Department of Statistics, Penn State University, USA.
[c] Department of Information Engineering, Computer Science and Mathematics, University of L'Aquila, Italy.

2022/21 July 2022
ISSN(ONLINE) 2284-0400

On the Impact of Serial Dependence on Penalized Regression Methods

Simone Tonini*

Institute of Economics & EMbeDS, Sant'Anna School
of Advanced Studies, Pisa, Italy
and
Francesca Chiaromonte
Institute of Economics & EMbeDS,
Sant'Anna School of Advanced Studies, Pisa, Italy.
Department of Statistics,
Penn State University, USA.
and
Alessandro Giovannelli
University of L'Aquila Department of Information Engineering,
Computer Science and Mathematics, L'Aquila, Italy

Abstract

This paper characterizes the impact of covariate serial dependence on the non-asymptotic estimation error bound of penalized regressions (PRs). Focusing on the direct relationship between the degree of cross-correlation between covariates and the estimation error bound of PRs, we show that orthogonal or weakly cross-correlated stationary AR processes can exhibit high spurious correlations caused by serial dependence. We provide analytical results on the distribution of the sample cross-correlation in the case of two orthogonal Gaussian AR(1) processes, and extend and validate them through an extensive simulation study. Furthermore, we introduce a new procedure to mitigate spurious correlations in a time series setting, applying PRs to pre-whitened (ARMA filtered) time series. We show that under mild assumptions our procedure allows both to reduce the estimation error and to develop an effective forecasting strategy. The estimation accuracy of our proposal is validated through additional simulations, as well as an empirical application to a large set of monthly macroeconomic time series relative to the Euro Area.

Keywords: serial dependence, spurious correlation, minimum eigenvalue, penalized regressions.

*Corresponding author: simone.tonini@santannapisa.it

Acknowledgments: The authors wish to thank Marco Lippi for suggesting to develop the theoretical part of this work and for the precious technical support. We are also grateful to Sebastiano Michele Zema and Luca Insolia for their helpful comments and for the stimulating dialogues.

1 Introduction

Much contemporary statistical literature is devoted to the problem of extracting information from large datasets, which are ubiquitous in many fields of science (Fan et al., 2014). In the context of high-dimensional regression problems, where the number of variables is comparable to or larger than the sample size, coefficient estimates produced by ordinary least squares (OLS) can be hindered by massive variance inflation. The use of penalized procedures that introduce a shrinkage in the OLS estimator is a widely accepted solution to this problem. In this paper we consider the most commonly used penalized regressions (PRs); namely, those based on the ℓ_1-penalty (Tibshirani, 1996; Zou, 2006), the ℓ_2-penalty (Hoerl and Kennard, 1970) and their combinations (Zou and Hastie, 2005). Depending on the form of the penalty, PRs can produce *dense solutions*, where coefficients may have small yet non-zero estimates, or *sparse solutions*, where less relevant predictors have coefficient estimates equal to zero.

From a theoretical standpoint, we utilize the work of Zhao and Yu (2006); Bickel et al. (2009); Lounici et al. (2009); Negahban et al. (2009); Zou and Zhang (2009); Negahban et al. (2012); Hastie (2015); Xin et al. (2017), extending some of their results to the time series setting. Specifically, studying the estimation properties of PRs, these authors showed that their non-asymptotic estimation error bound depends critically on the degree of cross-correlation between covariates. In summary, PRs are most effective when covariates are orthogonal or weakly cross-correlated, since the bound is inversely proportional to the minimum eigenvalue of the sample cross-correlation matrix of the covariates themselves. In this respect, two different situations may occur. In the first, the covariates are truly multicollinear; that is, cross-correlations exist at the population level. In the second, correlations are spurious; the covariates can be orthogonal or nearly orthogonal at the population level, but other mechanisms generate cross-correlations in the sample. This matter becomes more prominent and consequential in high dimension, leading to false scientific discoveries and wrong statistical inferences (Fan et al., 2014; Fan and Zhou, 2016; Fan et al., 2018).

While issues related to multicollinear time series have been extensively studied by Forni et al. (2000); Mario and Lippi (2001); Stock and Watson (2002a); De Mol et al. (2008); Medeiros and F.Mendes (2012); Fan et al. (2020), the effects of spurious correlations on PRs in the context of time series data has not been fully addressed to date.

The main objective of this paper is to shed light on the role that covariates' serial dependence can play for PRs through spurious cross-correlations. We show that, in addition to autocorrelation in the residuals (Bartlett, 1935; Granger and Newbold, 1974; Granger et al., 2001), also the serial dependence in the covariates is critical for the estimation of regression coefficients. Our work introduces two elements of novelty. First, we formalize the impact of covariates' serial dependence on the non-asymptotic estimation error bound of PRs, showing that orthogonal or weakly cross-correlated stationary AR processes can exhibit high spurious correlations caused by serial dependence. Specifically, we demonstrate that the probability of spurious correlation between stationary processes depends, in addition to the sample size, on their degree of serial dependence. To prove this we derive the density of the sample cross-correlation between orthogonal stationary Gaussian AR(1) processes. Notably, the result can be generalized to high-dimensional correlation matrices due to the fact that the minimum eigenvalue of a sample correlation matrix is bounded from above by the maximum absolute value of its off diagonal entries. Second, to improve the estimation performance of PRs in a time series context, we introduce a new procedure based on applying PRs to the pre-whitened variables. We show that, under mild assumptions, our proposal produces more accurate estimates of regression coefficients, as well as improved selection of relevant covariates in sparse regimes. To validate our procedure we conduct an extensive simulation study. Furthermore, to illustrate the validity of our results in a high-dimension context, we report an empirical exercise on forecasting the consumer price index by means of a set of 309 macroeconomic time series relative to the Euro Area.

Our main results can be summarized as follows. (*i*) Through our theoretical analysis we show that, whenever the autocorrelation coefficients of AR(1) processes have the same sign, an increase in the degree of serial dependence induces an increase in the probability of large spurious cross-correlations. (*ii*) Through simulations we show that the association between serial dependence and the probability of large spurious correlations holds much more generally, e.g., in cases where the processes are non-Gaussian, weakly cross-correlated or generated by models other than than AR(1). This highlights that a small minimum eigenvalue is more likely in finite realizations of serially dependent weakly cross-correlated (or orthogonal) processes, compared to the case of independent samples – and thus that serially dependent covariates can cause major problems for the estimation accuracy of PRs – something that is numerically corroborated when we apply our proposed procedure to simulated data. (*iii*) Our empirical exercise shows that LASSO (Tibshirani, 1996) applied to ARMA residuals generates more parsimonious models more accurate forecasts compared to LASSO applied directly to the time series under consideration.

The reminder of the paper is organized as follows. In Section 2 we describe the problem setup and our contribution. In Section 3 we present our theoretical result on the impact of covariates serial dependence on sample cross-correlation. In Section 4 we provide simulation studies to corroborate and extend the theoretical results of Section 3. In Section 5 we introduce and evaluate our proposal for mitigating the adverse effects of serial dependence using simulated and real data. In Section 6 we provide some final remarks.

The following notations will be used throughout the paper. For any dimension p, bold letters denote vectors and the corresponding regular letters their elements, for example $\boldsymbol{a} = (a_1, a_2, \ldots, a_p)'$. Supp($\boldsymbol{a}$) denotes the support of a vector, that is, $\{j \in \{1, 2, \ldots, p\} : a_j \neq 0\}$, and $|\text{Supp}(\boldsymbol{a})|$ the support cardinality. The ℓ_q norm of a vector is $||\boldsymbol{a}||_q \coloneqq \left(\sum_{j=1}^p |a_j|^q\right)^{1/q}$ for $0 < q < \infty$, with $||\boldsymbol{a}||_q^k \coloneqq \left(\sum_{j=1}^p |a_j|^q\right)^{k/q}$, and with the usual extension $||\boldsymbol{a}||_0 \coloneqq |\text{Supp}(\boldsymbol{a})|$. Bold capital letters denote matrices, for example $\boldsymbol{A}$, where $(\boldsymbol{A})_{ij} = a_{ij}$ denotes its i-row j-column element. Moreover, $\mathbf{0}_p$ denotes a p-length vector of zeros, while $\boldsymbol{I}_p$ denotes a $p \times p$ identity matrix. Finally, $Sign(r)$ indicates the sign of a real number r.

2 Problem Setup and Our Contribution

2.1 State-of-the-Art

Let $\mathbf{X} = \{\mathbf{x}_t\}_{t=1}^T$ denote an $n \times T$ rectangular array of observations on n covariates, and $\mathbf{y} = \{y_t\}_{t=1}^T$ a $1 \times T$ response vector. Assume that $\mathbf{y}$ and $\mathbf{X}$ are realizations of strictly Gaussian stationary and absolutely regular processes $\{(y_t, \mathbf{x}_t') \in \mathbb{R}^{1+n}, t \in \mathbb{Z}, n \in \mathbb{N}\}$ defined on the probability space $(\Omega, \mathcal{F}, P)$. Let $\mathbf{C}_x = E[\mathbf{x}_t \mathbf{x}_t']$ and $\widehat{\mathbf{C}}_x = \frac{1}{T-1}\mathbf{X}\mathbf{X}'$ denote the cross-covariance matrix and its estimate, with generic element $\widehat{c}_{ij}^x$ and eigenvalues $\widehat{\psi}_{max}^x \geq \ldots \geq \widehat{\psi}_{min}^x$. Finally, assume that each $\mathbf{x}_i$ is standardized so that $\frac{1}{T-1}\sum_t x_{it} = 0$ and $\frac{1}{T-1}\sum_t x_{it}^2 = 1$. We consider the following data generating process (DGP) for the response

$$\mathbf{y} = \mathbf{X}'\boldsymbol{\alpha}^* + \boldsymbol{\varepsilon},$$

where $\boldsymbol{\alpha}^*$ is the $n \times 1$ unknown s-sparse vector of regression coefficients, i.e. $||\boldsymbol{\alpha}^*||_0 = s < n$, and $\boldsymbol{\varepsilon} \in \mathbb{R}^T$ is a random noise vector. If $n \geq T$, $\boldsymbol{\alpha}^*$ is estimated solving a convex optimization that combines a quadratic loss and a regularization penalty:

$$\widehat{\boldsymbol{\alpha}} = \underset{\boldsymbol{\alpha}\in\mathbb{R}^n}{\operatorname{argmin}} \left\{ \frac{1}{2T} \ ||\mathbf{y} - \mathbf{X}'\boldsymbol{\alpha}||_2^2 + \lambda\ell(\boldsymbol{\alpha}) \right\}. \tag{1}$$

Here $\lambda > 0$ represents the weight of the penalty, and $\ell : \mathbb{R}^n \rightarrow \mathbb{R}^+$ is a norm. The following definitions will be used.

Definition 1: *(Strong Convexity): Given a differentiable function $\mathcal{L} : \mathbb{R}^n \rightarrow \mathbb{R}$ and the vector differential operator $\triangledown$, we say that $\mathcal{L}$ is strongly convex with parameter $\gamma > 0$ at $\boldsymbol{a} \in \mathbb{R}^n$ if, for all $\boldsymbol{b} \in \mathbb{R}^n$, $\mathcal{L}(\boldsymbol{b}) - \mathcal{L}(\boldsymbol{a}) \geq \triangledown\mathcal{L}(\boldsymbol{a})'(\boldsymbol{b} - \boldsymbol{a}) + \frac{\gamma}{2}||\boldsymbol{b} - \boldsymbol{a}||_2^2$.*

Strong Convexity guarantees a small coefficient estimation error (see Negahban et al. 2009, 2012). In particular, when $\mathcal{L}$ (the loss function) is "sharply curved" around its optimum $\widehat{\boldsymbol{\alpha}}$, a small $|\mathcal{L}(\widehat{\boldsymbol{\alpha}}) - \mathcal{L}(\boldsymbol{\alpha}^*)|$ guarantees that $||\widehat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}^*||_2$ is also small. The parameter γ governs the strength of convexity; when $\mathcal{L}$ is twice differentiable, strong convexity requires the minimum eigenvalue of the Hessian $\triangledown^2\mathcal{L}(\boldsymbol{\alpha})$ to be at least γ for all $\boldsymbol{\alpha}$ in a neighborhood of $\boldsymbol{\alpha}^*$. Thus, since its Hessian is $\triangledown^2\mathcal{L}(\boldsymbol{\alpha}) = \widehat{\mathbf{C}}_x$ for all $\boldsymbol{\alpha} \in \mathbb{R}^n$, the quadratic loss $\mathcal{L}(\boldsymbol{\alpha}) = \frac{1}{2T}||\mathbf{y} - \mathbf{X}'\boldsymbol{\alpha}||_2^2$ is strongly convex with parameter γ if and only if $\widehat{\psi}_{min} \geq \gamma$ (see Hastie 2015, p. 293). Consequently, in this case $||\widehat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}^*||_2$ depends on $\widehat{\psi}_{min}$. It is important to note that when $n > T$ the quadratic loss cannot be strongly convex since $\widehat{\mathbf{C}}_x$ is singular and thus $\widehat{\psi}_{min} = 0$. In this case Bickel et al. (2009) proposed a *Restricted Eigenvalue Condition*, which is essentially a restriction on the eigenvalues of $\widehat{\mathbf{C}}_x$ as a function of the degree of sparsity, s. The Restricted Eigenvalue Condition allows for strong convexity (Definition 1) to hold in the singular case, and we refer to this as *Restricted Strong Convexity* (see Negahban et al. 2012; we provide more details in Supplement A).

Definition 2: *(Dual Norm and Subspace Compatibility Constant): Given a norm ℓ and* *the inner product* $\langle \cdot, \cdot \rangle$*, we define the dual norm of* ℓ *as* $\ell^*(\boldsymbol{v}) := \sup_{\mathbf{u} \in \mathbb{R}^n \setminus \{0\}} \frac{\langle \mathbf{u}, \mathbf{v} \rangle}{\ell(\mathbf{u})}$*. For any subspace* $\mathcal{A}$ *of* $\mathbb{R}^n$ *that captures the constraints underlying* (1)*, we define the subspace compatibility constant with respect to the pair* $(\ell, ||\cdot||_2)$ *as* $\Psi(\mathcal{A}) := \sup_{\mathbf{u} \in \mathcal{A}: \mathbf{u} \neq 0} \frac{\ell(\mathbf{u})}{||\mathbf{u}||_2}$*.*

The following Proposition is derived from Corollary 1 of Negahban et al. (2012) and provides the non-asymptotic coefficient estimation error bound for PRs.

Proposition 1: *Consider the convex optimization problem in* (1)*. Suppose that the penalty parameter* λ *is strictly positive and* $\geq 2\ell^*\left(\frac{1}{T}\mathbf{X}\boldsymbol{\varepsilon}\right)$*, and that strong convexity holds with parameter* $\gamma > 0$*. Then, any optimal solution* $\widehat{\boldsymbol{\alpha}}$ *satisfies the bound* $||\widehat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}^*||_2 \leq 3\frac{\lambda}{\gamma}\Psi(\mathcal{A})$*.*

Proof: See Corollary 1 in Negahban et al. (2012). ■

The coefficient estimation error bound in Proposition 1 increases with the penalty parameter λ, which must be strictly positive and $\geq 2\ell^*\left(\frac{1}{T}\mathbf{X}\boldsymbol{\varepsilon}\right)$; increases with the subspace compatibility constant $\Psi(\mathcal{A})$, which in turn increases with the size of the model subspace $\mathcal{A}$; and decreases with the convexity parameter γ. Negahban et al. (2009, 2012) derive the bound for PRs in the case of independent observations (no serial dependence). To this end, the authors compute the probability that $\lambda \geq 2\ell^*\left(\frac{1}{T}\mathbf{X}\boldsymbol{\varepsilon}\right)$ when the entries of $\mathbf{X}$ and $\boldsymbol{\varepsilon}$ are sub-Gaussian, and assume that strong convexity (or restricted strong convexity) holds with parameter γ, i.e. that $\widehat{\psi}^x_{min} \geq \gamma$ (see Corollary 2 in Negahban et al. 2012 and Corollary 6 in Negahban et al. 2009 for examples of sparse and dense PRs, respectively).

This analysis shows the role of covariates cross-correlation in determining the estimation accuracy of PRs. In particular, Proposition 1 shows that PRs perform better if covariates are orthogonal or weakly correlated in the sample, since strong sample cross-correlations correspond to small $\widehat{\psi}^x_{min}$. As mentioned in the Introduction, strong sample cross-correlations may reflect true multicollinearities at the population level. In this case Fan et al. (2020) focus on the fact that time series multicollinearities can be captured with

Factor Models, and propose to apply PRs on the estimated idiosyncratic components obtained by filtering the observed time series through estimated factors. However, strong sample cross-correlation may also be spurious; this is the case we wish to tackle in the particular context of time series.

2.2 Our Contribution

We argue that spurious correlations are one of the causes that potentially limits the use of PRs in time series. In particular, we focus on the implications of covariates serial dependence on $\widehat{\psi}^x_{min}$, which determines the “strength” of strong convexity (see Definition 1) and is one of the main components of the PRs error bound in Proposition 1. In this respect, we relax the assumption that strong convexity (or restricted strong convexity) holds with parameter $\gamma > 0$, showing that the probability of getting $\widehat{\psi}^x_{min} \leq \gamma$ increases with the covariates' serial dependence. Note that in order to focus on $\widehat{\psi}^x_{min}$ we are assuming that $\widehat{\mathbf{C}}_x$ is positive definite. If $n > T$ the matrix $\widehat{\mathbf{C}}_x$ is singular, but our results are still valid considering the probability of a restricted eigenvalue $\leq \gamma$ (see Bickel et al. 2009).

The structure of our theoretical contribution is as follows. Given $\gamma = 1 - \tau$, $\tau \in [0, 1)$, and the upper bound $\widehat{\psi}^x_{min} \leq 1 - \max_{i \neq j} |\widehat{c}^x_{ij}|$, we emphasize the role of a generic off-diagonal element of $\widehat{\mathbf{C}}_x$ in determining the probability that $\widehat{\psi}^x_{min} \leq \gamma$ through the inequalities

$$\Pr\Big\{\widehat{\psi}^x_{min} \leq 1 - \tau\Big\} \geq \Pr\Big\{1 - \max_{i \neq j} |\widehat{c}^x_{ij}| \leq 1 - \tau\Big\} \geq \Pr\big\{1 - |\widehat{c}^x_{i \neq j}| \leq 1 - \tau\big\} = \Pr\big\{|\widehat{c}^x_{i \neq j}| \geq \tau\big\}. \quad (2)$$

Thus, $\Pr\{|\widehat{c}^x_{i \neq j}| \geq \tau\}$ plays a role in determining the probability of dealing with a small $\widehat{\psi}^x_{min}$ and consequently, through strong convexity, on the PRs estimation error bound presented in Proposition 1. It follows that any impact of the degree of serial dependence on $\Pr\{|\widehat{c}^x_{i \neq j}| \geq \tau\}$ results in an impact on such bound.

To better illustrate our reasoning, we introduce a toy example where we show numerically the impact of serial dependence on $\max_{i \neq j} |\widehat{c}_{ij}^x|$ and $\widehat{\psi}_{min}^x$. We generate 10 processes from the model $\mathbf{x}_t = \mathbf{D}_\phi \mathbf{x}_{t-1} + \mathbf{u}_t$, $t = 1, \ldots, 100$, where $\mathbf{D}_\phi$ is a 10×10 diagonal matrix with the same autocorrelation coefficient ϕ in all positions along the main diagonal, and $\mathbf{u}_t \sim N(\mathbf{0}_{10}, \boldsymbol{I}_{10})$. Note that for these AR(1) processes the degree of serial dependence is determined by $|\phi|$ and, since the processes are orthogonal, the minimum eigenvalue of the population cross-correlation matrix $\mathbf{C}_x$ is $\psi_{min}^x = 1$. We consider five values for ϕ, namely $0.0, 0.3, 0.6, 0.9, 0.95$, and for each we calculate the average and the standard deviation of both $\max_{i \neq j} |\widehat{c}_{ij}^x|$ and $\widehat{\psi}_{min}^x$ on 5000 Monte Carlo replications. Results are reported in Figure 1. We observe that the stronger the persistence of the process (ϕ closer to 1) the higher is the probability of a large spurious sample correlation (orange circle), which in turn leads to a small minimum eigenvalue of the sample cross-correlation matrix (blue triangle).

