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 $\ell_1$-penalty (Tibshirani, 1996; Zou, 2006), the $\ell_2$-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 $\mathbf{a} = (a_1, a_2, \ldots, a_p)'$. $\text{Supp}(\mathbf{a})$ denotes the support of a vector, that is, $\{j \in \{1, 2, \ldots, p\} : a_j \neq 0\}$, and $|\text{Supp}(\mathbf{a})|$ the support cardinality. The $\ell_q$ norm of a vector is $||\mathbf{a}||_q := \left(\sum_{j=1}^{p} |a_j|^q \right)^{1/q}$ for $0 < q < \infty$, with $||\mathbf{a}||_q^k := \left(\sum_{j=1}^{p} |a_j|^q \right)^{k/q}$, and with the usual extension $||\mathbf{a}||_0 := |\text{Supp}(\mathbf{a})|$. Bold capital letters denote matrices, for example $\mathbf{A}$, where $(\mathbf{A})_{ij} = a_{ij}$ denotes its $i$-row $j$-column element. Moreover, $\mathbf{0}_p$ denotes a $p$-length vector of zeros, while $\mathbf{I}_p$ denotes a $p \times p$ identity matrix. Finally, $\text{Sign}(r)$ 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 $\hat{\mathbf{C}}_x = \frac{1}{T-1} \mathbf{X} \mathbf{X}'$ denote the cross-covariance matrix and its estimate, with generic element $\hat{c}_{ij}^x$ and eigenvalues $\hat{\psi}_{max}^x \geq \ldots \geq \hat{\psi}_{min}^x$. 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

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

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

\[
\hat{\alpha} = \arg\min_{\alpha \in \mathbb{R}^n} \left\{ \frac{1}{2T} \|y - X'\alpha\|_2^2 + \lambda \ell(\alpha) \right\}
\]

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

**Definition 1:** (Strong Convexity): Given a differentiable function \( \mathcal{L} : \mathbb{R}^n \to \mathbb{R} \) and the vector differential operator \( \nabla \), we say that \( \mathcal{L} \) is strongly convex with parameter \( \gamma > 0 \) at \( \alpha \in \mathbb{R}^n \) if, for all \( \beta \in \mathbb{R}^n \),

\[
\mathcal{L}(\beta) - \mathcal{L}(\alpha) \geq \nabla \mathcal{L}(\alpha)'(\beta - \alpha) + \frac{\gamma}{2}\|\beta - \alpha\|_2^2.
\]

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 \( \hat{\alpha} \), a small \( |\mathcal{L}(\hat{\alpha}) - \mathcal{L}(\alpha^*)| \) guarantees that \( ||\hat{\alpha} - \alpha^*||_2 \) is also small. The parameter \( \gamma \) governs the strength of convexity; when \( \mathcal{L} \) is twice differentiable, strong convexity requires the minimum eigenvalue of the Hessian \( \nabla^2 \mathcal{L}(\alpha) \) to be at least \( \gamma \) for all \( \alpha \) in a neighborhood of \( \alpha^* \). Thus, since its Hessian is \( \nabla^2 \mathcal{L}(\alpha) = \hat{C}_x \) for all \( \alpha \in \mathbb{R}^n \), the quadratic loss \( \mathcal{L}(\alpha) = \frac{1}{2T}\|y - X'\alpha\|_2^2 \) is strongly convex with parameter \( \gamma \) if and only if \( \hat{\psi}_{\min} \geq \gamma \) (see Hastie 2015, p. 293).

Consequently, in this case \( ||\hat{\alpha} - \alpha^*||_2 \) depends on \( \hat{\psi}_{\min} \). It is important to note that when \( n > T \) the quadratic loss cannot be strongly convex since \( \hat{C}_x \) is singular and thus \( \hat{\psi}_{\min} = 0 \). In this case Bickel et al. (2009) proposed a Restricted Eigenvalue Condition, which is essentially a restriction on the eigenvalues of \( \hat{C}_x \) 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 \( \ell \) and
the inner product $\langle \cdot, \cdot \rangle$, we define the dual norm of $\ell$ as $\ell^*(v) := \sup_{u \in \mathbb{R}^n \setminus \{0\}} \frac{\langle u, v \rangle}{\ell(u)}$. For any subspace $A$ 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 $\Psi(A) := \sup_{u \in A : u \neq 0} \frac{\ell(u)}{|u|_2}$.

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 $\lambda$ is strictly positive and $\geq 2\ell^*\left(\frac{1}{T}X\varepsilon\right)$, and that strong convexity holds with parameter $\gamma > 0$. Then, any optimal solution $\hat{\alpha}$ satisfies the bound $||\hat{\alpha} - \alpha^*||_2 \leq 3\frac{\lambda}{\gamma} \Psi(A)$.

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

The coefficient estimation error bound in Proposition 1 increases with the penalty parameter $\lambda$, which must be strictly positive and $\geq 2\ell^*\left(\frac{1}{T}X\varepsilon\right)$; increases with the subspace compatibility constant $\Psi(A)$, which in turn increases with the size of the model subspace $A$; and decreases with the convexity parameter $\gamma$. 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 $\lambda \geq 2\ell^*\left(\frac{1}{T}X\varepsilon\right)$ when the entries of $X$ and $\varepsilon$ are sub-Gaussian, and assume that strong convexity (or restricted strong convexity) holds with parameter $\gamma$, i.e. that $\hat{\psi}_{min}^x \geq \gamma$ (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 $\hat{\psi}_{min}^x$. 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 $\hat{\psi}_x^{\min}$, 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 $\hat{\psi}_x^{\min} \leq \gamma$ increases with the covariates’ serial dependence. Note that in order to focus on $\hat{\psi}_x^{\min}$ we are assuming that $\hat{C}_x$ is positive definite. If $n > T$ the matrix $\hat{C}_x$ 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 $\hat{\psi}_x^{\min} \leq 1 - \max_{i \neq j} |c_{ij}|$, we emphasize the role of a generic off-diagonal element of $\hat{C}_x$ in determining the probability that $\hat{\psi}_x^{\min} \leq \gamma$ through the inequalities

$$
\Pr\{\hat{\psi}_x^{\min} \leq 1 - \tau\} \geq \Pr\{1 - \max_{i \neq j} |c_{ij}| \leq 1 - \tau\} \geq \Pr\{1 - |c_{i \neq j}| \leq 1 - \tau\} = \Pr\{|c_{i \neq j}| \geq \tau\}. \quad (2)
$$