In light of these results, our next task is to derive the finite sample density of $\widehat{c}_{ij}^x$ for the purpose of formalizing the impact of serial dependence on $\Pr\{|\widehat{c}_{i \neq j}^x| \geq \tau\}$. It is noteworthy that, when the covariates have a factor-based structure, strong spurious correlations in the sample may affect the idiosyncratic components if they are serially dependent, reducing the accuracy of the procedure proposed by Fan et al. (2020) (see Supplement B) .

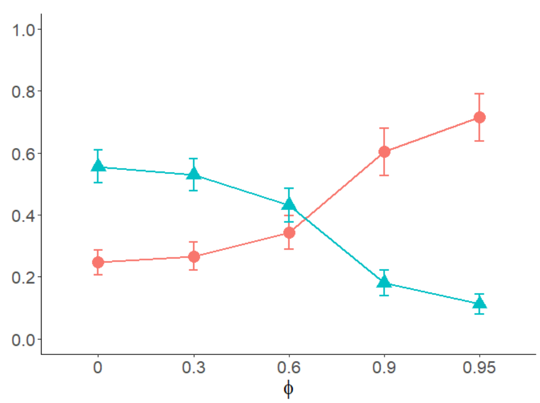


Figure 1: Results from a numerical toy example. The orange circles and bars represent means and standard deviations of $\max_{i \neq j} |\widehat{c}_{ij}^x|$ for various values of the autocorrelation ϕ, as obtained from 5000 Monte Carlo replications. Similarly, blue triangles and bars represent means and standard deviations of $\widehat{\psi}_{min}^x$.

3 Density of the Sample Correlation between two Orthogonal AR(1) Gaussian Processes

In this section we present our main theoretical contribution concerning the impact of covariates' serial dependence on the non-asymptotic estimation error bound of PRs, showing that the probability of incurring in strong spurious correlation increases with serial dependence.

Consider a first order bivariate autoregressive process $\mathbf{x}_t = \mathbf{D}_\phi \mathbf{x}_{t-1} + \mathbf{u}_t$, $t = 1, \dots, T$, where $\mathbf{D}_\phi$ is a 2×2 diagonal matrix with main diagonal elements $\phi_1, \phi_2 < 1$. We make the following assumption about the bivariate vector of autoregressive residuals:

Assumption 1: $\mathbf{u}_t \sim N(\mathbf{0}_2, \boldsymbol{I}_2)$.

Therefore $\boldsymbol{x}_t \sim N(\mathbf{0}_2, \mathbf{C}_x)$ with $(\mathbf{C}_x)_{ii} = \frac{1}{1-\phi_i^2}$, $i = 1, 2$, and $(\mathbf{C}_x)_{12} = c_{12}^x = 0$. In this setting we focus on the density of the sample correlation coefficient defined as

$$\widehat{c}_{12}^x = \frac{a_{12}}{\sqrt{a_{11}}\sqrt{a_{22}}}, \tag{3}$$

where $a_{i,j} = \sum_{t=1}^T (x_{it} - \overline{x}_i)(x_{jt} - \overline{x}_j) = \sum_{t=1}^T x_{it} x_{jt}$, since $\overline{x}_i = 0$, $i, j = 1, 2$. In particular, when $c_{12}^u = 0$, $b = a_{21}/a_{11}$ and $v = a_{22} - a_{21}^2/a_{11}$, then

$$\frac{\sqrt{a_{11}}\ b}{\sqrt{v/(T-2)}} = \sqrt{T-2}\frac{a_{12}/\sqrt{a_{11}a_{22}}}{\sqrt{1 - a_{12}^2/(a_{11}a_{22})}} = \sqrt{T-2}\frac{\widehat{c}_{12}^x}{\sqrt{1 - (\widehat{c}_{12}^x)^2}}. \tag{4}$$

Note that b is the least squares regression coefficient of x_{2t} on x_{1t}, and v is the sum of the square of residuals of such regression. Thus, to derive the finite sample density of $\widehat{c}_{12}^x$ we need the sample densities of b and v.

Remark 1: *In contrast to asymptotic statements, our theoretical analysis is intended to derive distributions and densities of estimators that hold for $T < \infty$. Hence we will not* *employ the usual concepts of convergence in probability and in distribution; rather we will use a notion of approximation, whose "precision" needs to be evaluated. The precision of our approximations will be extensively tested under several finite T scenarios in both the simulation study provided in Section 4 and in the Supplement.*

Sample Distribution of b. We start by deriving the sample distribution of b, the OLS regression coefficient for x_2 on x_1. The same holds if we regress x_1 on x_2.

Proposition 2: *Under Assumption 1 the sample distribution of b is approximately* $N\left(0, \frac{(1-\phi_1^2\phi_2^2)(1-\phi_1^2)}{(T-1)(1-\phi_2^2)(1-\phi_1\phi_2)^2}\right)$.

Proof: See Supplement C.1

Proposition 2 shows that the OLS estimate b is normally distributed with a variance that strongly depends on the degree of covariates serial dependence. In this context, it is common to adjust the standard error of the OLS to achieve consistency in the presence of heteroskedasticity and/or serial dependence; this leads, for instance, to the Heteroskedasticity and Autocorrelation Consistent (HAC) estimator of Newey and West (1987) (NW). However, NW estimates can be highly sub-optimal (or inefficient) in the presence of strong serial dependence (Baillie et al., 2022). In Supplement D we provide a simulation study to corroborate the result in Proposition 2.

Sample Distribution of v. Here we derive the sample distribution of the sum of the square of residuals obtained by regressing x_2 on x_1.

Proposition 3: *Under Assumption 1 the sample distribution of v is approximately* $\Gamma\left(\frac{T-2}{2}, \frac{2}{1-\phi_2^2}\right)$.

Proof: See Supplement C.2

Sample Density of $\widehat{c}^x_{12}$. Note that b and v are independent. Using Propositions 2 and 3 and Equation (4) we can now derive the density of the sample distribution of $\widehat{c}^x_{12}$.

Theorem 1: *Let $\mathbf{x}_t$ be a stationary bivariate Gaussian AR(1) process with autoregressive residuals distributed according to $N(\mathbf{0}_2, \boldsymbol{I}_2)$. Further, let $\phi_{12} = \phi_1\phi_2$ where ϕ_i, $i = 1, 2$, are the autoregressive coefficients. Then, the sample density of $\widehat{c}^x_{12}$ is approximated by*

$$\mathcal{D}_{\widehat{c}^x_{12}} = \frac{\Gamma\left(\frac{T-1}{2}\right)(1-\phi_{12})}{\Gamma\left(\frac{T-2}{2}\right)\sqrt{\pi}}(1-(\widehat{c}^x_{12})^2)^{\frac{T-4}{2}}\left(1-\phi_{12}^2\right)^{\frac{T-2}{2}}\left(\frac{1}{1-\phi_{12}^2+2(\widehat{c}^x_{12})^2\phi_{12}(\phi_{12}-1)}\right)^{\frac{T-1}{2}}. \quad (5)$$

Proof: See Supplement C.3

Remark 2: *$\mathcal{D}_{\widehat{c}^x_{12}}$ is the density of the sample correlation coefficient* (3) *based on a finite T, with serial dependence expressed by ϕ_1 and ϕ_2, and under the assumption of orthogonal Gaussian AR(1) processes.*

Remark 3: *From* (5) *we see that ϕ_{12} determines the density of $\widehat{c}^x_{12}$ trough both its magnitude and sign. More precisely, when $Sign(\phi_1) = Sign(\phi_2)$, the probability in the tails increases as $|\phi_{12}|$ grows. On the other hand, when $Sign(\phi_1) \neq Sign(\phi_2)$, an increase in $|\phi_{12}|$ leads to a density more concentrated around the origin. This peculiarity on the effect of $Sign(\phi_{12})$ will be numerically explored and validated in Section 4.*

Theorem 1 shows that, in a finite T context, the probability of observing sizeable spurious correlation between orthogonal Gaussian Autoregressive processes crucially depends on the degree of serial dependence. This has important consequences on the non-asymptotic performance of PRs for the reasons that we pointed out at the beginning of Section 2.2, related to the role of $\Pr\{|\widehat{c}^x_{12}| \geq \tau\}$ (see inequality (2), Definition 1 and Proposition 1). The implication of Theorem 1 for such probability can be summarized in the following remark.

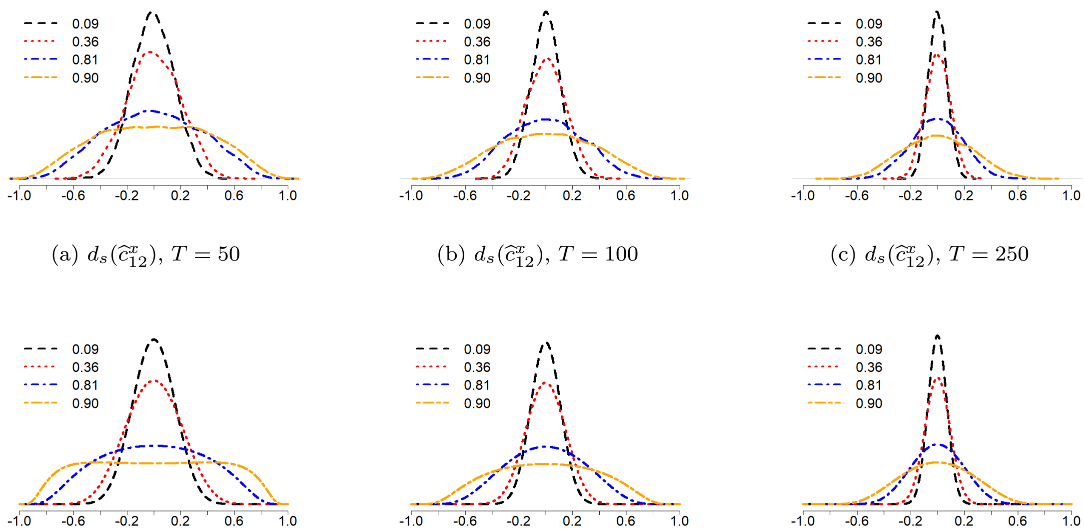
Remark 4: *Because of Theorem 1* $\Pr\{|\widehat{c}^x_{12}| \geq \tau\} \approx \int_{-1}^{-\tau} \mathcal{D}_{\widehat{c}^x_{12}} d\widehat{c}^x_{12} + \int_{\tau}^{1} \mathcal{D}_{\widehat{c}^x_{12}} d\widehat{c}^x_{12}$, *which depends on the degrees of serial dependence of the processes.*

4 Monte Carlo Experiments

In this Section we conduct Monte Carlo experiments to assess numerically the approximation of the density of $\widehat{c}_{12}^{x}$ described in Section 3. Then, we expand the theoretical results in more generic contexts, relaxing the assumption that the covariates are orthogonal Gaussian AR(1) processes. We indicate the density of $\widehat{c}_{12}^{x}$ obtained by simulations as $d_s(\widehat{c}_{12}^{x})$.

4.1 Numerical Approximation of $d_s(\widehat{c}_{12}^{x})$ to $\mathcal{D}_{\widehat{c}_{12}^{x}}$

We generate data from the bivariate process $\mathbf{x}_t = \mathbf{D}_\phi \mathbf{x}_{t-1} + \mathbf{u}_t$ for $t = 1, \ldots, T$, where $\mathbf{D}_\phi$ is a 2×2 diagonal matrix with the same autocorrelation coefficient ϕ in both positions along the diagonal, and $u_t \sim N(\mathbf{0}_2, \boldsymbol{I}_2)$. We consider $T = 50, 100, 250$ and $\phi = 0.3, 0.6, 0.9, 0.95$ – thus, the parameter ϕ_{12} in $\mathcal{D}_{\widehat{c}_{12}^{x}}$, here equal to ϕ^2, takes values 0.09, 0.36, 0.81, 0.90. The first row of Figure 2 (Plots (a), (b), (c)) shows, for various values of T and ϕ_{12}, the density $d_s(\widehat{c}_{12}^{x})$ generated through 5000 Monte Carlo replications. The second row of Figure 2 (Plots (d), (e), (f)) shows the corresponding $\mathcal{D}_{\widehat{c}_{12}^{x}}$. These were plotted using 5000 values of the argument starting at -1 and increasing by steps of size 0.0004 until 1. As expected, we observe that the degree of approximation of $d_s(\widehat{c}_{12}^{x})$ to $\mathcal{D}_{\widehat{c}_{12}^{x}}$ improves as T increases and/or ϕ_{12} decreases. In particular, Plots (a) and (d) in Figure 2, where $T = 50$, show that $\mathcal{D}_{\widehat{c}_{12}^{x}}$ approximates $d_s(\widehat{c}_{12}^{x})$ well for a low-to-intermediate degree of serial dependence ($\phi_{12} \leq 0.36$, i.e. $\phi \leq 0.6$). In contrast, in cases with high degree of serial dependence ($\phi_{12} \geq 0.81$, i.e. $\phi \geq 0.9$), $\mathcal{D}_{\widehat{c}_{12}^{x}}$ has larger tails compared to $d_s(\widehat{c}_{12}^{x})$; that is, the latter over-estimates the probability of large spurious correlations. However, it is noteworthy that the difference between the two densities is negligible for $T \geq 100$ (Figure 2, Plots (b) and (e) for $T = 100$, and Plots (c) and (f) for $T = 250$), also with high degree of serial dependence ($\phi_{12} \approx 0.90$, i.e. $\phi = 0.95$). These numerical experiments corroborate the fact that the sample cross-correlation between orthogonal Gaussian AR(1) processes is affected by the degree of serial dependence in a way that is well approximated by $\mathcal{D}_{\widehat{c}^x_{12}}$. In fact, for a sufficiently large finite T, we observe that $\Pr\{|\widehat{c}^x_{12}| \geq \tau\}$, $\tau > 0$, increases with ϕ_{12} in a similar way for $d_s(\widehat{c}^x_{12})$ and $\mathcal{D}_{\widehat{c}^x_{12}}$.



(a) $d_s(\widehat{c}^x_{12})$, $T = 50$ (b) $d_s(\widehat{c}^x_{12})$, $T = 100$ (c) $d_s(\widehat{c}^x_{12})$, $T = 250$

(d) $\mathcal{D}_{\widehat{c}^x_{12}}$, $T = 50$ (e) $\mathcal{D}_{\widehat{c}^x_{12}}$, $T = 100$ (f) $\mathcal{D}_{\widehat{c}^x_{12}}$, $T = 250$

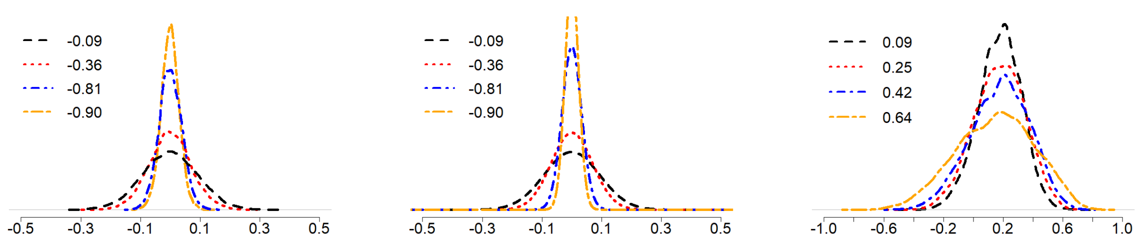
Figure 2: Monte Carlo densities for $\widehat{c}^x_{12}$ (top) and corresponding $\mathcal{D}_{\widehat{c}^x_{12}}$ (bottom) for various T and ϕ_{12}.

The Impact of $Sign(\phi_{12})$

In Remark 3 we pointed out that the impact of ϕ_{12} on $\mathcal{D}_{\widehat{c}^x_{12}}$ depends on $Sign(\phi_{12})$. In particular, when $-1 < \phi_{12} < 0$, an increment on $|\phi_{12}|$ makes the density of $\widehat{c}^x_{12}$ more concentrated around 0. In order to validate this result, we run simulations with $T = 100$ and different values for the second element of the diagonal of $\mathbf{D}_\phi$; namely, $-0.3, -0.6, -0.9, -0.95$. Results are shown in Plots (a) and (b) of Figure 3. In this case, we see that when $Sign(\phi_1) \neq Sign(\phi_2)$ and $|\phi_{12}|$ increases, $d_s(\widehat{c}^x_{12})$ increases its concentration around 0 in a way that is, again, well approximated by $\mathcal{D}_{\widehat{c}^x_{12}}$.

4.2 General Case

To generalize our findings to the case of non-Gaussian weakly correlated AR and ARMA processes, we generate covariates according to the following DGPs: $x_{1t} = (\phi + 0.1)x_{1t-1} + (\phi+0.1)x_{1t-2} - 0.2x_{1t-3} + u_{1t}$, and $x_{2t} = \phi x_{2t-1} + \phi x_{2t-2} + u_{2t} + 0.8u_{2t-1}$, where $t = 1, \ldots, 100$ and $\phi = 0.15, 0.3, 0.45, 0.475$. Moreover, we generate u_{1t} and u_{2t} from a bivariate Laplace distribution with means 0, variances 1, and $c^u_{12} = 0.2$. In these more general cases, we do not know an approximate theoretical density for $\widehat{c}^u_{12}$. Therefore, we rely entirely on simulations to show the effect of serial dependence on $\Pr\{|\widehat{c}^x_{12}| \geq \tau\}$. Plot (c) of Figure 3 shows $d_s(\widehat{c}^x_{12})$ obtained from 5000 Monte Carlo replications for the different values of ϕ. In short, also in the more general cases where covariates are non-Gaussian, weakly correlated AR(3) and ARMA(2,1) processes, the probability of getting large sample cross-correlations depends on the degree of serial dependence. More simulation results are provided in Supplement E.



(a) $d_s(\widehat{c}^x_{12})$, $Sign(\phi_1) \neq Sign(\phi_2)$ (b) $\mathcal{D}_{\widehat{c}^x_{12}}$, $Sign(\phi_1) \neq Sign(\phi_2)$ (c) $d_s(\widehat{c}^x_{12})$, AR and ARMA

Figure 3: Monte Carlo densities for $\widehat{c}^x_{12}$ (a) and corresponding $\mathcal{D}_{\widehat{c}^x_{12}}$ (b), for T=100 and various (negative) ϕ_{12}'s. Monte Carlo densities for $\widehat{c}^x_{12}$ in the case of Laplace weakly correlated AR(3) and ARMA(2,1) processes, for $T = 100$ and various ϕ's (c).

5 A Remedy for Serial Dependence-Induced Spurious Correlation

In this section we propose an approach to mitigate the issues caused by serial dependence-induced spurious correlations for the performance of PRs. Our proposal consists of a two-step procedure. In the first step, we estimate a univariate model on each covariate time series (for example, an ARMA model); in the second step, we run PRs using the residuals of the models fitted in the first step instead of the original covariates. In more detail, let x_{it} (the i-th time series at time t) be generated by the model

$$x_{it} = \sum_{l=1}^{p_i} \phi_{il} x_{it-l} + \sum_{k=1}^{q_i} \theta_{ik} u_{it-k} + u_{it} \tag{6}$$

where $i = 1, \ldots, n, t = 1 \ldots, T$. This describes an ARMA(p_i,q_i) process where p_i is the order of autocorrelation, which determines the order of the weighted moving average over past values of the covariate, and q_i is the order of the weighted moving average over past errors. Note that the AR (i.e. $p_i \geq 1$, $q_i = 0$) and MA (i.e. $p_i = 0$, $q_i \geq 1$) models are special cases of (6). For notational simplicity let $x_{it|t-1} = \sum_{l=1}^{p_i} \phi_{il} x_{it-l} + \sum_{k=1}^{q_i} u_{it-k}$ and let $\widehat{x}_{it|t-1}$ be an estimate of $x_{it|t-1}$. We propose to run PRs using the estimated residuals $\widehat{u}_{it} = x_{it} - \widehat{x}_{it|t-1}$.

5.1 The Working Model on ARMA residuals

Assume that response variable and covariates are generated by the following DGPs:

$$y_t = \sum_{i=1}^{n} \alpha_i^* x_{it-1} + \varepsilon_t \ , \tag{7}$$

$$x_{it} = \phi_i x_{it-1} + u_{it} \ , \tag{8}$$

$$\varepsilon_t = \phi_\varepsilon \varepsilon_{t-1} + \omega_t \ , \tag{9}$$

where $i = 1, \ldots, n$, $t = 1, \ldots, T$, $|\phi_i| < 1$ $|\phi_\varepsilon| < 1$, and u_{it} and ω_t are the *i.i.d.* random errors of the processes. Consider a regression model where one lag of y_t is included in the set of potential predictors

$$y_t = \sum_{i=1}^{n} \alpha_i^* x_{it-1} + \phi_y y_{t-1} + \omega_t \ . \tag{10}$$

This strategy is usually adopted to eliminate any residual serial correlation (Keele and Kelly, 2006). The following two assumptions are crucial for our proposal.

Assumption 2: $u_{it} \perp u_{jt-l}$ *for any* i*,* j*,* t *and* $l \neq 0$*.*

Assumption 3: $u_{it-l} \perp \omega_t$ *for any* i*,* t *and* l*.*

Moreover, we assume that n is comparable to or larger than T and, temporarily and for the sake of the argument, that the u_{it-1}'s are observable – so that we do not need to estimate them through $x_{it} - \widehat{x}_{it|t-1}$. Our proposal utilizes the *working model*

$$y_t = \sum_{i=1}^{n} \alpha_i^* u_{it-1} + \phi_y y_{t-1} + \omega_t \tag{11}$$

in place of model (10). In the following, we refer to PRs when estimating the coefficients of (10) and to u-PRs when estimating the coefficients of (11). Taking a step back, we consider the asymptotic behavior of the unpenalized OLS least squares estimates for the two models. The reason for this digression is that, in the presence of serial dependence in the covariates and in the error, OLS estimates for (10) are asymptotically biased (Achen, 2000; Keele and Kelly, 2006). We show that those for (11) are not. The following Proposition contrasts the asymptotic behavior of the two OLS estimates in terms of convergence in probability.

Proposition 4: *Let the data be generated by* (7)*,* (8) *and* (9)*, with* $\phi_i = \phi$*,* $i = 1, \ldots, n$*. Moreover, let* σ_y^2 *be the variance of* y *and* $R^2 = \boldsymbol{\alpha}^{*\prime}\mathbf{C}_x\boldsymbol{\alpha}^*/\sigma_y^2$ *be the asymptotic coefficient of determination, where* $\mathbf{XX}'/T \overset{p}{\to} \mathbf{C}_x$*. Under Assumptions 2 and 3, as* $T \to \infty$*, OLS* *estimates for model* (10) *converge as follows*

$$\widehat{\phi}_y \xrightarrow{p} \frac{\phi_\varepsilon(1-R^2)}{1-\phi^2R^2} \quad , \quad \widehat{\boldsymbol{\alpha}} \xrightarrow{p} \boldsymbol{\alpha}^* \left(1 - \frac{\phi\phi_\varepsilon(1-R^2)}{1-\phi^2R^2}\right)$$

while OLS estimates for model (11) *converge as follows*

$$\widehat{\phi}_y \xrightarrow{p} \phi R^2 + \phi_\varepsilon(1-R^2) \quad , \quad \widehat{\boldsymbol{\alpha}} \xrightarrow{p} \boldsymbol{\alpha}^*.$$

Proof: See Supplement C.4

Proposition 4 shows that applying OLS to the covariates induces an asymptotic bias in the estimation of $\boldsymbol{\alpha}^*$ which does not occur when applying OLS to the residuals. The same problem will be inherited by penalized versions of the least squares. Thus, we can articulate the impact of serial dependence on the coefficient estimation error of model (10) in relation to T. When T is fixed, serial dependence leads both to biased coefficients for the relevant variables and to the proliferation of false positives. The latter may be the consequence of serial dependence-induced spurious correlations between relevant and irrelevant variables (Fan and Lv, 2008; Fan et al., 2014). When T grows and the issues caused by serial dependence-induced spurious correlations fade, we still have a negative effect of serial dependence on the estimated coefficients, as shown in Proposition 4 (in Supplement F we compare OLS applied on ARMA residuals with some of the best-known OLS estimators used to address serial dependence in regression). Note that, regardless of whether $T << \infty$ or $T \to \infty$, the estimation error of $\widehat{u}$-PRs is smaller than that of PRs.

We point out that the working model (10) never corresponds to the "true model" (7) (i.e. the data generating process for the response), and its estimates are downwardly biased as both ϕ and ϕ_ε increase. Thus, we proceed by comparing (7) and (11). The latter allows us to estimate the true vector $\boldsymbol{\alpha}^*$ through the u_{it-1}'s, regardless of the possible issues in estimating the serial dependence of y_t. This is possible because $u_{it} \perp x_{it|t-1}$ for any specification of $x_{it|t-1}$, which is a consequence of Assumptions 2 and 3. For this reason, omitting autoregressive and/or moving average component(s) from (11) does not induce an omitted-variables bias for the estimated coefficients. However, it does reduce the explained variance of y_t – which is mitigated by including its lags among the regressors of the working model. How well the "true model" (7) and the working model (11) match depends on the DGPs of the covariates and the noise. We show this in Examples 1 and 2.

Example 1: *(Equal degrees of serial dependence). Suppose $\phi_i = \phi_\varepsilon = \phi$, $i = 1, \ldots, n$. Then, model* (7) *can be rewritten as*

$$y_t = \sum_{i=1}^{n} \alpha_i^* (\phi x_{it-2} + u_{it-1}) + \phi \varepsilon_{t-1} + \omega_{it} = \sum_{i=1}^{n} \alpha_i^* u_{1t-1} + \phi y_{t-1} + \omega_t \ .$$

Thus, in an "ideal regime" in terms of degree of serial dependence (also known as "common factor restriction"), the working model (11) *is equivalent to the true model* (7) *because of the decomposition of the AR(1) processes x_{1t-1}, x_{2t-1} and ε_t. Note that by Proposition 4, if the common factor restriction holds, $\widehat{\phi}_y \xrightarrow{p} \phi$.*

Example 2: *(Different degrees of serial dependence). Suppose $\phi_1 \neq \ldots \neq \phi_n \neq \phi_\varepsilon$. Then, with some simple steps, model* (7) *can be rewritten as*

$$y_t = \sum_{i=1}^{n} \alpha_i^* (\phi_i x_{it-2} + u_{it-1}) + \phi_\varepsilon \varepsilon_{t-1} + \omega_{it} = \sum_{i=1}^{n} \alpha_i^* u_{it-1} + \sum_{i=1}^{n} \alpha_i^* (\phi_i x_{it-2}) + \phi_\varepsilon \varepsilon_{t-1} + \omega_t \ .$$

Thus, in this perhaps more realistic regime, the working model (11) *is* not *equivalent to the true model* (7) *since the predictors and the error do not have the same degree of* *serial dependence, and therefore the use of y_{t-1} does not allow us to summarize the serial dependence of y_t.*

Two more examples (equal degrees of serial dependence and different models, either for the predictors or for the error) are provided in Supplement H. We note that, even when true and working models do not match, as in Example 2, u-PRs moves us from estimating coefficients in a context characterized by high spurious correlation, to one characterized by very weak (or absent) spurious correlations. Of course the parameters we estimate about the past of y_t change, but we can still formulate an effective estimation strategy, e.g., if $\varepsilon_t = \sum_{j=1}^{p_\varepsilon} \phi_{\varepsilon j} \varepsilon_{t-j} + \omega_t$, even in cases such as Example 2, the variability due to misspecification of the serial dependence of y_t is less than that introduced by estimating the model directly on the $\mathbf{x}_i$'s, $i = 1, \ldots, n$.

Since the error term is not correlated with the regressors included in the model, its serial dependence does not violate the assumption of exogeneity and the OLS estimator remains unbiased and consistent. To prevent serial dependence of the error term, a possible solution could be to increase the number of lags of y_t considered in the working model.

Of course, in practice, the $\boldsymbol{u}_i$'s, $i = 1, \ldots, n$, are not observable and need to be replaced by estimated residuals of ARMA, AR or MA processes. When fitting the working models with such residuals we refer to our proposal as $\widehat{u}$-PRs. In the following Sections we demonstrate the estimation and forecasting performance of $\widehat{u}$-LASSO (LASSO applied on ARMA residuals) through both simulations and an empirical application.

5.2 $\widehat{u}$-LASSO

5.2.1 Coefficient Estimation Error Bound

Here, we present Monte Carlo experiments to assess the effectiveness of $\widehat{u}$-LASSO in reducing the coefficient estimation error. We generate the response as $y_t = \sum_{i=1}^{n} \alpha_i^* x_{it-1} + \varepsilon_t$, where $\varepsilon_t = \phi\varepsilon_{t-1} + \omega_t$, and $\omega_t \sim N(0, \sigma_\omega^2)$. The coefficient vector $\boldsymbol{\alpha}^* = (\alpha_1^*, \ldots, \alpha_n^*)'$ is sparse with $||\boldsymbol{\alpha}^*||_0 = 10$. The active covariates are the first 10, followed by $n - 10$ inactive ones, and $\alpha_1 = \cdots = \alpha_{10} = 1$. We generate the n covariates as $x_{it} = \phi x_{it-1} + u_{it}$, $i = 1, \ldots, n$, $t = 1, \ldots, T$ with $T = 100$, where $u_{it} \sim N(0, 1)$ and $(\mathbf{C}_u)_{ij} = c_{ij}^u = 0.3^{|i-j|}$. We consider $n = 50, 150$ and $\phi = 0.3, 0.6, 0.9, 0.95$. Left panel of Figure 4 displays mean and standard deviation of the ratio between $\widehat{\psi}_{min}^{\widehat{u}}$ (the minimum eigenvalue of $\widehat{\mathbf{C}}_{\widehat{u}}$) and $\widehat{\psi}_{min}^{x}$ obtained from 1000 Monte Carlo simulations run with $n = 50$ (orange circles) and 150 (blu triangles); for $n = 150$ we consider the minimum eigenvalues of the correlation matrices restricted to the 10 relevant variables. As expected, the correlation matrix of the $\widehat{\mathbf{u}}_i$'s, $i = 1, \ldots, n$ does not suffer from spurious correlation induced by serial dependence, and this leads to an increment of $\widehat{\psi}_{min}^{\widehat{u}} / \widehat{\psi}_{min}^{x}$ as ϕ increases. To observe how this result translates into coefficients estimation accuracy, we compare the estimation error of LASSO ($||\widehat{\boldsymbol{\alpha}}_x - \boldsymbol{\alpha}^*||_2$) with that of $\widehat{u}$-LASSO ($||\widehat{\boldsymbol{\alpha}}_{\widehat{u}} - \boldsymbol{\alpha}^*||_2$), where the tuning parameter λ is selected by BIC. Right panel of Figure 4 shows how the mean and standard deviation of $||\widehat{\boldsymbol{\alpha}}_{\widehat{u}} - \boldsymbol{\alpha}^*||_2 / ||\widehat{\boldsymbol{\alpha}}_x - \boldsymbol{\alpha}^*||_2$ vary as a function of ϕ. Also here, as expected, the application of LASSO on serially uncorrelated data reduces the coefficient estimation error, with a gain in estimation accuracy that increases with ϕ. Of course the gains in accuracy shown in right panel of Figure 4 may in part be due to a reduction in bias, as the LASSO inherits the OLS bias illustrated in Proposition 4. However, bias is not the whole story here; the gains in accuracy are also linked to overcoming the spurious correlation induced by serial dependence. In this regard, Table 1 reports average percentages of true and false positives (%TP, %FP) with LASSO and $\widehat{u}$-LASSO; the latter clearly improves variable selection. Fan and Lv (2008) and Fan et al. (2014) pointed out the role of the spurious correlations between relevant and irrelevant variables in the proliferation of false positives; the improved variable selection performance of $\widehat{u}$-LASSO compared to LASSO can be interpreted as partial evidence of the fact that, concurrently with the bias illustrated in Proposition 4, spurious correlation induced by serial dependence strongly contributes to false positives.

In summary, results shown in Figure 4 and Table 1 corroborate the theoretical analysis according to which an increase in the degree of serial dependence leads to an increase in the probability of large spurious correlations, which in turn increases the probability of a small minimum eigenvalue for the sample correlation matrix. This negatively affects the estimation accuracy of PRs (see Proposition 1).

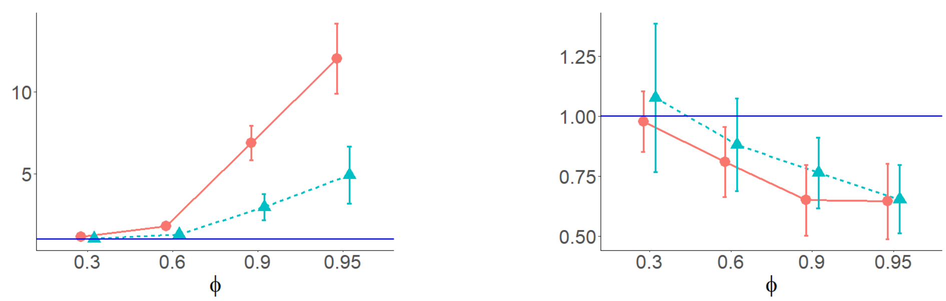


Figure 4: Average and standard deviation of $\widehat{\psi}^{\widehat{u}}_{min}/\widehat{\psi}^{x}_{min}$ (left) and of $||\widehat{\boldsymbol{\alpha}}_{\widehat{u}} - \boldsymbol{\alpha}^*||_2/||\widehat{\boldsymbol{\alpha}}_x - \boldsymbol{\alpha}^*||_2$ (right) across 1000 Monte Carlo replications, for various values of ϕ. Orange circles and bars represent means and standard deviations for $n/T = 0.5$, blue triangles and bars represent means and standard deviations for $n/T = 1.5$. In each panel, the horizontal blue line marks a ratio of 1.

Table 1: Average percentages of true positives (%TP) and false positives (%FP) by LASSO and $\widehat{u}$-LASSO across 1000 Monte Carlo replications, for various values of n/T and ϕ.

ϕ	$n/T = 0.5$				$n/T = 1.5$			
	LASSO		$\widehat{u}$-LASSO		LASSO		$\widehat{u}$-LASSO	
	% TP	% FP	% TP	% FP	% TP	% FP	% TP	% FP
0.3	100.0	13.9	100.0	13.0	100.0	40.4	100.0	48.7
0.6	100.0	25.7	100.0	12.8	100.0	63.5	100.0	28.7
0.9	99.9	45.2	100.0	13.7	99.8	55.7	100.0	20.3
0.95	99.9	47.5	100.0	13.9	99.2	43.8	100.0	19.9

Note: %TP=$\frac{1}{1000}\sum_{k=1}^{1000} \text{TP}_k/10$ and %FP=$\frac{1}{1000}\sum_{k=1}^{1000} \text{FP}_k/(n-10)$ where, for each Monte Carlo replication k, $\text{TP}_k = \#\{\widehat{\alpha}_j \neq 0 : j \in S\}_k$, $\text{FP}_k = \#\{\widehat{\alpha}_j \neq 0 : j \in S^c\}_k$, S=Supp($\boldsymbol{\alpha}^*$), S^c = complement.

5.2.2 Empirical Application

We consider Euro Area (EA) data composed by 309 monthly macroeconomic time series spanning the period between January 1997 and December 2018. The series are listed in Supplement J, grouped according to their measurement domain: Industry & Construction Survey (ICS), Consumer Confidence Indicators (CCI), Money & Interest Rates (M&IR), Industrial Production (IP), Harmonized Consumer Price Index (HCPI), Producer Price Index (PPI), Turnover & Retail Sale (TO), Harmonized Unemployment Rate (HUR), and Service Surveys (SI). Supplement J also reports transformations applied to the series to achieve stationarity (we did not attempt to identify or remove outliers). The target variable is the Overall EA Consumer Price Index (CPI), which is transformed as I(2) (i.e. integration of order 2) following Stock and Watson (2002b):

$$y_{t+h} = (1200/h)\log(CPI_{t+h}/CPI_t) - 1200\log(CPI_t/CPI_{t-1}) \ ,$$

where $y_t = 1200\log(CPI_t/CPI_{t-1}) - 1200\log(CPI_{t-1}/CPI_{t-2})$, and h is the forecasting horizon. We compute forecasts of y_{t+h} at horizons $h = 12$ and 24 using a rolling ω-year window $[t-\omega, t+1]$; the models are re-estimated at each t, adding one observation on the right of the window and removing one observation on the left. The last forecast is December 2018. The methods employed for our empirical exercise are:

- *Univariate AR(p):* the autoregressive forecasting model based on p lagged values of the target variable, i.e. $\widehat{y}_{t+h} = \widehat{\alpha}_0 + \sum_{i=1}^{p} \widehat{\phi}_i y_{t-i+1}$, which serves as a benchmark.
- *LASSO* (Tibshirani, 1996): forecasts are obtained from the equation $\widehat{y}^x_{t+h} = \widehat{\alpha}_0 + \widehat{\boldsymbol{\beta}}'_x \mathbf{v}_t$, where $\widehat{\boldsymbol{\beta}}_x = (\widehat{\phi}_1, \ldots, \widehat{\phi}_{12}, \widehat{\boldsymbol{\alpha}}'_x)'$ is the sparse vector of penalized regression coefficients estimated by the LASSO on the original time series, and $\mathbf{v}_t = (y_t, \ldots, y_{t-11}, \mathbf{x}'_t)'$.

- $\widehat{u}$-*LASSO*: our proposal, where LASSO is applied to the estimated ARMA residuals. Forecasts are obtained from the equation $\widehat{y}^x_{t+h} = \widehat{\alpha}_0 + \widehat{\boldsymbol{\beta}}'_{\widehat{u}}\mathbf{w}_t$, where $\widehat{\boldsymbol{\beta}}_{\widehat{u}} = (\widehat{\phi}_1, \ldots, \widehat{\phi}_{12}, \widehat{\boldsymbol{\alpha}}'_{\widehat{u}})'$ is the sparse vector of penalized regression coefficients estimated by the LASSO on the estimated ARMA residuals, and $\mathbf{w}_t = (y_t, \ldots, y_{t-11}, \widehat{\mathbf{u}}'_t)'$.