Thus, $\Pr\{|c_{i \neq j}| \geq \tau\}$ plays a role in determining the probability of dealing with a small $\hat{\psi}_x^{\min}$ 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 $\Pr\{|c_{i \neq j}| \geq \tau\}$ results in an impact on such bound.

To better illustrate our reasoning, we introduce a toy example where we show numer-
ically the impact of serial dependence on $\max_{i \neq j} |\hat{c}_{ij}^x|$ and $\hat{\psi}_{\min}^x$. We generate 10 processes from the model $x_t = D_\phi x_{t-1} + u_t$, $t = 1, \ldots, 100$, where $D_\phi$ is a $10 \times 10$ diagonal matrix with the same autocorrelation coefficient $\phi$ in all positions along the main diagonal, and $u_t \sim N(0, I_{10})$. 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 $C_x$ is $\psi_{\min}^x = 1$. We consider five values for $\phi$, namely $0.0, 0.3, 0.6, 0.9, 0.95$, and for each we calculate the average and the standard deviation of both $\max_{i \neq j} |\hat{c}_{ij}^x|$ and $\hat{\psi}_{\min}^x$ on 5000 Monte Carlo replications. Results are reported in Figure 1. We observe that the stronger the persistence of the process ($\phi$ 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 $\hat{c}_{ij}^x$ for the purpose of formalizing the impact of serial dependence on $\Pr\{||\hat{c}_{ij}^x| \geq \tau\}$. 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 $\max_{i \neq j} |\hat{c}_{ij}^x|$ for various values of the autocorrelation $\phi$, as obtained from 5000 Monte Carlo replications. Similarly, blue triangles and bars represent means and standard deviations of $\hat{\psi}_{\min}^x$.](image-url)
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 \( x_t = D_\phi x_{t-1} + u_t, \ t = 1, \ldots, T, \) where \( D_\phi \) is a 2 \( \times \) 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:** \( u_t \sim N(0, I_2). \)

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

\[
\hat{c}_{12}^x = \frac{a_{12}}{\sqrt{a_{11}a_{22}}},
\]

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

\[
\frac{\sqrt{a_{11}}b}{\sqrt{v/(T-2)}} = \sqrt{T-2} \frac{a_{12}/\sqrt{a_{11}a_{22}}}{\sqrt{1-a_{12}^2/(a_{11}a_{22})}} = \sqrt{T-2} \frac{\hat{c}_{12}^x}{\sqrt{1-(\hat{c}_{12}^x)^2}}.
\]

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 \( \hat{c}_{12}^x \) 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
\[
N\left(0, \frac{(1-\phi_1^2)(1-\phi_2^2)}{(T-1)(1-\phi_1^2)(1-\phi_2^2)^2} \right).
\]

**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 $\hat{c}_{12}$. 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 $\hat{c}_{12}$.

**Theorem 1:** Let $x_i$ be a stationary bivariate Gaussian AR(1) process with autoregressive residuals distributed according to $N(0_2, I_2)$. Further, let $\phi_{12} = \phi_1 \phi_2$ where $\phi_i$, $i = 1, 2$, are the autoregressive coefficients. Then, the sample density of $\hat{c}_{12}$ is approximated by

$$D_{\hat{c}_{12}} = \frac{\Gamma\left(\frac{T-1}{2}\right)\Gamma\left(\frac{T+1}{2}\right)}{\Gamma\left(\frac{T-2}{2}\right)\sqrt{\pi}} (1 - (\hat{c}_{12}^2)^2)\left(1 - \phi_{12}^2\right)^{\frac{T-4}{2}} \left(1 - \phi_{12}^2 + 2(\hat{c}_{12}^2)\phi_{12}(\phi_{12} - 1)\right)^{\frac{T-1}{2}}.$$  (5)

**Proof:** See Supplement C.3

**Remark 2:** $D_{\hat{c}_{12}}$ is the density of the sample correlation coefficient (3) based on a finite $T$, with serial dependence expressed by $\phi_1$ and $\phi_2$, and under the assumption of orthogonal Gaussian AR(1) processes.

**Remark 3:** From (5) we see that $\phi_{12}$ determines the density of $\hat{c}_{12}$ through both its magnitude and sign. More precisely, when $\text{Sign}(\phi_1) = \text{Sign}(\phi_2)$, the probability in the tails increases as $|\phi_{12}|$ grows. On the other hand, when $\text{Sign}(\phi_1) \neq \text{Sign}(\phi_2)$, an increase in $|\phi_{12}|$ leads to a density more concentrated around the origin. This peculiarity on the effect of $\text{Sign}(\phi_{12})$ 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 $\text{Pr}\{|\hat{c}_{12}| \geq \tau\}$ (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] $\text{Pr}\{|\hat{c}_{12}| \geq \tau\} \approx \int_{-1}^{\tau} D_{\hat{c}_{12}} d\hat{c}_{12} + \int_{\tau}^{1} D_{\hat{c}_{12}} d\hat{c}_{12}$, 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 $\hat{c}_{12}$ 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 $\hat{c}_{12}$ obtained by simulations as $d_s(\hat{c}_{12})$.

4.1 Numerical Approximation of $d_s(\hat{c}_{12})$ to $D\hat{c}_{12}$

We generate data from the bivariate process $x_t = D_\phi x_{t-1} + u_t$ for $t = 1, \ldots, T$, where $D_\phi$ is a $2 \times 2$ diagonal matrix with the same autocorrelation coefficient $\phi$ in both positions along the diagonal, and $u_t \sim N(0_2, I_2)$. We consider $T = 50, 100, 250$ and $\phi = 0.3, 0.6, 0.9, 0.95$ – thus, the parameter $\phi_{12}$ in $D\hat{c}_{12}$, here equal to $\phi^2$, 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 $\phi_{12}$, the density $d_s(\hat{c}_{12})$ generated through $5000$ Monte Carlo replications. The second row of Figure 2 (Plots (d), (e), (f)) shows the corresponding $D\hat{c}_{12}$. 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 $d_s(\hat{c}_{12})$ to $D\hat{c}_{12}$ improves as $T$ increases and/or $\phi_{12}$ decreases. In particular, Plots (a) and (d) in Figure 2, where $T = 50$, show that $D\hat{c}_{12}$ approximates $d_s(\hat{c}_{12})$ 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. $\phi \geq 0.9$), $D\hat{c}_{12}$ has larger tails compared to $d_s(\hat{c}_{12})$; 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 $D_{c_{12}}$. In fact, for a sufficiently large finite $T$, we observe that $\Pr\{|\widehat{c}_{12}| \geq \tau\}$, $\tau > 0$, increases with $\phi_{12}$ in a similar way for $d_s(\widehat{c}_{12})$ and $D_{c_{12}}$.

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

In Remark 3 we pointed out that the impact of $\phi_{12}$ on $D_{c_{12}}$ depends on $\text{Sign}(\phi_{12})$. In particular, when $-1 < \phi_{12} < 0$, an increment on $|\phi_{12}|$ makes the density of $\widehat{c}_{12}$ 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 $D_{\phi}$; 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 $\text{Sign}(\phi_1) \neq \text{Sign}(\phi_2)$ and $|\phi_{12}|$ increases, $d_s(\widehat{c}_{12})$ increases its concentration around 0 in a way that is, again, well approximated by $D_{c_{12}}$. 

Figure 2: Monte Carlo densities for $\widehat{c}_{12}$ (top) and corresponding $D_{c_{12}}$ (bottom) for various $T$ and $\phi_{12}$.
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:

\[ x_{1t} = (\phi + 0.1)x_{1t-1} + (\phi + 0.1)x_{1t-2} - 0.2x_{1t-3} + u_{1t}, \]
\[ x_{2t} = \phi x_{2t-1} + \phi x_{2t-2} + u_{2t} + 0.8u_{2t-1}, \]

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 \( \rho_{12}^u = 0.2 \). In these more general cases, we do not know an approximate theoretical density for \( \hat{c}_{12}^u \). Therefore, we rely entirely on simulations to show the effect of serial dependence on \( \Pr\{|\hat{c}_{12}^x| \geq \tau\} \). Plot (c) of Figure 3 shows \( d_s(\hat{c}_{12}^x) \) obtained from 5000 Monte Carlo replications for the different values of \( \phi \). 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.

![Figure 3: Monte Carlo densities for \( \hat{c}_{12}^x \) (a) and corresponding \( D_{\hat{c}_{12}^x} \) (b), for \( T=100 \) and various (negative) \( \phi_{12} \)'s. Monte Carlo densities for \( \hat{c}_{12}^x \) in the case of Laplace weakly correlated AR(3) and ARMA(2,1) processes, for \( T = 100 \) and various \( \phi \)'s (c).](image)

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

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

(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 autoregression, 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 $\hat{x}_{it|t-1}$ be an estimate of $x_{it|t-1}$. We propose to run PRs using the estimated residuals $\hat{u}_{it} = x_{it} - \hat{x}_{it|t-1}$.

5.1 The Working Model on ARMA residuals

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

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

(7)

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

(8)

$$\varepsilon_t = \phi^*_\varepsilon \varepsilon_{t-1} + \omega_t ,$$

(9)

where $i = 1, \ldots, n$, $t = 1, \ldots, T$, $|\phi_i| < 1$ $|\phi^*_\varepsilon| < 1$, and $u_{it}$ and $\omega_t$ 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

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

(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 \( x_{it} - \hat{x}_{it|t-1} \). Our proposal utilizes the *working model*

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

(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 \( \sigma_y^2 \) be the variance of \( y \) and \( R^2 = \alpha^*/\alpha^* \sigma_y^2 \sigma_y^2 \) be the asymptotic coefficient of determination, where \( XX'/T \overset{p}{\rightarrow} C_x \). Under Assumptions 2 and 3 as \( T \rightarrow \infty \), OLS
estimates for model (10) converge as follows

\[ \hat{\phi}_y \xrightarrow{p} \frac{\phi \epsilon (1 - R^2)}{1 - \phi^2 R^2}, \quad \hat{\alpha} \xrightarrow{p} \alpha^* \left( 1 - \frac{\phi \phi \epsilon (1 - R^2)}{1 - \phi^2 R^2} \right) \]

while OLS estimates for model (11) converge as follows

\[ \hat{\phi}_y \xrightarrow{p} \phi R^2 + \phi \epsilon (1 - R^2), \quad \hat{\alpha} \xrightarrow{p} \alpha^*. \]

Proof: See Supplement C.4

Proposition 4 shows that applying OLS to the covariates induces an asymptotic bias in the estimation of \( \alpha^* \) 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 \( T \ll \infty \) or \( T \to \infty \), the estimation error of \( \hat{u}-\)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 $\phi$ and $\phi_{\varepsilon}$ increase. Thus, we proceed by comparing (7) and (11). The latter allows us to estimate the true vector $\alpha^*$ 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 \textsuperscript{2} and \textsuperscript{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

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

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 $\varepsilon_t$. Note that by Proposition 4, if the common factor restriction holds, $\hat{\phi}_y \overset{p}{\to} \phi$.