For the AR(p) benchmark the lag order p is selected by BIC within $0 \leq p \leq 12$. For the $\widehat{u}$-LASSO, estimated residuals are obtained filtering each time series with an ARMA(p_i, q_i), where p_i and q_i are selected by BIC within $0 \leq p_i, q_i \leq 12$, $i = 1, \ldots, n$. The shrinkage parameter λ of LASSO and $\widehat{u}$-LASSO is selected with BIC and 10-folds cross-validation (CV). Forecasting accuracy for all three methods is evaluated using the root mean square forecast error (RMSFE), defined as

$$RMSFE = \sqrt{\frac{1}{T_1 - T_0} \sum_{\tau=T_0}^{T_1} \left(\widehat{y}_\tau - y_\tau\right)^2}$$

where T_0 and T_1 are the first and last time points used for the out of sample evaluation. For LASSO and $\widehat{u}$-LASSO we also consider the number of selected variables.

Table 2 reports ratios of RMSFEs between pairs of methods, as well as significance of the corresponding Diebold-Mariano test (Diebold and Mariano, 1995). We also report the ratio between the average number of selected variables with $\widehat{u}$-LASSO and with LASSO. Notably, $\widehat{u}$-LASSO produces significantly better forecasts than both the classical LASSO and the AR(p), and provides a more parsimonious model than the LASSO; the ratio between the average number of selected variables is much smaller than <1. This is, in principle, consistent with the theoretical analysis we provided earlier. The sparser $\widehat{u}$-LASSO output may be due to fewer false positives, as compared to the LASSO – since the latter suffers from the effects of spurious correlations induced by serial dependence. However, since in this real data application we do not know the true DGP, any comments regarding accuracy in variable selection is necessarily speculative.

Table 2: Left: ratios of RMSFE contrasting pairs of employed methods; for each ratio we perform a Diebold-Mariano test (alternative: the second method is less accurate in forecasting) and report p-values as 0 '***' 0.001 '**' 0.01 '*' 0.05 '•' 0.1". Right: average of the number of variables selected by $\hat{u}$-LASSO (left of the vertical bar) and LASSO (right of the vertical bar).

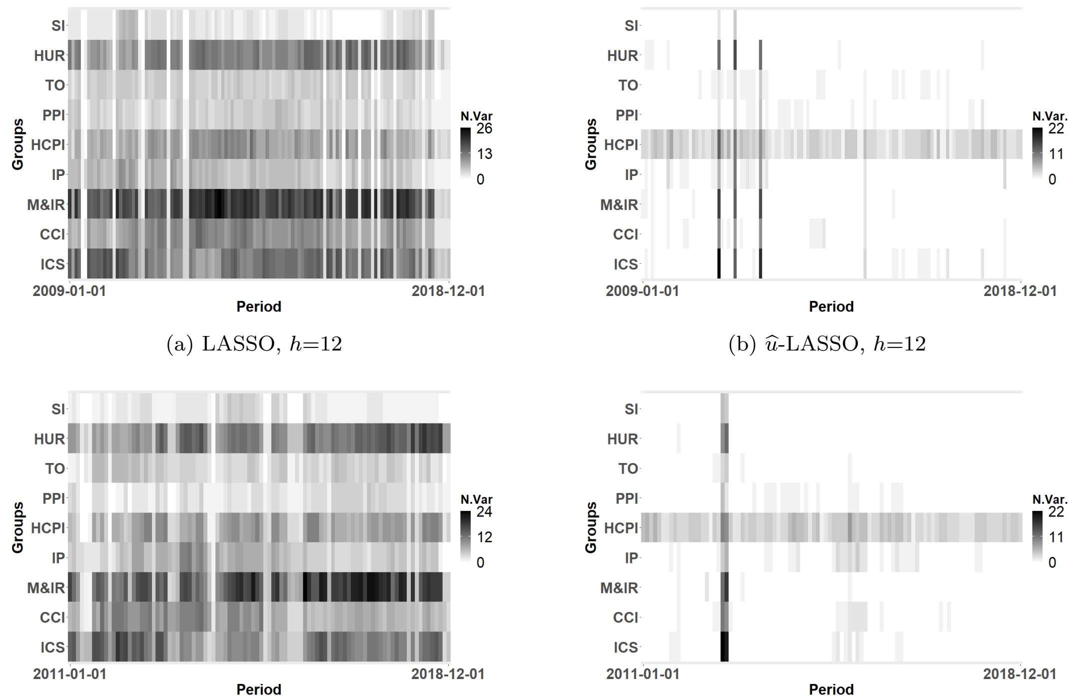
Method 1	Method 2	RMSFE (ratio)				Average of Selected Variables			
		h=12		h = 24		h=12		h = 24	
		BIC	CV	BIC	CV	BIC	CV	BIC	CV
$\widehat{u}$-LASSO	LASSO	0.69***	0.66***	0.82*	0.83**	6.0|67.9	14.8|56.8	6.2|60.9	14.7|57.9
$\widehat{u}$-LASSO	AR(p)	0.94	0.91*	0.89*	0.88**	–	–	–	–
LASSO	AR(p)	1.36	1.38	1.08	1.07	–	–	–	–

Note: For AR(p) coefficients are estimated using the R package *lm*. For $\widehat{u}$-LASSO estimated residuals are obtained by means of an ARMA(p_i, q_i) filter. The penalty parameter λ is selected with BIC using the R package *HDeconometrics*, and with 10-folds cross-validation (CV) using the R package *glmnet*.

In terms of selected variables, Figure 5 summarizes patterns over time obtained with BIC tuning (results obtained with CV are reported in Supplement I). The heatmaps represent the number of selected variables categorized according to the nine main domains (see above). LASSO selects variables largely, though not exclusively, from the domains ICS, M&IR and HUR. $\widehat{u}$-LASSO is more targeted, selecting variables in the HCPI domain. The top 5 variables in terms of selection frequency across forecasting samples are listed in Table 3. Regardless of tuning (BIC, CV) and forecasting horizon h, the top predictor for $\widehat{u}$-LASSO is the Goods Index. The other top predictors, also in the HCPI domain, include EA measurements (e.g., Services Index), or are specific to France and Germany (e.g., All-Items). In summary, $\widehat{u}$-LASSO exploits cross-sectional information mainly focusing on prices, and accrues a forecasting advantage – as LASSO uses many more variables to produce significantly worse forecasts.

Table 3: Five most frequently selected variables. Selection percentages are ratios between the number of times a variable appears in a forecast and total number of forecasts (120 for h=12 and 96 for h = 24).

Rank	Selected Variables			
	h=12		h = 24	
	BIC	CV	BIC	CV
I°	Goods, Index 85.8%	Goods, Index 80%	Goods, Index 85.4%	Goods, Index 85.4%
II°	Industrial Goods, Index 47.5%	Services, Index 77.5%	Services, Index 43.8%	Services, Index 82.3%
III°	Services, Index 40.8%	Industrial Goods, Index 56.7%	All-Items (De) 35.4%	All-Items (Fr) 62.5%
IV°	All-Items Excluding Tobacco, Index 32.5%	All-Items (Fr) 40%	All-Items Excluding Tobacco, Index 32.3%	All-Items (De) 42.7%
V°	All-Items (Fr) 24.2%	All-Items Excluding Tobacco, Index 39.2%	Industrial Goods, Index 30.2%	Industrial Goods, Index 35.4%



(a) LASSO, h=12
(b) $\widehat{u}$-LASSO, h=12
(c) LASSO, h=24
(d) $\widehat{u}$-LASSO, h=24

Figure 5: Heatmaps representing the number of variables selected by LASSO (left) and $\widehat{u}$-LASSO (right) in the nine main domains. The tuning procedure is BIC.

6 Concluding Remarks

In this paper we demonstrated that the probability of spurious correlations between stationary orthogonal or weakly cross-correlated processes depends not only on the sample size, but also on the degree of covariates serial dependence. Through this result, we pointed out that serial dependence negatively affects the behavior of the sample cross-correlation matrix, leading to a large probability of getting a small minimum eigenvalue. Considering the role of the minimum eigenvalue in the non-asymptotic estimation error bounds of PRs, our findings highlight the limitations of these methods in a time series context. In order to improve the estimation performance of PRs in such context, we propose an approach based on applying PRs to pre-whitened (ARMA filter) time series. This proposal allows us to solve the problem of large spurious correlation as well as the problem of biased estimates due to the inclusion of response lags (Achen, 2000).

We assessed the performance of our proposal through Monte Carlo simulations and an empirical application to Euro Area macroeconomic time series. Through simulations we observed that $\widehat{u}$-LASSO, i.e. the LASSO applied on ARMA residuals, reduces the probability of large spurious correlation, performing better than LASSO applied on the original covariates in terms of coefficients estimation. Through the empirical application we observed that $\widehat{u}$-LASSO improves the forecasting performance of LASSO, and produces more parsimonious models. These findings encourage us to further investigate the potential of $\widehat{u}$-PRs – and especially sparse $\widehat{u}$-PRs.

References

Achen, C. H. (2000). why lagged dependent variables can suppress the explanatory power of other independent variables. *Annual Meeting of thePolitial Methodology Section of the American Political Science Association, UCLA*.

Anderson, T. W. (2003). *An Introduction to Multivariate Statistical Analysis* (3rd ed.). New York.

Baillie, R. T., F. X. Diebold, G. Kapetanios, and K. H. Kim (2022, March). On Robust Inference in Time Series Regression.

Bartlett, M. S. (1935). Some aspects of the time-correlation problem in regard to tests of significance. *j-J-R-STAT-SOC-SUPPL 98*(3), 536–543.

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

Cochrane, D. and G. H. Orcutt (1949). Application of least squares regression to relationships containing auto- correlated error terms. *Journal of the American Statistical Association 44*(245), 32–61.

De Mol, C., D. Giannone, and L. Reichlin (2008). Forecasting using a large number of predictors: Is bayesian shrinkage a valid alternative to principal components? *Journal of Econometrics 146*(2), 318–328.

Diebold, F. X. and R. S. Mariano (1995, July). Comparing Predictive Accuracy. *Journal of Business & Economic Statistics 13*(3), 253–263.

Fan, J., F. Han, and H. Liu (2014, feb). Challenges of big data analysis. *National Science Review 1*(2), 293–314.

Fan, J., Y. Ke, and K. Wang (2020). Factor-adjusted regularized model selection. *Journal of Econometrics 216*(1), 71 – 85. Annals Issue in honor of George Tiao: Statistical Learning for Dependent Data.

Fan, J. and J. Lv (2008). Sure independence screening for ultra-high dimensional feature space. *Journal of Royal Statistical Society B*.

Fan, J., Q.-M. Shao, and W.-X. Zhou (2018, Jun). Are discoveries spurious? distributions of maximum spurious correlations and their applications. *The Annals of Statistics 46*(3).

Fan, J. and W.-X. Zhou (2016). Guarding against spurious discoveries in high dimensions. *Journal of Machine Learning Research 17*(203), 1–34.

Forni, M., M. Hallin, M. Lippi, and L. Reichlin (2000). The generalized dynamic-factor model: Identification and estimation. *The Review of Economics and Statistics 82*(4), 540–554.

Glen, A. G., L. M. Leemis, and J. H. Drew (2004). Computing the distribution of the product of two continuous random variables. *Computational Statistics & Data Analysis 44*(3), 451–464.

Granger, C., N. Hyung, and Y. Jeon (2001). Spurious regressions with stationary series. *Applied Economics 33*(7), 899–904.

Granger, C. and P. Newbold (1974). Spurious regressions in econometrics. *Journal of Econometrics 2*(2), 111 – 120.

Hastie, T. (2015). *Statistical learning with sparsity : the lasso and generalizations*. Chapman & Hall/CRC monographs on statistics & applied probability ; 143. Boca Raton, FL: CRC Press.

Hoerl, A. and W. Kennard (1970). Ridge regression: Biased estimation for nonorthogonal problems. pp. 55–67.

Keele, L. and N. J. Kelly (2006). Dynamic models for dynamic theories: The ins and outs of lagged dependent variables. *Political analysis 14*(2), 186–205.

Lounici, K., M. Pontil, A. B. Tsybakov, and S. van de Geer (2009). Taking advantage of sparsity in multi-task learning. *arXiv preprint arXiv:0903.1468*.

Mario, F. and M. Lippi (2001, 02). The generalized dynamic factor model: Representation theory. *Econometric Theory 17*, 1113–1141.

Medeiros, M. C. and E. F.Mendes (2012, August). Estimating High-Dimensional Time Series Models. Textos para discussão 602, Department of Economics PUC-Rio (Brazil).

Negahban, S., B. Yu, M. J. Wainwright, and P. Ravikumar (2009). A unified framework for high-dimensional analysis of m-estimators with decomposable regularizers. In Y. Bengio,

D. Schuurmans, J. Lafferty, C. Williams, and A. Culotta (Eds.), *Advances in Neural Information Processing Systems*, Volume 22. Curran Associates, Inc.

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

Newey, W. K. and K. D. West (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica 55*(3), 703–708.

Proietti, T. and A. Giovannelli (2021). Nowcasting monthly gdp with big data: A model averaging approach. *Journal of the Royal Statistical Society Series A 184*(2), 683–706.

Stock, J. H. and M. W. Watson (2002a). Forecasting using principal components from a large number of predictors. *Journal of the American Statistical Association 97*(460), 1167–1179.

Stock, J. H. and M. W. Watson (2002b). Macroeconomic forecasting using diffusion indexes. *Journal of Business & Economic Statistics 20*(2), 147–162.

Stock, J. H. and M. W. Watson (2008). *Introduction to econometrics / James H. Stock, Mark W. Watson.* (Brief ed. ed.). The Addison-Wesley series in economics. Pearson/Addison-Wesley.

Stuart, A. and K. Ord (1998). *Kendall's advanced theory of statistics* (Sixth ed.), Volume 1, Classical Inference and Relationship.

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

Xin, X., H. Jianhua, and L. Liangyuan (2017). On the oracle property of a generalized adaptive elastic-net for multivariate linear regression with a diverging number of parameters. *Journal of Multivariate AnalysisVolume 162*, 16 – 31.

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

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

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

Zou, H. and H. H. Zhang (2009). On the adaptive elastic-net with a diverging number of parameters. *The Annals of Statistics 37*(4), 1733 – 1751.

Supplement

A Restricted Eigenvalue

In the specific case of $n > T$ the loss function $\mathcal{L}(\boldsymbol{\alpha})$ cannot be strongly convex since $\mathbf{XX}'/T$ is not positive definite. In this specific case Bickel et al. (2009) proposed a solution based on a kind of strong convexity for some subset $\mathcal{C} \subset \mathbb{R}^n$ of possible perturbation vectors $\Delta = |\widehat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}^*| \in \mathbb{R}^n$, named *Restricted Eigenvalue Condition*. In particular, for any subset $S \subseteq \{1, 2, \ldots, n\}$ with cardinality s, let $\Delta_S \in \mathbb{R}^S$ and $\Delta_{S^c} \in \mathbb{R}^{S^c}$. The restricted eigenvalue condition requires that there exists a positive number ν such that

$$\min_{\Delta \in \mathbb{R}^n : \Delta \neq 0} \frac{||\mathbf{X}'\Delta||_2}{\sqrt{T}||\Delta_S||_2} \geq \nu.$$

Such condition is essentially a restriction on the eigenvalues of $\mathbf{XX}'/T$ as a function of sparsity, which allows for the strong convexity to hold whit parameter $\gamma = \nu$, which characterizes how strong the covariates depend on each other. According to Bickel et al. (2009), the restricted eigenvalue condition restricts the LASSO error to a set of the form:

$$\mathcal{C}(S) := \left\{ \widehat{\Delta} \in \mathbb{R}^n : ||\widehat{\Delta}_{S^c}||_1 \leq 3||\widehat{\Delta}_S||_1 \right\}.$$

B On the Population Cross-Correlation in Time Series

Let the response variable be generated according to the following DGP: $y_t = \sum_{i=1}^n \alpha_i^* x_{it-1} + \varepsilon_t$. We consider the case where the covariates are generated as follows

$$x_{it} = \lambda_i F_t + u_{it}, \tag{12}$$

with $i = 1, \ldots, n$, $t = 1, \ldots, T$, where n is comparable to or larger than T and therefore PRs are used in order to estimate α's. F_t represents a common factor that introduce population cross-correlation between covariates, λ_i is the factor loading relative to x_i, and u_{it} is the idiosyncratic component relative to x_i at time t.

In this case, Fan et al. (2020) propose a method to reduce the cross-correlation between covariates in order to improve the estimation accuracy of sparse PRs. It consists in using the principal component analysis to obtain $\widehat{\lambda}_i$, $\widehat{F}_t$ and $\widehat{u}_{it} = x_{it} - \widehat{\lambda}_i \widehat{F}_t$, i.e. estimates of λ_i, F_t and u_t. Hence, when covariates are generated by model (12), the procedure proposed by Fan et al. (2020) allows us to deal with the problem from PRs estimation with highly cross-correlated covariates $\mathbf{x}_1, \ldots, \mathbf{x}_n$ to PRs estimation with weakly or orthogonal covariates $\widehat{\mathbf{u}}_1, \ldots, \widehat{\mathbf{u}}_n$.

However, we stress that if the idiosyncratic components are orthogonal or weakly cross-correlated AR processes, the methodology proposed by Fan et al. (2020) would not solve the problem of the high spurious cross-correlation caused by serial dependence.

C Proofs

To achieve the theoretical results we generalize the approach of Anderson (2003) to our time series context. In particular the equation

$$\frac{\sqrt{a_{11}}\ b}{\sqrt{v/(T-2)}} = \sqrt{T-2}\frac{a_{12}/\sqrt{a_{11}a_{22}}}{\sqrt{1-a_{12}^2/(a_{11}a_{22})}} = \sqrt{T-2}\frac{\widehat{c}_{12}^x}{\sqrt{1-(\widehat{c}_{12}^x)^2}},$$

where $c_{12}^u = 0$, $b = a_{21}/a_{11}$ and $v = a_{22} - a_{21}^2/a_{11}$, is obtained by Anderson (2003, p. 119).

C.1 Proposition 2

We first focus on the distribution of the sample covariance between x_1 and x_2, which is

$$\widehat{Cov}(x_1, x_2) = \frac{a_{12}}{(T-1)} = \\ = \left[\sum_{l=1}^{T-1} \phi_1^l \widehat{Cov}(u_{1[-l]}, u_2) + \sum_{l=1}^{T-1} \phi_2^l \widehat{Cov}(u_{2[-l]}, u_1) + \widehat{Cov}(u_1, u_2)\right](1 - \phi_1\phi_2)^{-1}, \quad (13)$$

where $\widehat{Cov}(u_{i[-l]}, u_j) = \sum_{t=l+1}^{T}(u_{it-l} - \overline{u}_i)(u_{jt} - \overline{u}_j)/(T-l-1)$, for $i \neq j = 1, 2$, $\overline{u}_i = \frac{1}{T-l-1}\sum_{t=l+1}^{T} u_{it-l}$ and $\overline{u}_j = \frac{1}{T-l-1}\sum_{t=l+1}^{T} u_{jt}$. Since u_1 and u_2 are standard Normal, we have (see Glen et al. 2004 and Supplement K)

$$\widehat{Cov}(u_1, u_2) \approx N\left(0\,,\ \frac{1}{(T-1)}\right).$$

Moreover, the quantity

$$\eta_{12} = \sum_{l=1}^{T-1} \phi_1^l \widehat{Cov}(u_{1[-l]}, u_2) + \sum_{l=1}^{T-1} \phi_2^l \widehat{Cov}(u_{2[-l]}, u_1),$$

is a linear combination of the sample covariances between the residual of a time series at time t and the lagged residuals of the other time series. Note that η_{12} is a linear combination of $N\left(0, \frac{\phi_i^{2l}}{T-l-1}\right)$, $i = 1, 2$. However, because $|\phi_i| < 1$, we can approximate η_{12} as a linear combination of centered Normals with variance $\frac{1}{T-1}$, so that

$$\eta_{12} \approx N\left(0, \frac{\phi_1^2 + \phi_2^2 - 2\phi_1^2\phi_2^2}{(T-1)(1-\phi_1^2)(1-\phi_2^2)}\right)$$

and

$$\widehat{Cov}(x_1, x_2) \approx N\left(0\,,\; \frac{1-\phi_1^2\phi_2^2}{(T-1)(1-\phi_1^2)(1-\phi_2^2)(1-\phi_1\phi_2)^2}\right).$$

Since a_{11} is $T-1$ times the sample variance of x_1, $a_{11} \approx \frac{T-1}{1-\phi_1^2}$. Therefore, $b = \frac{a_{21}}{a_{11}}$ is Normally distributed and, based on the approximation of mean and variance of a ratio (see Stuart and Ord 1998), we have $E(b) = 0$ and

$$\begin{aligned} Var(b) = (T-1)^2\frac{Var(a_{12})}{E(a_{11})^2} &\approx (T-1)^2 Var\left(\widehat{Cov}(x_1, x_2)\right)\left(\frac{1-\phi_1^2}{T-1}\right)^2 \\ &= \frac{(1-\phi_1^2\phi_2^2)(1-\phi_1^2)}{(T-1)(1-\phi_2^2)(1-\phi_1\phi_2)^2} \quad . \end{aligned}$$

■

Remark C.1: *For $T < \infty$ the quantity η_{12} has a variance that increases with the degree of serial dependence. This quantity strongly affects the impact of the degree of serial dependence on the variance of a_{12} and, as a consequence, on the variance of both $\widehat{Cov}(x_1, x_2)$ and b.*

Remark C.2: *It is important to note that if x_1 and x_2 are generated by independent MA(q) processes then serial dependence increases the variance of b as the order q increases; see Granger et al. (2001). This is due to the fact that any MA(∞) can be represented as an AR(1). Thus, increasing q, we are faced with the same spurious component η_{12} that impacts the sample covariance between orthogonal AR(1) processes.*

C.2 Proposition 3

To obtain the sample distribution of v we adapt Theorem 3.3.1 in Anderson (2003, p. 75) to the case of AR(1) processes.