**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

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

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 se-
rial 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 $\varepsilon_t = \sum_{j=1}^{p_e} \phi_{ej} \varepsilon_{t-j} + \omega_t$, 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 $x_i$'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 $u_i$'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 $\hat{u}$-PRs. In the following Sections we demonstrate the estimation and forecasting performance of $\hat{u}$-LASSO (LASSO applied on ARMA residuals) through both simulations and an empirical application.
5.2 $\hat{u}$-LASSO

5.2.1 Coefficient Estimation Error Bound

Here, we present Monte Carlo experiments to assess the effectiveness of $\hat{u}$-LASSO in reducing the coefficient estimation error. We generate the response as $y_t = \sum_{i=1}^{n} \alpha^*_i x_{it-1} + \epsilon_t$, where $\epsilon_t = \phi \epsilon_{t-1} + \omega_t$, and $\omega_t \sim N(0, \sigma^2)$. The coefficient vector $\alpha^* = (\alpha^*_1, \ldots, \alpha^*_n)'$ is sparse with $||\alpha^*||_0 = 10$. 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 $(C_u)_{ij} = c_{ij}^u = 0.3^{|i-j|}$. 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 $\hat{\psi}^u_{\text{min}}$ (the minimum eigenvalue of $\hat{C}_u$) and $\hat{\psi}^x_{\text{min}}$ 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 $\hat{u}_i$’s, $i = 1, \ldots, n$ does not suffer from spurious correlation induced by serial dependence, and this leads to an increment of $\hat{\psi}^u_{\text{min}} / \hat{\psi}^x_{\text{min}}$ as $\phi$ increases. To observe how this result translates into coefficients estimation accuracy, we compare the estimation error of LASSO ($||\hat{\alpha}_x - \alpha^*||_2$) with that of $\hat{u}$-LASSO ($||\hat{\alpha}_u - \alpha^*||_2$), where the tuning parameter $\lambda$ is selected by BIC. Right panel of Figure 4 shows how the mean and standard deviation of $||\hat{\alpha}_u - \alpha^*||_2 / ||\hat{\alpha}_x - \alpha^*||_2$ vary as a function of $\phi$. 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 $\phi$. 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 \( \hat{u} \)-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 \( \hat{u} \)-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 4).

![Figure 4: Average and standard deviation of \( \hat{\psi}_{\text{min}} / \hat{\psi}_{\text{min}}\) (left) and of \( ||\hat{\alpha}_x - \alpha^*||_2 / ||\hat{\alpha}_x - \alpha^*||_2\) (right) across 1000 Monte Carlo replications, for various values of \( \phi \). 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.](image)

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

| \( \phi \) | \n/T = 0.5 | \n/T = 1.5 |
| --- | --- | --- |
| LASSO | \( \hat{u} \)-LASSO | LASSO | \( \hat{u} \)-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 | 38.7 |
| 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: \( %\text{TP} = \frac{1}{1000} \sum_{k=1}^{1000} \text{TP}_k / 10 \) and \( %\text{FP} = \frac{1}{1000} \sum_{k=1}^{1000} \text{FP}_k / (n-10) \) where, for each Monte Carlo replication \( k \), \( \text{TP}_k = \# \{ \hat{\alpha}_j \neq 0 : j \in S \}_k \), \( \text{FP}_k = \# \{ \hat{\alpha}_j \neq 0 : j \in S^c \}_k \), \( S = \text{Supp}(\alpha^*) \), \( S^c = \text{complement} \).
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):

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

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 \(\omega\)-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. \(\hat{y}_{t+h} = \hat{\alpha}_0 + \sum_{i=1}^{p} \hat{\phi}_i y_{t-i+1}\), which serves as a benchmark.

- **LASSO (Tibshirani, 1996)**: forecasts are obtained from the equation \(\hat{y}_{t+h}^x = \hat{\alpha}_0 + \hat{\beta}_x^\prime v_t\), where \(\hat{\beta}_x = (\hat{\phi}_1, \ldots, \hat{\phi}_{12}, \hat{\alpha}^\prime)\) is the sparse vector of penalized regression coefficients estimated by the LASSO on the original time series, and \(v_t = (y_t, \ldots, y_{t-11}, x_t)^\prime\).
• \(\hat{u}\)-LASSO: our proposal, where LASSO is applied to the estimated ARMA residuals. Forecasts are obtained from the equation \(\hat{y}_{t+h} = \hat{\alpha}_0 + \hat{\beta}_u' w_t\), where \(\hat{\beta}_u = (\hat{\phi}_1, \ldots, \hat{\phi}_{12}, \hat{\alpha}_u')'\) is the sparse vector of penalized regression coefficients estimated by the LASSO on the estimated ARMA residuals, and \(w_t = (y_t, \ldots, y_{t-11}, \hat{u}_t')'\).

For the AR(\(p\)) benchmark the lag order \(p\) is selected by BIC within \(0 \leq p \leq 12\). For the \(\hat{u}\)-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 \(\lambda\) of LASSO and \(\hat{u}\)-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

\[
RMSFE = \sqrt{\frac{1}{T_1 - T_0} \sum_{r = T_0}^{T_1} (\hat{y}_r - y_r)^2}
\]

where \(T_0\) and \(T_1\) are the first and last time points used for the out of sample evaluation.

For LASSO and \(\hat{u}\)-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 \(\hat{u}\)-LASSO and with LASSO. Notably, \(\hat{u}\)-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 \(\hat{u}\)-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 |
|----------|-------------|---------------|-------------------------------|
| \( \hat{u} \)-LASSO | LASSO       | \( h=12 \) | \( h=24 \) | \( h=12 \) | \( h=24 \) |
| BIC      | CV          | BIC           | CV                           | BIC           | CV                           |
| \( \hat{u} \)-LASSO | AR(\( p \)) | 0.69***       | 0.66***                      | 0.82*         | 0.83**                      |
| LASSO    | AR(\( p \)) | 1.36          | 1.38                         | 1.08          | 1.07                         |

Note: For AR(\( p \)) coefficients are estimated using the R package lm. For \( \hat{u} \)-LASSO estimated residuals are obtained by means of an ARMA(\( p_i, q_i \)) filter. The penalty parameter \( \lambda \) 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. \( \hat{u} \)-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 \( \hat{u} \)-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, \( \hat{u} \)-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 | BIC Selected Variables | CV Selected Variables | BIC Selected Variables | CV Selected Variables |
|------|------------------------|-----------------------|------------------------|-----------------------|
| 1*   | Goods, index           | Goods, index          | Goods, index           | Goods, index          |
| 85.8%| 80%                    | 85.4%                 | 85.4%                  |
| 2*   | Industrial Goods, index| Services, index       | Services, index        | Services, index       |
| 47.5%| 77.5%                  | 43.8%                 | 82.3%                  |
| 3*   | Services, Index        | Industrial Goods, Index| All-Items (De)         | All-Items (Fr)        |
| 40.8%| 56.7%                  | 35.4%                 | 62.5%                  |
| 4*   | All-Items Excluding Tobacco, Index| All-Items (Fr)      | All-Items Excluding Tobacco, Index | All-Items (De) |
| 32.5%| 40%                    | 32.3%                 | 42.7%                  |
| V*   | All-Items (Fr)         | All-Items Excluding Tobacco, Index | Industrial Goods, Index | Industrial Goods, Index |
| 24.2%| 39.2%                  | 30.2%                 | 35.4%                  |
Figure 5: Heatmaps representing the number of variables selected by LASSO (left) and $\hat{u}$-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 $\hat{u}$-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 $\hat{u}$-LASSO improves the forecasting performance of LASSO, and produces more parsimonious models. These findings encourage us to further investigate the potential of $\hat{u}$-PRs – and especially sparse $\hat{u}$-PRs.

References

Achen, C. H. (2000). why lagged dependent variables can suppress the explanatory power of other independent variables. *Annual Meeting of the Political 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-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,
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 Analysis* Volume 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.
A Restricted Eigenvalue

In the specific case of \( n > T \) the loss function \( \mathcal{L}(\alpha) \) cannot be strongly convex since \( \mathbf{X}\mathbf{X}'/T \) 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 \( \Delta = |\tilde{\alpha} - \alpha^*| \in \mathbb{R}^n \), 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 \( \nu \) 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 \( \mathbf{X}\mathbf{X}'/T \) as a function of sparsity, which allows for the strong convexity to hold with 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:

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

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

\[
x_{it} = \lambda_i F_t + u_{it},
\]

(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 $\alpha$’s. $F_t$ represents a common factor that introduce population cross-correlation between covariates, $\lambda_i$ 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 $\hat{\lambda}_i$, $\hat{F}_t$ and $\hat{u}_{it} = x_{it} - \hat{\lambda}_i \hat{F}_t$, i.e. estimates of $\lambda_i$, $F_t$ 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 $x_1, \ldots, x_n$ to PRs estimation with weakly or orthogonal covariates $\hat{u}_1, \ldots, \hat{u}_n$.

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

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

where $c_{12} = 0$, $b = a_{21}/a_{11}$ 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

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

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

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

Moreover, the quantity

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

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 $\eta_{12}$ is a linear combination of $N\left(0, \frac{\phi_i^2}{T-l-1}\right)$, $i = 1, 2$. However, because $|\phi_i| < 1$, we can approximate $\eta_{12}$ as a linear combination of centered Normals with variance $\frac{1}{T-l-1}$, 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
\[ \hat{\text{Cov}}(x_1, x_2) \approx N\left(0, \frac{1 - \phi_1^2 \phi_2^2}{(T - 1)(1 - \phi_1^2)(1 - \phi_2^2)(1 - \phi_1 \phi_2)^2}\right). \]

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
\[
Var(b) = (T - 1)^2 \frac{\text{Var}(a_{12})}{E(a_{11})^2} \approx (T - 1)^2 \text{Var}\left(\hat{\text{Cov}}(x_1, x_2)\right) \left(\frac{1 - \phi_1^2}{T - 1}\right)^2
\]
\[
\quad = \frac{(1 - \phi_1^2 \phi_2^2)(1 - \phi_1^2)}{(T - 1)(1 - \phi_2^2)(1 - \phi_1 \phi_2)^2}.
\]

\[ \blacksquare \]

**Remark C.1:** For \( T < \infty \) the quantity \( \eta_{12} \) 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 \( \hat{\text{Cov}}(x_1, x_2) \) 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(\( \infty \)) can be represented as an AR(1). Thus, increasing \( q \), we are faced with the same spurious component \( \eta_{12} \) 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 \( x_1'/\sqrt{a_{11}} \) and let
\[ s_h = \sum_{t=1}^{T-1} d_{ht}x_{2t}, \quad h = 1, \ldots, (T - 1), \quad t = 1, \ldots, (T - 1). \]