Consider a $(T-1) \times (T-1)$ orthogonal matrix $D = (d_{ht})$ with first row $\mathbf{x}_1'/\sqrt{a_{11}}$ and let

$s_h = \sum_{t=1}^{T-1} d_{ht} x_{2t}$, $h = 1, \ldots, (T-1)$, $t = 1, \ldots, (T-1)$. We have

$$b = \frac{\sum_{t=1}^{T-1} x_{1t} x_{2t}}{\sum_{t=1}^{T-1} x_{1t}^2} = \frac{\sum_{t=1}^{T-1} d_{1t} x_{2t}}{\sqrt{a_{11}}} = \frac{s_1}{\sqrt{a_{11}}} \quad .$$

Then, from Lemma 3.3.1 in Anderson (2003, p. 76), we have

$$v = \sum_{t=1}^{T-1} x_{2t}^2 - b^2 \sum_{t=1}^{T-1} x_{1t}^2 = \sum_{t=1}^{T-1} s_t^2 - s_1^2 = \sum_{t=2}^{T-1} s_t^2 \quad .$$

Thus, v approximates the sum of $T-2$ Normal variables with variance $1/(1-\phi_2^2)$. Now, let z_t be the variable obtained by standardizing x_{2t}. We have

$$v = \sum_{t=2}^{T-1} s_t^2 \approx \sum_{t=2}^{T-1} \frac{z_t^2}{1-\phi_2^2} \quad . \tag{14}$$

The right side of (14) is a Gamma distribution with shape parameter $\frac{T-2}{2}$ and rate parameter $\frac{2}{1-\phi_2^2}$. ■

C.3 Theorem 1

Because of Proposition 2, $\sqrt{a_{11}}b$ is approximately $N\left(0, \frac{1-\phi_1^2\phi_2^2}{(1-\phi_2^2)(1-\phi_1\phi_2)^2}\right)$. Let $\delta^2 = \frac{1-\phi_1^2\phi_2^2}{(1-\phi_2^2)(1-\phi_1\phi_2)^2}$, $\theta^2 = \frac{1}{1-\phi_2^2}$ and $t = \frac{\sqrt{a_{11}}\ b}{\sqrt{v/(T-2)}}$. In the reminder of the proof, we consider the distributions of b and v in Propositions 2 and 3 as exact, not approximate. Thus, we have the densities

$$g(\sqrt{a_{11}}b) = \frac{1}{\delta\sqrt{2\pi}} \exp\left(-\frac{a_{11}b^2}{2\delta^2}\right) , \tag{15}$$

$$h(v) = \frac{1}{(2\theta^2)^{\frac{T-2}{2}} \Gamma\left(\frac{T-2}{2}\right)} v^{\frac{T-2}{2}-1} \exp\left(-\frac{v}{2\theta^2}\right) . \tag{16}$$

We focus on

$$
\begin{aligned}
f(t) &= \int \sqrt{\frac{v}{T-2}}\, g\left(\sqrt{\frac{v}{T-2}}\, t\right) h(v) dv \\
&= \int_0^\infty \sqrt{\frac{v}{T-2}} \frac{1}{\delta\sqrt{2\pi}} \exp\left(-\frac{vt^2}{(T-2)2\delta^2}\right) \frac{v^{\frac{T-2}{2}-1}\exp\left(-\frac{v}{2\theta^2}\right)}{(2\theta^2)^{\frac{T-2}{2}}\Gamma\left(\frac{T-2}{2}\right)} dv \\
&= \frac{1}{\sqrt{2\pi(T-2)}\delta(2\theta^2)^{\frac{T-2}{2}}\Gamma\left(\frac{T-2}{2}\right)} \int_0^\infty v^{\frac{1}{2}} v^{\frac{T-2}{2}-1} \exp\left(-\frac{vt^2}{(T-2)2\delta^2}\right) \exp\left(-\frac{v}{2\theta^2}\right) dv \\
&= \frac{1}{\sqrt{2\pi(T-2)}\delta(2\theta^2)^{\frac{T-2}{2}}\Gamma\left(\frac{T-2}{2}\right)} \int_0^\infty v^{\frac{T-3}{2}} \exp\left(-\left(\frac{1}{\theta^2}\frac{t^2}{(T-2)\delta^2}\right)\frac{v}{2}\right) dv \; .
\end{aligned}
$$

Now define $\Upsilon = \frac{1}{\sqrt{2\pi(T-2)}\delta(2\theta^2)^{\frac{T-2}{2}}\Gamma\left(\frac{T-2}{2}\right)}$ and $x = \left(\frac{1}{\theta^2} + \frac{t^2}{(T-2)\delta^2}\right)\frac{v}{2}$. Then

$$
\begin{aligned}
f(t) &= \Upsilon \int_0^\infty \left[2x\left(\frac{1}{\theta^2} + \frac{t^2}{(T-2)\delta^2}\right)^{-1}\right]^{\frac{T-3}{2}} \exp(-x) dx \\
&= \Upsilon \; 2^{\frac{T-1}{2}} \left(\frac{1}{\theta^2} + \frac{t^2}{(T-2)\delta^2}\right)^{-\frac{T-1}{2}} \int_0^\infty x^{\frac{T-1}{2}-1} \exp(-x) dx \; .
\end{aligned}
$$

The integral on the right hand side can be represented by using the gamma function

$$
\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} \exp(-x) dx \; .
$$

Thus we obtain

$$
\begin{aligned}
f(t) &= \Upsilon \; 2^{\frac{T-1}{2}} \left[\frac{(T-2)\delta^2 + t^2\theta^2}{\theta^2(T-2)\delta^2}\right]^{-\frac{T-1}{2}} \Gamma\left(\frac{T-1}{2}\right) \\
&= \frac{\Gamma\left(\frac{T-1}{2}\right) 2^{\frac{T-1}{2}}}{\sqrt{2\pi(T-2)}\delta(2\theta^2)^{\frac{T-2}{2}}\Gamma\left(\frac{T-2}{2}\right)} \left[\frac{(T-2)\delta^2 + t^2\theta^2}{\theta^2(T-2)\delta^2}\right]^{-\frac{T-1}{2}} \\
&= \frac{\Gamma\left(\frac{T-1}{2}\right)\theta}{\sqrt{\pi(T-2)}\delta\Gamma\left(\frac{T-2}{2}\right)} \left[\frac{(T-2)\delta^2 + t^2\theta^2}{(T-2)\delta^2}\right]^{-\frac{T-1}{2}} .
\end{aligned}
$$

Substituting δ^2 with $\frac{1-\phi_1^2\phi_2^2}{(1-\phi_2^2)(1-\phi_1\phi_2)^2}$ and θ^2 with $\frac{1}{1-\phi_2^2}$, we obtain the density

$$f(t) = \frac{\Gamma\left(\frac{T-1}{2}\right)(1-\phi_1\phi_2)\sqrt{(1-\phi_2^2)}}{\Gamma\left(\frac{T-2}{2}\right)\sqrt{\pi(T-2)(1-\phi_1^2\phi_2^2)(1-\phi_2^2)}}\left[1+\frac{t^2(1-\phi_1\phi_2)^2(1-\phi_2^2)}{(T-2)(1-\phi_1^2\phi_2^2)(1-\phi_2^2)}\right]^{-\frac{T-1}{2}}$$
$$= \frac{\Gamma\left(\frac{T-1}{2}\right)(1-\phi_1\phi_2)}{\Gamma\left(\frac{T-2}{2}\right)\sqrt{\pi(T-2)(1-\phi_1^2\phi_2^2)}}\left[1+\frac{t^2(1-\phi_1\phi_2)^2}{(T-2)(1-\phi_1^2\phi_2^2)}\right]^{-\frac{T-1}{2}} .$$

The density of $w = \widehat{c}_{12}^x[1-(\widehat{c}_{12}^x)^2]^{-\frac{1}{2}}$ is thus

$$f(w) = \frac{\Gamma\left(\frac{T-1}{2}\right)(1-\phi_1\phi_2)}{\Gamma\left(\frac{T-2}{2}\right)\sqrt{\pi(1-\phi_1^2\phi_2^2)}}\left[1+\frac{w^2(1-\phi_1\phi_2)^2}{(1-\phi_1^2\phi_2^2)}\right]^{-\frac{T-1}{2}} .$$

Next, define $\kappa(\widehat{c}_{12}^x) = w = \widehat{c}_{12}^x[1-(\widehat{c}_{12}^x)^2]^{-\frac{1}{2}}$, from which $\kappa'(\widehat{c}_{12}^x) = [1-(\widehat{c}_{12}^x)^2]^{-\frac{3}{2}}$, $\phi_{12} = \phi_1\phi_2$ and $\Theta = \left[\Gamma\left(\frac{T-1}{2}\right)(1-\phi_{12})\right]/\left[\Gamma\left(\frac{T-2}{2}\right)\sqrt{\pi(1-\phi_{12}^2)}\right]$. We can use these quantities to write

$$\mathcal{D}_{\widehat{c}_{12}^x} = f_w(\kappa(\widehat{c}_{12}^x))\kappa'(\widehat{c}_{12}^x) = \Theta\left[1+\left(\widehat{c}_{12}^x(1-(\widehat{c}_{12}^x)^2)^{-\frac{1}{2}}\right)^2\frac{(1-\phi_{12})^2}{(1-\phi_{12}^2)}\right]^{-\frac{T-1}{2}}\left[1-(\widehat{c}_{12}^x)^2\right]^{-\frac{3}{2}}$$
$$= \Theta\left[1+\frac{(\widehat{c}_{12}^x)^2(1-\phi_{12})^2}{(1-(\widehat{c}_{12}^x)^2)(1-\phi_{12}^2)}\right]^{-\frac{T-1}{2}}\left[1-(\widehat{c}_{12}^x)^2\right]^{-\frac{3}{2}}$$
$$= \Theta\left[\frac{(1-(\widehat{c}_{12}^x)^2)(1-\phi_{12}^2)+(\widehat{c}_{12}^x)^2(1-\phi_{12})^2}{(1-(\widehat{c}_{12}^x)^2)(1-\phi_{12}^2)}\right]^{-\frac{T-1}{2}}\left[1-(\widehat{c}_{12}^x)^2\right]^{-\frac{3}{2}}$$
$$= \Theta\left[\frac{1-\phi_{12}^2+2(\widehat{c}_{12}^x)^2\phi_{12}(\phi_{12}-1)}{(1-(\widehat{c}_{12}^x)^2)(1-\phi_{12}^2)}\right]^{-\frac{T-1}{2}}\left[1-(\widehat{c}_{12}^x)^2\right]^{-\frac{3}{2}}$$
$$= \Theta\left[1-(\widehat{c}_{12}^x)^2\right]^{\frac{T-4}{2}}\left[\frac{(1-\phi_{12}^2)}{1-\phi_{12}^2+2(\widehat{c}_{12}^x)^2\phi_{12}(\phi_{12}-1)}\right]^{\frac{T-1}{2}} .$$

Thus, the (finite) sample density of $\widehat{c}_{12}^x$, taking the densities in (15) and (16) as exact, is

$$\mathcal{D}_{\widehat{c}_{12}^x} = \frac{\Gamma\left(\frac{T-1}{2}\right)(1-\phi_{12})}{\Gamma\left(\frac{T-2}{2}\right)\sqrt{\pi}}\left[1-(\widehat{c}_{12}^x)^2\right]^{\frac{T-4}{2}}\left(1-\phi_{12}^2\right)^{\frac{T-2}{2}}\left[\frac{1}{1-\phi_{12}^2+2(\widehat{c}_{12}^x)^2\phi_{12}(\phi_{12}-1)}\right]^{\frac{T-1}{2}} .$$

■

C.4 Proof of Proposition 4

In this Section we show that the probability limits of the OLS coefficients are biased as a consequence of serial dependence. To simplify notation, we report results in matrix form, where the subscript $-j$ denotes the corresponding matrix or vector of j period lagged values. Then, we consider the following equations

$$\mathbf{y} = \phi_y \mathbf{y}_{-1} + \mathbf{X}'_{-h}\boldsymbol{\alpha}^* + \boldsymbol{\varepsilon}, \tag{17}$$

$$\mathbf{X} = \phi \mathbf{X}_{-1} + \mathbf{U}, \tag{18}$$

$$\boldsymbol{\varepsilon} = \phi_\varepsilon \boldsymbol{\varepsilon}_{-1} + \boldsymbol{\omega} \tag{19}$$

We provide the convergence in probability of the OLS estimates of ϕ_y and $\boldsymbol{\alpha}^*$, namely $\widehat{\phi}_y$ and $\widehat{\boldsymbol{\alpha}}$, when $\phi_y = 0$, i.e. when we incorrectly include y_{t-1} in our model. This results are based on the contibution of Achen (2000). Note that the following results hold for any lag period h in which $\mathbf{X}_{-h}$ loads on $\mathbf{y}$.

As a consequence of Assumptions 2 and 3 in the text of the main paper:

$$\mathbf{X}_{-h}\boldsymbol{\varepsilon}/T \xrightarrow{p} 0; \qquad \phi\mathbf{X}_{-1}\mathbf{U}'/T \xrightarrow{p} 0; \qquad \phi_\varepsilon \boldsymbol{\varepsilon}'_{-1}\boldsymbol{\omega}/T \xrightarrow{p} 0; \qquad \mathbf{U}_{-h}\boldsymbol{\omega}/T \xrightarrow{p} 0.$$

Let:

$$\mathbf{X}_{-h}\mathbf{X}'_{-h}/T \xrightarrow{p} \mathbf{C}_x; \qquad \boldsymbol{\varepsilon}'\boldsymbol{\varepsilon}/T \xrightarrow{p} c_\varepsilon.$$

Preliminary results:

$$\mathbf{X}_{-h}\mathbf{y}/T \xrightarrow{p} \mathbf{C}_x\boldsymbol{\alpha}^*; \qquad \mathbf{X}_{-h}\mathbf{y}_{-1}/T \xrightarrow{p} \phi\mathbf{C}_x\boldsymbol{\alpha}^*;$$

$$\boldsymbol{\varepsilon}'\mathbf{y}_{-1}\boldsymbol{\omega}/T \xrightarrow{p} \phi_\varepsilon c_\varepsilon; \qquad \mathbf{y}'_{-1}\mathbf{y}/T \xrightarrow{p} \phi\boldsymbol{\alpha}^{*\prime}\mathbf{C}_x\boldsymbol{\alpha}^* + \phi_\varepsilon c_\varepsilon.$$

Moreover, $var(y) = \sigma_y^2 = \boldsymbol{\alpha}^{*\prime}\mathbf{C}_x\boldsymbol{\alpha}^* + c_\varepsilon$.

$$\begin{bmatrix} \widehat{\phi}_y \\ \widehat{\boldsymbol{\alpha}} \end{bmatrix} = \begin{bmatrix} \mathbf{y}'_{-1}\mathbf{y}_{-1}/T & \mathbf{y}'_{-1}\mathbf{X}'_{-h}/T \\ \mathbf{X}_{-h}\mathbf{y}_{-1}/T & \mathbf{X}_{-h}\mathbf{X}'_{-h}/T \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{y}'_{-1}\mathbf{y}/T \\ \mathbf{X}_{-h}\mathbf{y}/T \end{bmatrix}.$$

Applying the results above gives:

$$\begin{bmatrix} \widehat{\phi}_y \\ \widehat{\boldsymbol{\alpha}} \end{bmatrix} \xrightarrow{p} \begin{bmatrix} \sigma_y^2 & \phi\boldsymbol{\alpha}^{*\prime}\mathbf{C}_x \\ \phi\mathbf{C}_x\boldsymbol{\alpha}^* & \mathbf{C}_x \end{bmatrix}^{-1} \begin{bmatrix} \phi\boldsymbol{\alpha}^{*\prime}\mathbf{C}_x\boldsymbol{\alpha}^* + \phi_\varepsilon c_\varepsilon \\ \mathbf{C}_x\boldsymbol{\alpha}^* \end{bmatrix}.$$

Setting $s = \sigma_y^2 - \phi^2\boldsymbol{\alpha}^{*\prime}\mathbf{C}_x\boldsymbol{\alpha}^*$, we have

$$\begin{bmatrix} \widehat{\phi}_y \\ \widehat{\boldsymbol{\alpha}} \end{bmatrix} \xrightarrow{p} \frac{1}{s}\begin{bmatrix} 1 & -\phi\boldsymbol{\alpha}^{*\prime} \\ -\phi\boldsymbol{\alpha}^* & s\mathbf{C}_x^{-1} + \phi^2\boldsymbol{\alpha}^*\boldsymbol{\alpha}^{*\prime} \end{bmatrix} \begin{bmatrix} \phi\boldsymbol{\alpha}^{*\prime}\mathbf{C}_x\boldsymbol{\alpha}^* + \phi_\varepsilon c_\varepsilon \\ \mathbf{C}_x\boldsymbol{\alpha}^* \end{bmatrix}.$$

After some algebra we get:

$$\widehat{\phi}_y \xrightarrow{p} \frac{\phi_\varepsilon c_\varepsilon}{s}, \qquad \widehat{\boldsymbol{\alpha}} \xrightarrow{p} \boldsymbol{\alpha}^*\left(1 - \frac{\phi\phi_\varepsilon c_\varepsilon}{s}\right).$$

Considering that $c_\varepsilon = \sigma_y^2 - \boldsymbol{\alpha}^{*\prime}\mathbf{C}_x\boldsymbol{\alpha}^*$, and $\boldsymbol{\alpha}^{*\prime}\mathbf{C}_x\boldsymbol{\alpha}^* = R^2\sigma_y^2$, then

$$\widehat{\phi}_y \xrightarrow{p} \frac{\phi_\varepsilon(1-R^2)}{1-\phi^2R^2}, \qquad \widehat{\boldsymbol{\alpha}} \xrightarrow{p} \boldsymbol{\alpha}^*\left(1 - \frac{\phi\phi_\varepsilon(1-R^2)}{1-\phi^2R^2}\right).$$