We have

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

Then, from Lemma 3.3.1 in [Anderson 2003, p. 76], we have

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

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

\[ v = \sum_{t=2}^{T-1} s_t^2 \approx \sum_{t=2}^{T-1} \frac{z_t^2}{1 - \phi_2^2}. \]

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

### C.3 Theorem 1

Because of Proposition 2, \( \sqrt{a_{11}}b \) is approximately \( N\left(0, \frac{1 - \phi_1^2 \phi_2^4}{(1 - \phi_2^2)(1 - \phi_1 \phi_2)^2}\right) \). Let \( \delta^2 = \frac{1 - \phi_1^2 \phi_2^4}{(1 - \phi_2^2)(1 - \phi_1 \phi_2)^2} \), \( \theta^2 = \frac{1}{1 - \phi_2^2} \) 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

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

\[ h(v) = \frac{1}{(2\theta^2)^{\frac{T-2}{2}} \Gamma\left(\frac{T-2}{2}\right)} v^{\frac{T-2}{2} - 1} \exp\left(-\frac{v}{2\theta^2}\right). \quad (16) \]
We focus on

\[ f(t) = \int \sqrt{\frac{v}{T-2}} g\left(\sqrt{\frac{v}{T-2}} t\right) h(v) dv \]

\[ = \int_{0}^{\infty} \sqrt{\frac{v}{T-2}} \frac{1}{\delta \sqrt{2\pi}} \exp\left(-\frac{vt^2}{(T-2)\delta^2}\right) \frac{v^{T-2}}{2} \exp\left(-\frac{v}{2\theta^2}\right) dv \]

\[ = \frac{1}{\sqrt{2\pi(T-2)\delta(2\theta^2) \frac{T-2}{2} \Gamma\left(\frac{T-2}{2}\right)}} \int_{0}^{\infty} v^{\frac{T-2}{2}} \exp\left(-\frac{vt^2}{(T-2)\delta^2}\right) \exp\left(-\frac{v}{2\theta^2}\right) dv .\]

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

\[ f(t) = \Upsilon \int_{0}^{\infty} 2x \left(\frac{1}{\theta^2} + \frac{t^2}{(T-2)\delta^2}\right)^{-1} x^{\frac{T-3}{2}} \exp(-x) dx \]

\[ = \Upsilon 2^{\frac{T-1}{2}} \left(\frac{1}{\theta^2} + \frac{t^2}{(T-2)\delta^2}\right)^{-\frac{T-1}{2}} \int_{0}^{\infty} x^{\frac{T-1}{2}-1} \exp(-x) dx .\]

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

\[ f(t) = \Upsilon 2^{\frac{T-1}{2}} \left[\frac{(T-2)\delta^2 + t^2\theta^2}{\theta^2(T-2)\delta^2}\right]^{-\frac{T-1}{2}} \frac{\Gamma\left(\frac{T-1}{2}\right)}{\Gamma\left(\frac{T-2}{2}\right)} \]

\[ = \frac{\Gamma\left(\frac{T-1}{2}\right) \theta^{\frac{T-1}{2}}}{\sqrt{2\pi(T-2)\delta(2\theta^2) \frac{T-2}{2} \Gamma\left(\frac{T-2}{2}\right)}} \frac{(T-2)\delta^2 + t^2\theta^2}{\theta^2(T-2)\delta^2}^{-\frac{T-1}{2}} \]

\[ = \frac{\Gamma\left(\frac{T-1}{2}\right) \theta^{\frac{T-1}{2}}}{\sqrt{\pi(T-2)\delta \Gamma\left(\frac{T-2}{2}\right)}} \frac{(T-2)\delta^2 + t^2\theta^2}{(T-2)\delta^2}^{-\frac{T-1}{2}} .\]
Substituting $\delta^2$ with $\frac{1-\phi_1^2}{(1-\phi_2)^2(1-\phi_1\phi_2)^2}$ and $\theta^2$ with $\frac{1}{1-\phi_1^2}$, we obtain the density

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

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

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

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

$$\mathcal{D}_{\tilde{c}^{x}_{12}} = f_w(\kappa(\tilde{c}^{x}_{12}))(\kappa'(\tilde{c}^{x}_{12})) = \Theta \left[1 + \left(\tilde{c}^{x}_{12}(1 - (\tilde{c}^{x}_{12})^2)^{-\frac{1}{2}}\right)^{2}(1-\phi_1)^2 \right]^{\frac{T-1}{2}} [1 - (\tilde{c}^{x}_{12})^2]^{-\frac{3}{2}}$$

$$= \Theta \left[1 + \frac{(\tilde{c}^{x}_{12})^2(1-\phi_1)^2}{(1 - (\tilde{c}^{x}_{12})^2)(1-\phi_1^2)}\right]^{\frac{T-1}{2}} [1 - (\tilde{c}^{x}_{12})^2]^{-\frac{3}{2}}$$

$$= \Theta \left[\frac{(1 - (\tilde{c}^{x}_{12})^2)(1-\phi_1^2) + (\tilde{c}^{x}_{12})^2(1-\phi_1)^2}{(1 - (\tilde{c}^{x}_{12})^2)(1-\phi_1^2)}\right]^{\frac{T-1}{2}} [1 - (\tilde{c}^{x}_{12})^2]^{-\frac{3}{2}}$$

$$= \Theta \left[1 - \phi_1^2 + 2(\tilde{c}^{x}_{12})^2\phi_1(\phi_1 - 1)\right]^{\frac{T-1}{2}} [1 - (\tilde{c}^{x}_{12})^2]^{-\frac{3}{2}}$$

$$= \Theta \left[1 - (\tilde{c}^{x}_{12})^2\right]^{\frac{T-1}{2}} \left[\frac{(1-\phi_1^2)}{1 - \phi_1^2 + 2(\tilde{c}^{x}_{12})^2\phi_1(\phi_1 - 1)}\right]^{\frac{T-1}{2}}.$$

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

$$\mathcal{D}_{\tilde{c}^{x}_{12}} = \frac{\Gamma(T-1/2)(1-\phi_{12})}{\Gamma(T-2)\sqrt{\pi}} \frac{[1 - (\tilde{c}^{x}_{12})^2]^{\frac{T-1}{2}} (1-\phi_1^2)^{\frac{T-1}{2}}}{[1 - \phi_1^2 + 2(\tilde{c}^{x}_{12})^2\phi_1(\phi_1 - 1)]^{\frac{T-1}{2}}}.$$

38
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

\[
y = \phi_y y_{-1} + X'_{-h} \alpha^* + \varepsilon, \tag{17}
\]
\[
X = \phi X_{-1} + U, \tag{18}
\]
\[
\varepsilon = \phi \varepsilon_{-1} + \omega \tag{19}
\]

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

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

\[
X_{-h} \varepsilon / T \xrightarrow{p} 0; \quad \phi X_{-1} U' / T \xrightarrow{p} 0; \quad \phi \varepsilon_{-1} \omega / T \xrightarrow{p} 0; \quad U_{-h} \omega / T \xrightarrow{p} 0.
\]

Let:

\[
X_{-h} X'_{-h} / T \xrightarrow{p} C_x; \quad \varepsilon' \varepsilon / T \xrightarrow{p} c_\varepsilon.
\]

Preliminary results:

\[
X_{-h} y / T \xrightarrow{p} C_x \alpha^*; \quad X_{-h} y_{-1} / T \xrightarrow{p} \phi C_x \alpha^*;
\]
\[ \varepsilon' y_{-1} \omega / T \overset{p}{\rightarrow} \phi \varepsilon c \varepsilon; \quad y'_{-1} y / T \overset{p}{\rightarrow} \phi \alpha^* C_x \alpha^* + \phi \varepsilon c \varepsilon. \]

Moreover, \( \text{var}(y) = \sigma^2_y = \alpha'^* C_x \alpha^* + c \varepsilon. \)

\[
\begin{bmatrix}
\hat{\phi}_y \\
\hat{\alpha}
\end{bmatrix} \quad \overset{p}{\rightarrow} \quad \begin{bmatrix}
y_{-1} y_{-1} / T & y'_{-1} X'_{-h} / T \\
X_{-h} y_{-1} / T & X_{-h} X'_{-h} / T
\end{bmatrix}^{-1} \begin{bmatrix}
y'_{-1} y / T \\
X_{-h} y / T
\end{bmatrix}.
\]

Applying the results above gives:

\[
\begin{bmatrix}
\hat{\phi}_y \\
\hat{\alpha}
\end{bmatrix} \quad \overset{p}{\rightarrow} \quad \begin{bmatrix}
\sigma^2_y \\
\phi C_x \alpha^*
\end{bmatrix}^{-1} \begin{bmatrix}
\phi \alpha^* C_x \alpha^* + \phi \varepsilon c \varepsilon \\
C_x \alpha^*
\end{bmatrix}.
\]

Setting \( s = \sigma^2_y - \phi^2 \alpha'^* C_x \alpha^* \), we have

\[
\begin{bmatrix}
\hat{\phi}_y \\
\hat{\alpha}
\end{bmatrix} \quad \overset{p}{\rightarrow} \quad \frac{1}{s} \begin{bmatrix}
1 & -\phi \alpha^* \\
-\phi \alpha^* & s C_x^{-1} + \phi^2 \alpha'^* \alpha'^*
\end{bmatrix} \begin{bmatrix}
\phi \alpha'^* C_x \alpha^* + \phi \varepsilon c \varepsilon \\
C_x \alpha^*
\end{bmatrix}.
\]

After some algebra we get:

\[
\begin{align*}
\hat{\phi}_y & \overset{p}{\rightarrow} \frac{\phi \varepsilon c \varepsilon}{s}, \\
\hat{\alpha} & \overset{p}{\rightarrow} \alpha^* \left( 1 - \frac{\phi \varepsilon c \varepsilon}{s} \right).
\end{align*}
\]

Considering that \( c \varepsilon = \sigma^2_y - \alpha'^* C_x \alpha^* \), and \( \alpha'^* C_x \alpha^* = R^2 \sigma^2_y \), then

\[
\begin{align*}
\hat{\phi}_y & \overset{p}{\rightarrow} \frac{\phi \varepsilon (1 - R^2)}{1 - \phi^2 R^2}, \\
\hat{\alpha} & \overset{p}{\rightarrow} \alpha^* \left( 1 - \frac{\phi \varepsilon (1 - R^2)}{1 - \phi^2 R^2} \right).
\end{align*}
\]

If we replace \( X_{-h} \) with \( U_{-h} \) in the estimated model, then we have the following preliminary
results:

\[ U_{-h}U^*_h/T \xrightarrow{P} C_u; \quad U_{-h}y/T \xrightarrow{P} C_u\alpha^*; \quad U_{-h}y_{-1}/T \xrightarrow{P} 0. \]