If we replace $\mathbf{X}_{-h}$ with $\mathbf{U}_{-h}$ in the estimated model, then we have the following preliminary results:

$$\mathbf{U}_{-h}\mathbf{U}'_{-h}/T \xrightarrow{p} \mathbf{C}_u; \qquad \mathbf{U}_{-h}\mathbf{y}/T \xrightarrow{p} \mathbf{C}_u\boldsymbol{\alpha}^*; \qquad \mathbf{U}_{-h}\mathbf{y}_{-1}/T \xrightarrow{p} 0.$$

Thus,

$$\begin{bmatrix} \widehat{\phi}_y \\ \widehat{\boldsymbol{\alpha}} \end{bmatrix} = \begin{bmatrix} \mathbf{y}'_{-1}\mathbf{y}_{-1}/T & \mathbf{y}'_{-1}\mathbf{U}'_{-h}/T \\ \mathbf{U}_{-h}\mathbf{y}_{-1}/T & \mathbf{U}_{-h}\mathbf{U}'_{-h}/T \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{y}'_{-1}\mathbf{y}/T \\ \mathbf{U}_{-h}\mathbf{y}/T \end{bmatrix},$$

and applying the results above gives:

$$\begin{bmatrix} \widehat{\phi}_y \\ \widehat{\boldsymbol{\alpha}} \end{bmatrix} \xrightarrow{p} \begin{bmatrix} \sigma_y^2 & 0 \\ 0 & \mathbf{C}_u \end{bmatrix}^{-1} \begin{bmatrix} \phi\boldsymbol{\alpha}^{*\prime}\mathbf{C}_x\boldsymbol{\alpha}^* + \phi_\varepsilon c_\varepsilon \\ \mathbf{C}_u\boldsymbol{\alpha}^* \end{bmatrix},$$

from which

$$\begin{bmatrix} \widehat{\phi}_y \\ \widehat{\boldsymbol{\alpha}} \end{bmatrix} \xrightarrow{p} \frac{1}{\sigma_y^2} \begin{bmatrix} 1 & 0 \\ 0 & \dfrac{\sigma_y^2}{\mathbf{C}_u} \end{bmatrix} \begin{bmatrix} \phi\boldsymbol{\alpha}^{*\prime}\mathbf{C}_x\boldsymbol{\alpha}^* + \phi_\varepsilon c_\varepsilon \\ \mathbf{C}_u\boldsymbol{\alpha}^* \end{bmatrix}.$$

After some algebra we get:

$$\widehat{\phi}_y \xrightarrow{p} \phi R^2 + \phi_\varepsilon(1 - R^2)\ , \qquad \widehat{\boldsymbol{\alpha}} \xrightarrow{p} \boldsymbol{\alpha}^*.$$

■

D Distribution of b

Consider two orthogonal Gaussian AR(1) processes generated according to the model $x_{it} = \phi_i x_{it-1} + u_{it}$, where $u_{it} \sim N(0,1)$, $i = 1,2$, $t = 1,\ldots,100$ and $\phi_1 = \phi_2 = \phi$. In this simulation exercise we run the model

$$x_{2t} = \beta x_{1t} + e_t, \tag{20}$$

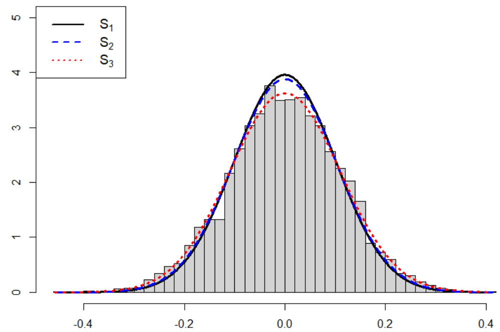
where $e_t \sim N(0, \sigma_e^2)$, and study the distribution of the OLS estimator b of β in the following four cases in terms of degrees of serial dependence: $\phi = 0.3, 0.6, 0.9, 0.95$. Figure 6 reports the density of b across the ϕ values obtained on 5000 Monte Carlo replications. We compare this density with that of three zero-mean Gaussian variables where the variances are respectively:

- $S_1^2 = \frac{\widehat{\sigma}_{\widehat{e}}^2}{\sum_{t=1}^{T}(x_{1t}-\overline{x}_1)^2}$, where $\widehat{\sigma}_{\widehat{e}}^2$ is the sample variance of the estimated residual $\widehat{e}_t = x_{2t} - bx_{1t}$. This is the OLS estimator for the variance of β.

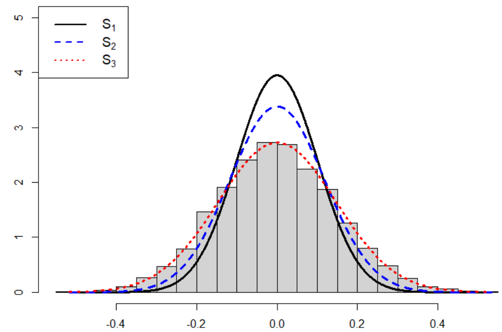
- $S_2^2 = \frac{1}{T}\frac{\frac{1}{T-2}\sum_{t=1}^{T}(x_{1t}-\overline{x}_1)^2\widehat{e}_t^2}{\left[\frac{1}{T}\sum_{t=1}^{T}(x_{1t}-\overline{x}_1)^2\right]^2}\widehat{f}_t$, is the Newey-West (NW) HAC estimator (Newey and West, 1987; Stock and Watson, 2008), where $\widehat{f}_t = \left(1 + 2\sum_{j=1}^{m-1}\left(\frac{m-j}{m}\right)\widehat{\rho}_j\right)$ is the correction factor that adjusts for serially correlated errors and involves estimates of $m-1$ autocorrelation coefficients $\widehat{\rho}_j$, and $\widehat{\rho}_j = \frac{\sum_{t=j+1}^{T}\widehat{v}_t\widehat{v}_{t-j}}{\sum_{t=1}^{T}\widehat{v}_t^2}$, with $\widehat{v}_t = (x_{1t} - \overline{x}_1)\widehat{e}_t$. A rule of thumb for choosing m is $m = [0.75T^{1/3}]$.

- $S_3^2 = \frac{(1-\phi_1^2\phi_2^2)(1-\phi_1^2)}{(T-1)(1-\phi_2^2)(1-\phi_1\phi_2)^2}$, is the theoretical variance of b obtained in Proposition 2.

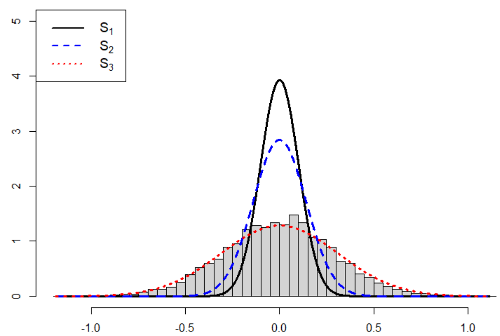
From Figure 6 we observe that the variance of b increases with the degree of serial dependence (ϕ) in a way that is well approximated by the distribution derived in Proposition 2 (see dotted line). On the contrary, OLS (solid line) and NW (dashed line), are highly sub-optimal in the presence of strong serial dependence, underestimating the variability of b as the serial dependence increases.



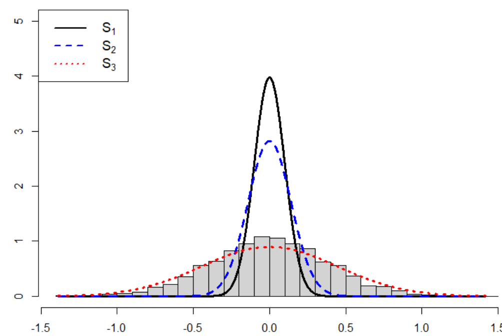
(a) $\phi = 0.3,\ T = 100$



(b) $\phi = 0.6,\ T = 100$



(c) $\phi = 0.9,\ T = 100$



(d) $\phi = 0.95,\ T = 100$

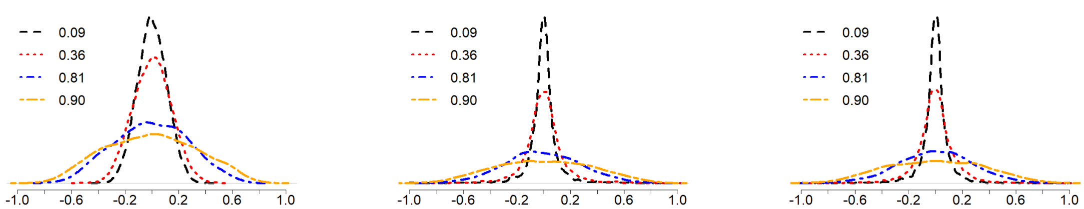
Figure 6: Density of b between uncorrelated AR(1) Gaussian processes. Solid line indicates the approximated density obtained by using the classical OLS estimator, dashed line indicates the approximated density obtained by using the NW estimator, and, finally, dotted line shows the theoretical approximated density obtained in Proposition 2.

E More General Cases

We study the density of $\widehat{c}^{x}_{12}$ in three different cases: non-Gaussian processes; weakly and high cross-correlated processes; and ARMA processes with different order. Note that for the first two cases the variables are AR(1) processes with $T = 100$ and autocorrelation coefficient $\phi = 0.3, 0.6, 0.9, 0.95$. Since we do not have $\mathcal{D}_{\widehat{c}^{x}_{12}}$ for these cases, we rely on the densities obtained on 5000 Monte Carlo replications, i.e. $d_s(\widehat{c}^{x}_{12})$, to show the effect of serial dependence on $\Pr\{|\widehat{c}^x_{12}| \geq \tau\}$.

The Impact of non-Gaussianity

The theoretical contribution reported in Section 3 requires the Gaussianity of u_1 and u_2. With the following simulation experiments we show that the impact of ϕ_{12} on the density of $\widehat{c}^x_{12}$ is relevant also when u_{1t} and u_{2t} are non-Gaussian random variables. To this end, we generate u_{1t} and u_{2t} from the following distributions: Laplace with mean 0 and variance 1 (case (a)); Cauchy with location parameter 0 and scale parameter 1 (case (b)); and from a t-student with 1 degree of freedom (case (c)). Figure 7 reports the results of the simulation experiment. We can state that regardless the distribution of the processes, whenever $Sign(\phi_1) = Sign(\phi_2)$, the probability of large values of $\widehat{c}^x_{12}$ increases with ϕ_{12}. As a curiosity, this result is more evident for the case of Laplace variables, whereas for Cauchy and t-student the effect of ϕ_{12} declines.



(a) Laplace (b) Cauchy (c) t-Student

Figure 7: Simulated density of $\widehat{c}^x_{12}$ in the case of non-Gaussian processes, for $T = 100$ and various values of ϕ.

The Impact of Population Cross-Correlation

Since orthogonality is an unrealistic assumption for most economic applications, here we admit population cross-correlation. In Figure 8 we report $d_s(\widehat{c}^x_{12})$ when the processes are weakly cross-correlated with $c^u_{12} = 0.2$, and when the processes are multicollinear with

$c^u_{12} = 0.8$ (usually we refer to multicollinearity when $c^u_{12} \geq 0.7$). We observe that the impact of ϕ_{12} on $d_s(\widehat{c}^x_{12})$ depends on the degree of (population) cross-correlation as follows. In the case of weakly correlated processes, an increase in ϕ_{12} yields a high probability of observing large sample correlations in absolute value. In the case of multicollinear processes, on the other hand, an increase in ϕ_{12} leads to a high probability of underestimating the true population cross-correlation.

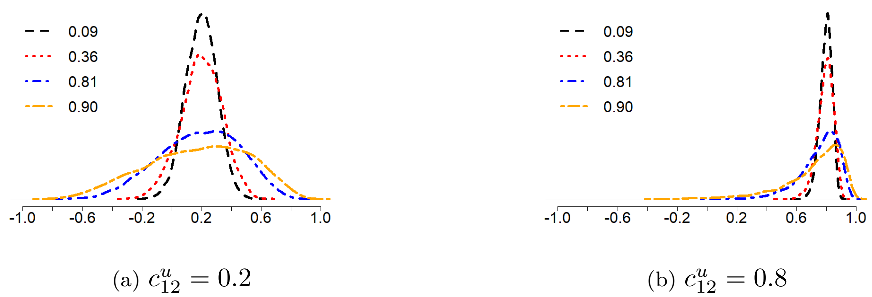


Figure 8: $d_s(\widehat{c}^x_{12})$ obtained through simulations in the case of $c^x_{12} = 0.2$ (a) and $c^x_{12} = 0.8$ (b), for $T = 100$ and various values of ϕ.

Density of $\widehat{c}^x_{12}$ in the case of ARMA(p_i, q_i) processes

To show the effect of serial dependence on a more general case, we generate x_1 and x_2 through the following ARMA processes

$$x_{1t} = \phi x_{1t-1} + \phi x_{1t-2} - \phi x_{1t-3} u_{1t} + 0.5 u_{1t-1},$$

$$x_{2t} = \phi x_{2t-1} + \phi x_{2t-2} + u_{2t} + 0.7 u_{2t-1} - 0.4 u_{3t-2},$$

where $t = 1, \dots, 100$ and $u_i \sim N(0, 1)$. In Figure 9 we report the density of $\widehat{c}^x_{12}$ in the case of $T = 100$ and $\phi = 0.1, 0.2, 0.3, 0.33$. With no loss of generality we can observe that $d_s(\widehat{c}^x_{12})$ gets larger as ϕ increases, that is $\Pr\{|\widehat{c}^x_{12}| \geq \tau\}$ increases with $|\phi|$.

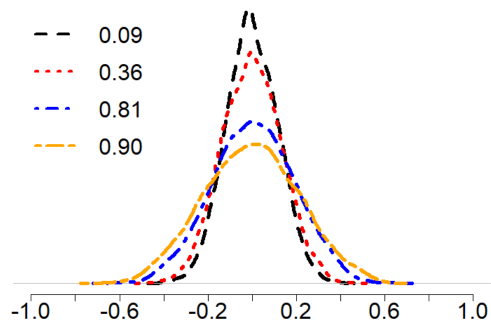


Figure 9: Densities of $d_s(\widehat{c}_{12}^x)$ between two uncorrelated ARMA Gaussian processes, for $T = 100$ and various values of ϕ.

F $\widehat{u}$-OLS: Coefficients Estimation and Prediction Accuracy in Low Dimension

We simulate the response through the equation $y_t = \alpha_1^* x_{1t-1} + \alpha_2^* x_{2t-1} + \varepsilon_t$, and consider three different scenarios:

DGP 1: Equal degrees of serial dependence, with $x_{it} = \phi_i x_{it-1} + u_{it}$, $\varepsilon_t = \phi_\varepsilon \varepsilon_{t-1} + \omega_t$, for $i = 1, 2$, $t = 1, \dots, T$, where $\phi_1 = \phi_2 = \phi_\varepsilon = 0.7$ (as in Example 1). This is the ideal regime in terms of degree of serial dependence, where the working model estimated through $\widehat{u}$-OLS is equivalent to the true model.

DGP 2: Different degrees of serial dependence, with $x_{it} = \phi_i x_{it-1} + u_{it}$, $\varepsilon_t = \phi_\varepsilon \varepsilon_{t-1} + \omega_t$, for $i = 1, 2$, $t = 1, \dots, T$, where $\phi_1 = 0.75$, $\phi_2 = 0.6$, and $\phi_\varepsilon = 0.9$ (as in Example 2). Here the common factor restriction does not hold.

DGP 3: Different models for predictors and error. Here we change DGP of covariates. In particular, we consider $x_{1t} = 0.6x_{1t-1} + u_{1t} + 0.5u_{1t-1}$, $x_{2t} = 0.75x_{2t-1} + u_{2t}$, $\varepsilon_t = 0.6\varepsilon_{t-1} + 0.3\varepsilon_{t-2} + \omega_t$, $t = 1, dots, T$, that is x_{1t}, x_{2t} and ε_t are ARMA(1,1), AR(1) and AR(2) processes, respectively.

For all DGPs the *i.i.d.* errors u_{it} and ω_t are standard Normal random variables for which

Assumptions 2 and 3 hold. Here we compare coefficients estimation and forecasting performance of the following methods:

- NW: the Heteroskedasticity and Autocorrelation Consistent (HAC) Newey-West estimator (Newey and West, 1987), which accommodates autocorrelation and heteroskedasticity of the error terms. The forecasting equation, in terms of the projection of y_t on the hyperplane spanned by the covariates, is $y_t^{(x)} = \text{Proj}(y_t | y_{t-1}, x_{1t-1}, x_{2t-1})$.

- CO: the Cochrane-Orcutt generalized least squares (GLS) estimator (Cochrane and Orcutt, 1949), which adjusts a linear model for serial correlation in the error terms iterating two steps, one to estimate the first order autocorrelation on OLS residuals, and one to transform the variables to eliminate serial dependence in the errors, until a certain criterion is satisfied (e.g., the estimated autocorrelation has converged); transformations are applied from the second observation onward, i.e. for $t = 2, \dots, T$. The forecasting equation is $y_t^{*(x)} = \text{Proj}(y_t^* | x_{1t-1}^*, x_{2t-1}^*)$, where $y_t^* = y_t - \widehat{\phi}_\varepsilon^* y_{t-1}$, $x_{it-1}^* = x_{it-1} - \widehat{\phi}_\varepsilon^* x_{it-2}$, and $\widehat{\phi}_\varepsilon^*$ is the CO estimate of ϕ_ε.

- DynReg: the dynamic regression method (Baillie et al., 2022), which includes lags of the variables as predictors; if the error is an AR(p), one adds to the model p lagged values of y_t and x_{it-1}, $i = 1, \dots, n$. The forecasting equation is $y_t^{(x)} = \text{Proj}(y_t | y_{t-1}, x_{1t-1}, x_{1t-2}, x_{2t-1}, x_{2t-2})$ in Scenarios 1 and 2; and $y_t^{(x)} = \text{Proj}(y_t | y_{t-1}, y_{t-2}, x_{1t-1}, x_{1t-2}, x_{1t-3}, x_{2t-1}, x_{2t-2}, x_{2t-3})$ in Scenario 3.

- $\widehat{u}$-OLS: our proposal, which applies OLS using as predictors $\widehat{u}_{it-1} = x_{it} - \widehat{x}_{it|t-1}$, $i = 1, \dots, n$. The forecasting equation is $y_t^{(\widehat{u})} = \text{Proj}(y_t | y_{t-1}, \widehat{u}_{1t-1}, \widehat{u}_{2t-1})$.

Note that for DGP 1, i.e. under the common factor restriction, also for CO and DynReg the true model and the estimated one coincide. Table 4 reports, for each method, the average and standard deviation of the coefficient estimation error $||\widehat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}^*||_2$ and of the coefficient of determination (R^2) over 1000 Monte Carlo replications, considering $T = 100$ (panel (a)) and $T = 1000$ (panel (b)). Unsurprisingly, NW has the largest coefficient estimation error (it retains OLS estimates and only adjusts standard errors). CO, DynReg and $\widehat{u}$-OLS have smaller and similar coefficient estimation errors. However, in terms of R^2, CO is outperformed by DynReg and $\widehat{u}$-OLS, which both include y_{t-1} as predictor in their forecasting equation. We note that, while DynReg and $\widehat{u}$-OLS have similar estimation and prediction performance, DynReg requires the estimation of more parameters. In fact, to express y_t through n covariates and an AR(p) error, DynReg estimates $p+n+np$ parameters (where p refers to the number of lags of y_t). In contrast, $\widehat{u}$-OLS always estimates $p+n$ parameters. This fact highlights the advantage of using our proposal when n is comparable to or larger than T. In the next Section we also provide an analysis of the t-statistics associated with these methods in the case of spurious regression between uncorrelated autoregressive processes.

DGP	Metric	Stat.	(a) $T = 100$				(b) $T = 1000$			
			NW	CO	DynReg	$\widehat{u}$-OLS	NW	CO	DynReg	$\widehat{u}$-OLS
1	$\|\|\widehat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}^*\|\|_2$	ave.	0.317	0.128	0.129	0.129	0.341	0.040	0.040	0.040
		s.d.	0.132	0.069	0.069	0.068	0.046	0.020	0.020	0.020
	R^2	ave.	0.747	0.682	0.824	0.817	0.747	0.668	0.829	0.829
		s.d.	0.066	0.056	0.045	0.046	0.021	0.017	0.014	0.014
2	$\|\|\widehat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}^*\|\|_2$	ave.	0.351	0.124	0.132	0.134	0.379	0.037	0.039	0.040
		s.d.	0.227	0.066	0.069	0.071	0.238	0.020	0.020	0.021
	R^2	ave.	0.761	0.704	0.836	0.814	0.768	0.695	0.845	0.828
		s.d.	0.066	0.057	0.059	0.065	0.022	0.019	0.046	0.049
3	$\|\|\widehat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}^*\|\|_2$	ave.	0.474	0.126	0.134	0.148	0.579	0.038	0.040	0.044
		s.d.	0.184	0.066	0.070	0.078	0.072	0.020	0.021	0.024
	R^2	ave.	0.789	0.701	0.888	0.846	0.791	0.684	0.900	0.868
		s.d.	0.072	0.054	0.039	0.051	0.034	0.016	0.017	0.020

Table 4: Coefficient estimation error and coefficient of determination (R^2) of Newey-West HAC estimator (NW), Cochrane-Orcutt GLS estimator (CO), Dynamic Regression (DynReg) and OLS applied to the estimated ARMA residuals ($\widehat{u}$-OLS) across the three simulation scenarios (DGPs). Panel (a) $T = 100$, panel (*b*) $T = 1000$. Results are obtained on 1000 Monte Carlo replications.

G Detecting Spurious Regression

Here we generate data as in Supplement D and compare the t-statistics of the Newey-West-style HAC estimators (NW), Cochrane-Orcutt GLS estimator (CO), Dynamic Regression (DynReg) and the ordinary least squares applied on the estimated AR residuals ($\widehat{u}$-OLS) to evaluate their ability in avoiding spurious regressions. In Table 5 we report the percentage of times that the t-statistics are greater than 1.96 in absolute value. Note that according to statistical theory $|t_b| > 1.96$ will occur approximately 5% of the time. The main results from this analysis are: (*i*) OLS estimator (Table 5 panel (a)) suffers of spurious regressions for any $\phi > 0$, which occurs about %50 when $\phi = 0.9$. (*ii*) NW estimator (Table 5 panel (b)) reduces the problem, but for large value of ϕ spurious regression occurs frequently. Note that these two results are in line with those in Granger et al. (2001). (*iii*) CO, DynReg and $\widehat{u}$-OLS (Table 5 panel (c)-(e)) solve the problem of spurious regression due to serial dependence, by making the variance of b independent of ϕ. However, $\widehat{u}$-OLS keeps the advantage already mentioned with respect to CO and DynReg, that are a better prediction accuracy and the estimation of less parameters (see Section 5.2 in the text of the main paper).

Table 5: Percentage of t-statistics over 1.96 in absolute value obtained on 1000 Monte Carlo replications.

	T	$\phi=0.0$	$\phi=0.3$	$\phi=0.6$	$\phi=0.9$	$\phi=0.95$
	50	5.96	8.00	17.16	47.58	56.12
	100	5.38	7.44	18.00	50.54	60.36
(a) $\lvert t_b^{ols}\rvert > 1.96$	250	6.06	7.20	18.26	51.16	64.90
	1000	4.94	7.28	17.72	51.82	66.48
	10000	5.12	7.08	19.00	53.62	65.76
	50	7.32	9.36	16.48	41.82	50.58
	100	6.96	7.58	12.48	36.36	47.72
(b) $\lvert t_b^{nw}\rvert > 1.96$	250	6.96	6.08	9.34	29.00	43.62
	1000	5.00	5.24	7.36	18.88	31.72
	10000	5.32	4.68	6.08	9.48	17.00
	50	6.58	6.76	7.16	8.24	8.42
	100	5.78	5.60	5.92	5.86	6.28
(c) $\lvert t_b^{co}\rvert > 1.96$	250	6.16	4.76	5.42	4.52	5.04
	1000	5.00	5.06	4.74	5.02	5.20
	10000	5.22	5.14	4.80	4.86	5.56
	50	5.88	5.94	6.16	6.12	5.52
	100	5.36	5.16	5.30	5.04	5.52
(d) $\lvert t_b^{dr}\rvert > 1.96$	250	5.86	4.62	5.14	4.56	4.86
	1000	4.86	5.02	4.84	4.88	5.12
	10000	5.22	5.10	4.82	4.86	5.58
	50	6.08	6.48	5.58	6.08	5.06
	100	5.36	5.40	5.34	4.84	5.26
(e) $\lvert t_b^{\widehat{u}-ols}\rvert > 1.96$	250	6.02	4.66	5.10	4.52	4.70
	1000	4.94	5.00	4.78	4.96	5.16
	10000	5.12	5.16	4.86	4.84	5.54

As a further analysis, the following Proposition shows that the variability of the limiting distribution of t_b^{ols} depends only on the degree of serial dependence of the processes.

Proposition G.1: *Let* $S_{ols}^2 = \frac{\widehat{\sigma}_{\widehat{e}}^2}{\sum_{t=1}^T (x_{1t}-\overline{x}_1)^2}$, *where* $\widehat{\sigma}_{\widehat{e}}^2$ *is the estimated variance of the residual of model* (20). *Then*

$$\frac{b}{S_{ols}} \xrightarrow{d} N\left(0, \frac{1-\phi_1^2\phi_2^2}{(1-\phi_1\phi_2)^2}\right).$$

Proof: From Proposition 2 in main text we know that $b \approx N\left(0, \frac{(1-\phi_1^2\phi_2^2)(1-\phi_1^2)}{(T-1)(1-\phi_2^2)(1-\phi_1\phi_2)^2}\right)$. Then, considering $S_{ols}^2 \approx \frac{1-\phi_1^2}{(T-1)(1-\phi_2^2)}$, we have $\frac{b}{S_{ols}} \xrightarrow{\mathrm{d}} N\left(0, \frac{1-\phi_1^2\phi_2^2}{(1-\phi_1\phi_2)^2}\right)$. ■

Note that the result in Proposition G.1 has been also derived in Granger et al. (2001). This result show that the misspecification of t_b^{ols} is only due to the degree of serial dependence. To confirm this, look at the columns of Table 5 and consider that the value of $\lvert t_b^{ols}\rvert$ increases with the degree of serial dependence ϕ, but stay quite constant regardless of the sample size T.

H More Examples with Different Models

For the sake of clarity, models (7) and (11) refer to the true and working model presented in the main text.

Example H.1: *(Equal degrees of serial dependence and different models for the predictors). Consider* x_{it} *and* x_{jt} *generated through models* $x_{it} = \phi x_{it-1} + \phi x_{it-2} + u_{it}$ *and* $x_{jt} = \phi x_{jt-1} + \theta u_{jt-1} + u_{jt}$*, for* $i = 1, \ldots, q$ *and* $j = q+1, \ldots, n$*, where* $2|\phi| < 1$*. Model* (7) *can be rewritten as*

$$
\begin{aligned}
y_t &= \sum_{i=1}^{q} \alpha_i^* (\phi x_{it-2} + \phi x_{it-3} + u_{it-1}) + \sum_{j=1}^{(n-q)} \alpha_j^* (\phi x_{jt-2} + \theta u_{jt-2} + u_{jt-1}) + \phi_\varepsilon \varepsilon_{t-1} + \omega_{it} \\
&= \sum_{i=1}^{q} \alpha_i^* u_{it-1} + \sum_{j=1}^{(n-q)} \alpha_j^* u_{jt-1} + \phi y_{t-1} + \sum_{i=1}^{q} \phi \alpha_i^* x_{it-3} + \sum_{j=1}^{(n-q)} \alpha_j^* \theta u_{jt-2} + \omega_t \ .
\end{aligned}
$$

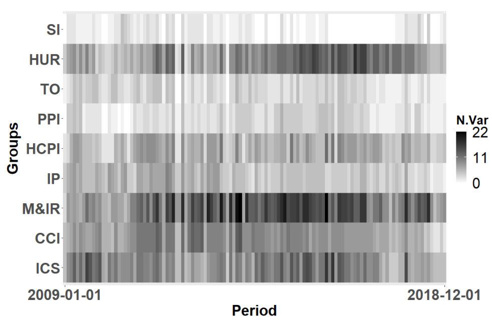
Thus, if we have an "ideal regime" in terms of degree of serial dependence, but different models for the predictors, the working model (11) *is not equivalent to the true model* (7)*. Here, the difference between true and working model is due to the differences between the mechanisms generating* $x_{it|t-1}$ *and* $x_{jt|t-1}$*. Again, this makes* y_{t-1} *not suitable for summarizing the serial dependence of* y_t*.*

Example H.2: *(Equals degrees of serial dependence and different model for the error). Consider now the case where* $\varepsilon_t = \phi \varepsilon_{t-1} + \phi \varepsilon_{t-2} + \omega_t$ *with* $2|\phi| < 1$*. Model* (7) *becomes*

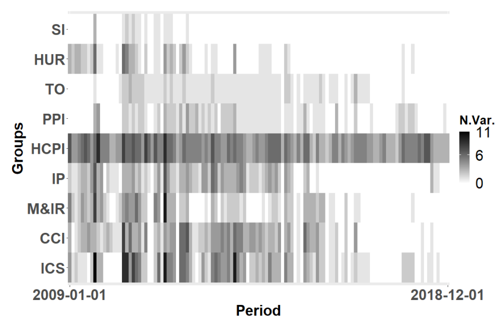
$$
y_t = \sum_{i=1}^{n} \alpha_i^* (\phi x_{it-2} + u_{it-1}) + \phi \varepsilon_{t-1} + \phi \varepsilon_{t-2} + \omega_{it} = \sum_{i=1}^{n} \alpha_i^* u_{1t-1} + \phi y_{t-1} + \phi \varepsilon_{t-2} + \omega_t \ .
$$

Thus, if we have an "ideal regime" in terms of degree of serial dependence, but a different model for the error, the working model (11) *is not equivalent to the true model* (7)*. Here, the difference between true and working model is due to the differences between the mechanism generating $\varepsilon_{t|t-1}$ and the mechanism generating the predictors. In this case, the residual of the working model would have an autoregressive component.*

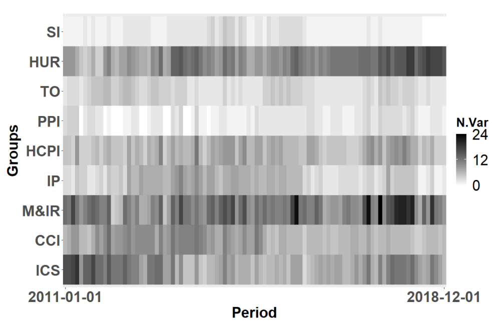
I Heatmaps CV



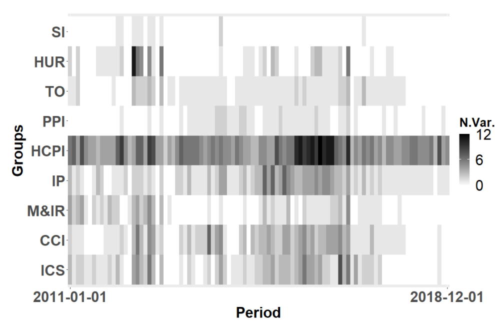
(a) LASSO, h=12



(b) $\widehat{u}$-LASSO, h=12



(c) LASSO, h=24



(d) $\widehat{u}$-LASSO, h=24

Figure 10: Heatmaps of the variables selected by LASSO (left column) and $\widehat{u}$-LASSO (right column) when tuning parameter is selected with CV, categorized by groups.

J List of Time Series in the Euro Area Data

We report the list of series for the Euro Area dataset adopted in the forecasting exercise (obtained from Proietti and Giovannelli (2021)). As for the FRED data, the column tcode denotes the data transformation for a given series x_t: (1) no transformation; (2) Δx_t; (3)$\Delta^2 x_t$; (4) $\log(x_t)$; (5) $\Delta\log(x_t)$; (6) $\Delta^2\log(x_t)$. (7) $\Delta(x_t/x_{t-} - 1.0)$.

The acronyms for the sectors refer to:

- ICS: Industry & Construction Survey
- CCI: Consumer Confidence Indicators
- M&IR: Money & Interest Rates
- IP: Industrial Production
- HCPI: Harm. Consumer Price Index
- PPI: Producer Price Index
- TO: Turnover & Retail Sale
- HUR: Harm. Unemployment rate
- SI: Service Svy.

Table 6: A sample long table.

ID	**Description**	**Area**	**Sector**	**Tcode**
1	Ind Svy: Employment Expectations	EA	ICS	1
2	Ind Svy: Export Order-Book Levels	EA	ICS	1
3	Ind Svy: Order-Book Levels	EA	ICS	1
4	Ind Svy: Mfg - Selling Price Expectations	EA	ICS	1
5	Ind Svy: Production Expectations	EA	ICS	1
6	Ind Svy: Production Trend	EA	ICS	1
7	Ind Svy: Mfg - Stocks Of Finished Products	EA	ICS	1
8	Constr. Svy: Price Expectations	EA	ICS	1

Continued on next page

Table 6 – continued from previous page

ID	Description	Area	Sector	Tcode
9	Ind Svy: Export Order Book Position	EA	ICS	1
10	Ind Svy: Production Trends In Recent Mth.	EA	ICS	1
11	Ind Svy: Selling Prc. Expect. Mth. Ahead	EA	ICS	1
12	Ret. Svy: Employment	EA	ICS	1
13	Ret. Svy: Orders Placed With Suppliers	EA	ICS	1
14	Constr. Svy: Synthetic Bus. Indicator	FR	ICS	1
15	Bus. Svy: Constr. Sector - Capacity Utilisation Rate	FR	ICS	1
16	Constr. Svy: Activity Expectations	FR	ICS	1
17	Constr. Svy: Price Expectations	FR	ICS	1
18	Constr. Svy: Unable To Increase Capacity	FR	ICS	1
19	Constr. Svy: Workforce Changes	FR	ICS	1
20	Constr. Svy: Workforce Forecast Changes	FR	ICS	1
21	Svy: Mfg Output - Order Book & Demand	FR	ICS	1
22	Svy: Mfg Output - Order Book & Foreign Demand	FR	ICS	1
23	Svy: Mfg Output - Personal Outlook	FR	ICS	1
24	Svy: Auto Ind - Order Book & Demand	FR	ICS	1
25	Svy: Auto Ind - Personal Outlook	FR	ICS	1
26	Svy: Basic & Fab Met Pdt Ex Mach & Eq - Personal Outlook	FR	ICS	1
27	Svy: Ele & Elec Eq, Mach Eq - Order Book & Demand	FR	ICS	1
28	Svy: Ele & Elec Eq, Mach Eq - Order Book & Foreign Demand	FR	ICS	1
29	Svy: Ele & Elec Eq, Mach Eq - Personal Outlook	FR	ICS	1
30	Svy: Mfg Output - Price Outlook	FR	ICS	1
31	Svy: Mfg Of Chemicals & Chemical Pdt - Order Book & Demand	FR	ICS	1
32	Svy: Mfg Of Chemicals & Chemical Pdt - Personal Outlook	FR	ICS	1
33	Svy: Mfg Of Food Pr & Beverages - Order Book & Demand	FR	ICS	1
34	Svy: Mfg Of Food Pr & Beverages - Order Book & Foreign Demand	FR	ICS	1

Continued on next page

Table 6 – continued from previous page

ID	Description	Area	Sector	Tcode
35	Svy: Mfg Of Trsp Eq - Finished Goods Inventories	FR	ICS	1
36	Svy: Mfg Of Trsp Eq - Order Book & Demand	FR	ICS	1
37	Svy: Mfg Of Trsp Eq - Order Book & Foreign Demand	FR	ICS	1
38	Svy: Mfg Of Trsp Eq - Personal Outlook	FR	ICS	1
39	Svy: Oth Mfg, Mach & Eq Rpr & Instal - Ord Book & Demand	FR	ICS	1
40	Svy: Oth Mfg, Mach & Eq Rpr & Instal - Ord Book & Fgn Demand	FR	ICS	1
41	Svy: Oth Mfg, Mach & Eq Rpr & Instal - Personal Outlook	FR	ICS	1
42	Svy: Other Mfg - Order Book & Demand	FR	ICS	1
43	Svy: Rubber, Plastic & Non Met Pdt - Order Book & Demand	FR	ICS	1
44	Svy: Rubber, Plastic & Non Met Pdt - Order Book & Fgn Demand	FR	ICS	1
45	Svy: Rubber, Plastic & Non Met Pdt - Personal Outlook	FR	ICS	1
46	Svy: Total Ind - Order Book & Demand	FR	ICS	1
47	Svy: Total Ind - Order Book & Foreign Demand	FR	ICS	1
48	Svy: Total Ind - Personal Outlook	FR	ICS	1
49	Svy: Total Ind - Price Outlook	FR	ICS	1
50	Svy: Wood & Paper, Print & Media - Ord Book & Fgn Demand	FR	ICS	1
51	Trd. & Ind: Bus Sit	DE	ICS	1
52	Trd. & Ind: Bus Expect In 6Mo	DE	ICS	1
53	Trd. & Ind: Bus Sit	DE	ICS	1
54	Trd. & Ind: Bus Climate	DE	ICS	1
55	Cnstr Ind: Bus Climate	DE	ICS	1
56	Mfg: Bus Climate	DE	ICS	1
57	Mfg: Bus Climate	DE	ICS	1
58	Mfg Cons Gds: Bus Climate	DE	ICS	1
59	Mfg (Excl Fbt): Bus Climate	DE	ICS	1
60	Whsle (Incl Mv): Bus Climate	DE	ICS	1

Continued on next page

Table 6 – continued from previous page

ID	Description	Area	Sector	Tcode
61	Mfg: Bus Sit	DE	ICS	1
62	Mfg: Bus Sit	DE	ICS	1
63	Mfg (Excl Fbt): Bus Sit	DE	ICS	1
64	Mfg (Excl Fbt): Bus Sit	DE	ICS	1
65	Cnstr Ind: Bus Expect In 6Mo	DE	ICS	1
66	Cnstr Ind: Bus Expect In 6Mo	DE	ICS	1
67	Mfg: Bus Expect In 6Mo	DE	ICS	1
68	Mfg: Bus Expect In 6Mo	DE	ICS	1
69	Mfg Cons Gds: Bus Expect In 6Mo	DE	ICS	1
70	Mfg (Excl Fbt): Bus Expect In 6Mo	DE	ICS	1
71	Mfg (Excl Fbt): Bus Expect In 6Mo	DE	ICS	1
72	Rt (Incl Mv): Bus Expect In 6Mo	DE	ICS	1
73	Whsle (Incl Mv): Bus Expect In 6Mo	DE	ICS	1
74	Bus. Conf. Indicator	IT	ICS	1
75	Order Book Level: Ind	ES	ICS	1
76	Order Book Level: Foreign - Ind	ES	ICS	1
77	Order Book Level: Investment Goods	ES	ICS	1
78	Order Book Level: Int. Goods	ES	ICS	1
79	Production Level - Ind	ES	ICS	1
80	Cons. Confidence Indicator	EA	CCI	1
81	Cons. Svy: Economic Situation Last 12 Mth. - Emu 11/12	EA	CCI	1
82	Cons. Svy: Possible Savings Opinion	FR	CCI	1
83	Cons. Svy: Future Financial Situation	FR	CCI	1
84	Svy - Households, Economic Situation Next 12M	FR	CCI	1
85	Cons. Confidence Indicator - DE	DE	CCI	1
86	Cons. Confidence Index	DE	CCI	5

Continued on next page

Table 6 – continued from previous page

ID	Description	Area	Sector	Tcode
87	Gfk Cons. Climate Svy - Bus. Cycle Expectations	DE	CCI	1
88	Cons.S Confidence Index	DE	CCI	5
89	Cons. Confidence Climate (Balance)	DE	CCI	1
90	Cons. Svy: Economic Climate Index (N.West It)	IT	CCI	5
91	Cons. Svy: Economic Climate Index (Southern It)	IT	CCI	5
92	Cons. Svy: General Economic Situation (Balance)	IT	CCI	1
93	Cons. Svy: Prices In Next 12 Mths. - Lower	IT	CCI	5
94	Cons. Svy: Unemployment Expectations (Balance)	IT	CCI	1
95	Cons. Svy: Unemployment Expectations - Approx. Same	IT	CCI	5
96	Cons. Svy: Unemployment Expectations - Large Increase	IT	CCI	5
97	Cons. Svy: Unemployment Expectations - Small Increase	IT	CCI	5
98	Cons. Svy: General Economic Situation (Balance)	IT	CCI	1
99	Cons. Svy: Household Budget - Deposits To/Withdrawals	ES	CCI	5
100	Cons. Svy: Household Economy (Cpy) - Much Worse	FR	CCI	5
101	Cons. Svy: Italian Econ.In Next 12 Mths.- Much Worse	FR	CCI	5
102	Cons. Svy: Major Purchase Intentions - Balance	FR	CCI	1
103	Cons. Svy: Major Purchase Intentions - Much Less	FR	CCI	5
104	Cons. Svy: Households Fin Situation - Balance	FR	CCI	1
105	Indl. Prod. - Excluding Constr.	EA	IP	5
106	Indl. Prod. - Cap. Goods	EA	IP	5
107	Indl. Prod. - Cons. Non-Durables	EA	IP	5
108	Indl. Prod. - Cons. Durables	EA	IP	5
109	Indl. Prod. - Cons. Goods	EA	IP	5
110	Indl. Prod.	FR	IP	5
111	Indl. Prod. - Mfg	FR	IP	5
112	Indl. Prod. - Mfg (2010=100)	FR	IP	5

Continued on next page

Table 6 – continued from previous page

ID	Description	Area	Sector	Tcode
113	Indl. Prod. - Manuf. Of Motor Vehicles, Trailers, Semitrailers	FR	IP	5
114	Indl. Prod. - Int. Goods	FR	IP	5
115	Indl. Prod. - Indl. Prod. - Constr.	FR	IP	5
116	Indl. Prod. - Manuf. Of Wood And Paper Products	FR	IP	5
117	Indl. Prod. - Manuf. Of Computer, Electronic And Optical Prod	FR	IP	5
118	Indl. Prod. - Manuf. Of Electrical Equipment	FR	IP	5
119	Indl. Prod. - Manuf. Of Machinery And Equipment	FR	IP	5
120	Indl. Prod. - Manuf. Of Transport Equipment	FR	IP	5
121	Indl. Prod. - Other Mfg	FR	IP	5
122	Indl. Prod. - Manuf. Of Chemicals And Chemical Products	FR	IP	5
123	Indl. Prod. - Manuf. Of Rubber And Plastics Products	FR	IP	5
124	Indl. Prod. - Investment Goods	IT	IP	5
125	Indl. Prod.	IT	IP	5
126	Indl. Prod.	IT	IP	5
127	Indl. Prod. - Cons. Goods - Durable	IT	IP	5
128	Indl. Prod. - Investment Goods	IT	IP	5
129	Indl. Prod. - Int. Goods	IT	IP	5
130	Indl. Prod. - Chemical Products & Synthetic Fibres	IT	IP	5
131	Indl. Prod. - Machines & Mechanical Apparatus	IT	IP	5
132	Indl. Prod. - Means Of Transport	IT	IP	5
133	Indl. Prod. - Metal & Metal Products	IT	IP	5
134	Indl. Prod. - Rubber Items & Plastic Materials	IT	IP	5
135	Indl. Prod. - Wood & Wood Products	IT	IP	5
136	Indl. Prod.	IT	IP	5
137	Indl. Prod. - Computer, Electronic And Optical Products	IT	IP	5
138	Indl. Prod. - Basic Pharmaceutical Products	IT	IP	5

Continued on next page

Table 6 – continued from previous page

ID	Description	Area	Sector	Tcode
139	Indl. Prod. - Constr.	DE	IP	5
140	Indl. Prod. - Ind Incl Cnstr	DE	IP	5
141	Indl. Prod. - Mfg	DE	IP	5
142	Indl. Prod. - Rebased To 1975=100	DE	IP	5
143	Indl. Prod. - Chems & Chem Prds	DE	IP	5
144	Indl. Prod. - Ind Excl Cnstr	DE	IP	5
145	Indl. Prod. - Ind Excl Energy & Cnstr	DE	IP	5
146	Indl. Prod. - Mining & Quar	DE	IP	5
147	Indl. Prod. - Cmptr, Eleccl & Opt Prds, Elecl Eqp	DE	IP	5
148	Indl. Prod. - Interm Goods	DE	IP	5
149	Indl. Prod. - Cap. Goods	DE	IP	5
150	Indl. Prod. - Durable Cons Goods	DE	IP	5
151	Indl. Prod. - Tex & Wearing Apparel	DE	IP	5
152	Indl. Prod. - Pulp, Paper&Prds, Pubshg&Print	DE	IP	5
153	Indl. Prod. - Chem Prds	DE	IP	5
154	Indl. Prod. - Rub&Plast Prds	DE	IP	5
155	Indl. Prod. - Basic Mtls	DE	IP	5
156	Indl. Prod. - Cmptr, Eleccl & Opt Prds, Elecl Eqp	DE	IP	5
157	Indl. Prod. - Motor Vehicles, Trailers&Semi Trail	DE	IP	5
158	Indl. Prod. - Tex & Wearing Apparel	DE	IP	5
159	Indl. Prod. - Paper & Prds, Print, Reprod Of Recrd Media	DE	IP	5
160	Indl. Prod. - Chems & Chem Prds	DE	IP	5
161	Indl. Prod. - Basic Mtls, Fab Mtl Prds, Excl Mach&Eqp	DE	IP	5
162	Indl. Prod. - Repair & Install Of Mach & Eqp	DE	IP	5
163	Indl. Prod. - Mfg Excl Cnstr & Fbt	DE	IP	5
164	Indl. Prod. - Mining & Ind Excl Fbt	DE	IP	5

Continued on next page

Table 6 – continued from previous page

ID	Description	Area	Sector	Tcode
165	Indl. Prod. - Ind Excl Fbt	DE	IP	5
166	Indl. Prod. - Interm & Cap. Goods	DE	IP	5
167	Indl. Prod. - Fab Mtl Prds Excl Mach & Eqp	ES	IP	5
168	Indl. Prod.	ES	IP	5
169	Indl. Prod. - Cons. Goods	ES	IP	5
170	Indl. Prod. - Cap. Goods	ES	IP	5
171	Indl. Prod. - Int. Goods	ES	IP	5
172	Indl. Prod. - Energy	ES	IP	5
173	Indl. Prod. - Cons. Goods, Non-Durables	ES	IP	5
174	Indl. Prod. - Mining	ES	IP	5
175	Indl. Prod. - Mfg Ind	ES	IP	5
176	Indl. Prod. - Other Mining & Quarrying	ES	IP	5
177	Indl. Prod. - Textile	ES	IP	5
178	Indl. Prod. - Chemicals & Chemical Products	ES	IP	5
179	Indl. Prod. - Plastic & Rubber Products	ES	IP	5
180	Indl. Prod. - Other Non-Metal Mineral Products	ES	IP	5
181	Indl. Prod. - Metal Processing Ind	ES	IP	5
182	Indl. Prod. - Metal Products Excl. Machinery	ES	IP	5
183	Indl. Prod. - Electrical Equipment	ES	IP	5
184	Indl. Prod. - Automobile	ES	IP	5
185	Euro Interbank Offered Rate - 3-Month (Mean)	EA	M&IR	5
186	Money Supply: Loans To Other Ea Residents Excl. Govt.	EA	M&IR	5
187	Money Supply: M3	EA	M&IR	5
188	Euro Short Term Repo Rate	FR	M&IR	5
189	Datastream Euro Share Price Index (Mth. Avg.)	FR	M&IR	1
190	Euribor: 3-Month (Mth. Avg.)	FR	M&IR	5

Continued on next page

Table 6 – continued from previous page

ID	Description	Area	Sector	Tcode
191	Mfi Loans To Resident Private Sector	FR	M&IR	5
192	Money Supply - M1	FR	M&IR	5
193	Money Supply - M3	FR	M&IR	5
194	Share Price Index - Sbf 250	DE	M&IR	1
195	Fibor - 3 Month (Mth.Avg.)	DE	M&IR	5
196	Money Supply - M3	DE	M&IR	5
197	Money Supply - M2	DE	M&IR	5
198	Bank Prime Lending Rate / Ecb Marginal Lending Facility	DE	M&IR	5
199	Dax Share Price Index, Ep	IT	M&IR	1
200	Interbank Deposit Rate-Average On 3-Months Deposits	IT	M&IR	5
201	Official Reserve Assets	ES	M&IR	5
202	Money Supply: M3 - Spanish	ES	M&IR	5
203	Madrid S.E - General Index	ES	M&IR	5
204	Hicp - Overall Index	EA	HCPI	6
205	Hicp - All-Items Excluding Energy, Index	EA	HCPI	6
206	Hicp - Food Incl. Alcohol And Tobacco, Index	EA	HCPI	6
207	Hicp - Processed Food Incl. Alcohol And Tobacco, Index	EA	HCPI	6
208	Hicp - Unprocessed Food, Index	EA	HCPI	6
209	Hicp - Goods, Index	EA	HCPI	6
210	Hicp - Industrial Goods, Index	EA	HCPI	6
211	Hicp - Industrial Goods Excluding Energy, Index	EA	HCPI	6
212	Hicp - Services, Index	EA	HCPI	6
213	Hicp - All-Items Excluding Tobacco, Index	EA	HCPI	6
214	Hicp - All-Items Excluding Energy And Food, Index	EA	HCPI	6
215	Hicp - All-Items Excluding Energy And Unprocessed Food, Index	EA	HCPI	6
216	All-Items Hicp	DE	HCPI	6

Continued on next page

Table 6 – continued from previous page

ID	Description	Area	Sector	Tcode
217	All-Items Hicp	ES	HCPI	6
218	All-Items Hicp	FR	HCPI	6
219	All-Items Hicp	IT	HCPI	6
220	Goods (Overall Index Excluding Services)	DE	HCPI	6
221	Goods (Overall Index Excluding Services)	FR	HCPI	6
222	Processed Food Including Alcohol And Tobacco	DE	HCPI	6
223	Processed Food Including Alcohol And Tobacco	ES	HCPI	6
224	Processed Food Including Alcohol And Tobacco	FR	HCPI	6
225	Processed Food Including Alcohol And Tobacco	IT	HCPI	6
226	Unprocessed Food	DE	HCPI	6
227	Unprocessed Food	ES	HCPI	6
228	Unprocessed Food	FR	HCPI	6
229	Unprocessed Food	IT	HCPI	6
230	Non-Energy Industrial Goods	DE	HCPI	6
231	Non-Energy Industrial Goods	FR	HCPI	6
232	Services (Overall Index Excluding Goods)	DE	HCPI	6
233	Services (Overall Index Excluding Goods)	FR	HCPI	6
234	Overall Index Excluding Tobacco	DE	HCPI	6
235	Overall Index Excluding Tobacco	FR	HCPI	6
236	Overall Index Excluding Energy	DE	HCPI	6
237	Overall Index Excluding Energy	FR	HCPI	6
238	Overall Index Excluding Energy And Unprocessed Food	DE	HCPI	6
239	Overall Index Excluding Energy And Unprocessed Food	FR	HCPI	6
240	Ppi: Ind Excluding Constr. & Energy	EA	PPI	6
241	Ppi: Cap. Goods	EA	PPI	6
242	Ppi: Non-Durable Cons. Goods	EA	PPI	6

Continued on next page

Table 6 – continued from previous page

ID	Description	Area	Sector	Tcode
243	Ppi: Int. Goods	EA	PPI	6
244	Ppi: Non Dom. - Mining, Mfg & Quarrying	EA	PPI	6
245	Ppi: Non Dom. Mfg	DE	PPI	6
246	Ppi: Int. Goods Excluding Energy	DE	PPI	6
247	Ppi: Cap. Goods	DE	PPI	6
248	Ppi: Cons. Goods	DE	PPI	6
249	Ppi: Fuel	DE	PPI	6
250	Ppi: Indl. Products (Excl. Energy)	DE	PPI	6
251	Ppi: Machinery	DE	PPI	6
252	Deflated T/O: Ret. Sale In Non-Spcld Str With Food, Bev & Tob	DE	T/O	5
253	Deflated T/O: Oth Ret. Sale In Non-Spcld Str	DE	T/O	5
254	Deflated T/O: Sale Of Motor Vehicle Pts & Acces	DE	T/O	5
255	Deflated T/O: Wholesale Of Agl Raw Matls & Live Animals	DE	T/O	5
256	Deflated T/O: Wholesale Of Household Goods	IT	T/O	5
257	T/O: Ret. Trd, Exc Of Mv , Motorcyles & Fuel	ES	T/O	5
258	T/O: Ret. Sale Of Clth & Leath Gds In Spcld Str	ES	T/O	5
259	T/O: Ret. Sale Of Non-Food Prds (Exc Fuel)	ES	T/O	5
260	T/O: Ret. Sale Of Info, Househld & Rec Eqp In Spcld Str	ES	T/O	5
261	Ek Unemployment: All	EA	HUR	5
262	Ek Unemployment: Persons Over 25 Years Old	EA	HUR	5
263	Ek Unemployment: Women Under 25 Years Old	EA	HUR	5
264	Ek Unemployment: Women Over 25 Years Old	EA	HUR	5
265	Ek Unemployment: Men Over 25 Years Old	EA	HUR	5
266	Fr Hur All Persons (All Ages)	FR	HUR	5
267	Fr Hur Femmes (Ages 15-24)	FR	HUR	5
268	Fr Hur Femmes (All Ages)	FR	HUR	5

Continued on next page

Table 6 – continued from previous page

ID	Description	Area	Sector	Tcode
269	Fr Hur Hommes (Ages 15-24)	FR	HUR	5
270	Fr Hur Hommes (All Ages)	FR	HUR	5
271	Fr Hur All Persons (Ages 15-24)	FR	HUR	5
272	Fr Hurall Persons(Ages 25 And Over)	FR	HUR	5
273	Fr Hur Females (Ages 25 And Over)	FR	HUR	5
274	Fr Hur Males (Ages 25 And Over)	FR	HUR	5
275	Bd Hur All Persons (All Ages)	DE	HUR	5
276	Bd Hur Femmes (Ages 15-24)	DE	HUR	5
277	Bd Hur Femmes (All Ages)	DE	HUR	5
278	Bd Hur Hommes (Ages 15-24)	DE	HUR	5
279	Bd Hur Hommes (All Ages)	DE	HUR	5
280	Bd Hur All Persons (Ages 15-24)	DE	HUR	5
281	Bd Hurall Persons(Ages 25 And Over)	DE	HUR	5
282	Bd Hur Females (Ages 25 And Over)	DE	HUR	5
283	Bd Hur Males (Ages 25 And Over)	DE	HUR	5
284	It Hur All Persons (All Ages)	IT	HUR	5
285	It Hur Femmes (All Ages)	IT	HUR	5
286	It Hur Hommes (All Ages)	IT	HUR	5
287	It Hur All Persons (Ages 15-24)	IT	HUR	5
288	It Hurall Persons(Ages 25 And Over)	IT	HUR	5
289	Es Hur All Persons (All Ages)	ES	HUR	5
290	Es Hur Femmes (Ages 16-24)	ES	HUR	5
291	Es Hur Femmes (All Ages)	ES	HUR	5
292	Es Hur Hommes (Ages 16-24)	ES	HUR	5
293	Es Hur Hommes (All Ages)	ES	HUR	5
294	Es Hur All Persons (Ages 16-24)	ES	HUR	5

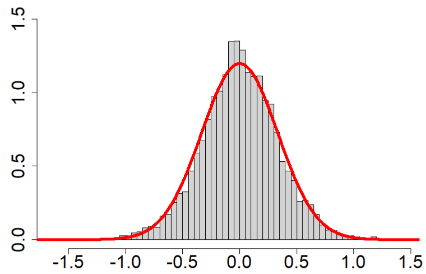
Continued on next page

Table 6 – continued from previous page

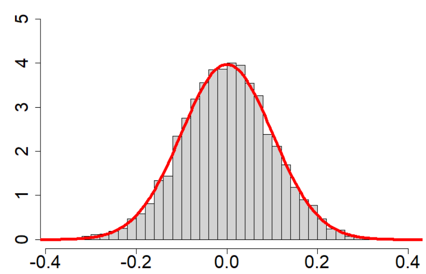
ID	Description	Area	Sector	Tcode
295	Es Hurall Persons(Ages 25 And Over)	ES	HUR	5
296	Es Hur Females (Ages 25 And Over)	ES	HUR	5
297	Es Hur Males (Ages 25 And Over)	ES	HUR	5
298	De - Service Confidence Indicator	DE	SI	1
299	De Services - Buss. Dev. Past 3 Months	DE	SI	1
300	De Services - Evol. Demand Past 3 Months	DE	SI	1
301	De Services - Exp. Demand Next 3 Months	DE	SI	1
302	De Services - Evol. Employ. Past 3 Months	DE	SI	1
303	Fr - Service Confidence Indicator	FR	SI	1
304	Fr Services - Buss. Dev. Past 3 Months	FR	SI	1
305	Fr Services - Evol. Demand Past 3 Months	FR	SI	1
306	Fr Services - Exp. Demand Next 3 Months	FR	SI	1
307	Fr Services - Evol. Employ. Past 3 Months	FR	SI	1
308	Fr Services - Exp. Employ. Next 3 Months	FR	SI	1
309	Fr Services - Exp. Prices Next 3 Months	FR	SI	1

K Distribution of $\widehat{Cov}(u_1, u_2)$

In Figure 11 we report the density of $\widehat{Cov}(u_1, u_2)$ when u_1 and u_2 are standard Normal in the cases of $T = 10$ and 100. Red line shows the density of $N\left(0, \frac{1}{T-1}\right)$. Observations are obtained on 5000 Monte Carlo replications. We observe that the approximation of $\widehat{Cov}(u_1, u_2)$ to $N(0, \frac{1}{T-1})$ holds also when T is small (see Figure 11 (a) relative to T=10). This analysis corroborate numerically the results in Glen et al. (2004), which show that if x and y are $N(0, 1)$, then the probability density function of xy is $\frac{K_0(|xy|)}{pi}$, where $K_0(|xy|)$ is the Bessel function of the second kind.



(a) $T = 10$



(b) $T = 100$

Figure 11: Density of $\widehat{Cov}(u_1, u_2)$ between two uncorrelated standard Normal variables for $T = 10$ (a) and $T = 100$ (b).