Thus,

\[
\begin{bmatrix}
\hat{\phi}_y \\
\hat{\alpha}
\end{bmatrix} =
\begin{bmatrix}
y_{-1}y_{-1}/T & y_{-1}U_{-h}/T \\
U_{-h}y_{-1}/T & U_{-h}U_{-h}/T
\end{bmatrix}^{-1}
\begin{bmatrix}
y_{-1}y/T \\
U_{-h}y/T
\end{bmatrix},
\]

and applying the results above gives:

\[
\begin{bmatrix}
\hat{\phi}_y \\
\hat{\alpha}
\end{bmatrix} \xrightarrow{P} \frac{1}{\sigma_y^2}
\begin{bmatrix}
1 & 0 \\
\sigma_y^2 & C_u
\end{bmatrix}^{-1}
\begin{bmatrix}
\phi\alpha^*C_\alpha^* + \phi_c\varepsilon \\
C_u\alpha^*
\end{bmatrix},
\]

from which

\[
\begin{bmatrix}
\hat{\phi}_y \\
\hat{\alpha}
\end{bmatrix} \xrightarrow{P} \frac{1}{\sigma_y^2}
\begin{bmatrix}
1 & 0 \\
\sigma_y^2 & C_u
\end{bmatrix}^{-1}
\begin{bmatrix}
\phi\alpha^*C_\alpha^* + \phi_c\varepsilon \\
C_u\alpha^*
\end{bmatrix}.
\]

After some algebra we get:

\[
\begin{aligned}
\hat{\phi}_y &\xrightarrow{P} \phi R^2 + \phi_c(1 - R^2), \\
\hat{\alpha} &\xrightarrow{P} \alpha^*.
\end{aligned}
\]

\[ \blacksquare \]

D Distribution of $b$

Consider two orthogonal Gaussian AR(1) processes generated according to the model $x_{it} = \phi_i x_{i,t-1} + u_{it}$, 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

\[ x_{2t} = \beta x_{1t} + e_t, \]  

where \(e_t \sim N(0, \sigma_e^2)\), and study the distribution of the OLS estimator \(b\) of \(\beta\) 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 \(\phi\) values obtained on 5000 Monte Carlo replications. We compare this density with that of three zero-mean Gaussian variables where the variances are respectively:

- \(S_1^2 = \frac{\hat{\sigma}_e^2}{\sum_{i=1}^{T} (x_{1t} - \bar{x}_1)^2}\), where \(\hat{\sigma}_e^2\) is the sample variance of the estimated residual \(\hat{e}_t = x_{2t} - b x_{1t}\). This is the OLS estimator for the variance of \(\beta\).

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

- \(S_3^2 = \frac{(1-\phi_1^2 \phi_2^2)(1-\phi_1^2)}{(T-1)(1-\sigma_e^2)(1-\phi_1 \phi_2)^2}\), 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 (\(\phi\)) 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.
More General Cases

We study the density of $\hat{C}_{12}$ 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 $D_{\hat{C}_{12}}$ for these cases, we rely on the densities obtained on 5000 Monte Carlo replications, i.e. $d_4(\hat{C}_{12})$, to show the effect of serial
dependence on $\Pr\{|\hat{c}_{12}^2| \geq \tau\}$.

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 $\phi_{12}$ on the density of $\hat{c}_{12}$ 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 $\text{Sign}(\phi_1) = \text{Sign}(\phi_2)$, the probability of large values of $\hat{c}_{12}$ increases with $\phi_{12}$. As a curiosity, this result is more evident for the case of Laplace variables, whereas for Cauchy and $t$-student the effect of $\phi_{12}$ declines.

![Figure 7: Simulated density of $\hat{c}_{12}$ in the case of non-Gaussian processes, for $T = 100$ and various values of $\phi$.](image)

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 $d_s(\hat{c}_{12})$ when the processes are weakly cross-correlated with $c_{12}^w = 0.2$, and when the processes are multicollinear with...
$c_{12}^u = 0.8$ (usually we refer to multicollinearity when $c_{12}^u \geq 0.7$). We observe that the impact of $\phi_{12}$ on $d_s(\hat{c}_{12}^u)$ depends on the degree of (population) cross-correlation as follows. In the case of weakly correlated processes, an increase in $\phi_{12}$ 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 $\phi_{12}$ leads to a high probability of underestimating the true population cross-correlation.

![Figure 8: $d_s(\hat{c}_{12}^u)$ obtained through simulations in the case of $c_{12}^u = 0.2$ (a) and $c_{12}^u = 0.8$ (b), for $T = 100$ and various values of $\phi$.](image)

**Density of $\hat{c}_{12}^u$ 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 $t = 1, \ldots, 100$ and $u_i \sim N(0,1)$. In Figure 9 we report the density of $\hat{c}_{12}^u$ 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 $d_s(\hat{c}_{12}^u)$ gets larger as $\phi$ increases, that is $\Pr\{|\hat{c}_{12}^u| \geq \tau\}$ increases with $|\phi|$. 

45
Figure 9: Densities of $d_s(\hat{c}_{12})$ between two uncorrelated ARMA Gaussian processes, for $T = 100$ and various values of $\phi$.

\section*{F \, $\hat{u}$-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:

\textbf{DGP 1: Equal degrees of serial dependence}, with $x_{it} = \phi_1 x_{it-1} + u_{it}$, $\varepsilon_t = \phi_\varepsilon \varepsilon_{t-1} + \omega_t$, for $i = 1, 2$, $t = 1, \ldots, T$, 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 $\hat{u}$-OLS is equivalent to the true model.

\textbf{DGP 2: Different degrees of serial dependence}, with $x_{it} = \phi_1 x_{it-1} + u_{it}$, $\varepsilon_t = \phi_\varepsilon \varepsilon_{t-1} + \omega_t$, for $i = 1, 2$, $t = 1, \ldots, T$, 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.

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

For all DGPs the \textit{i.i.d.} errors $u_{it}$ and $\omega_t$ 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 \cite{Newey: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 \cite{Cochrane: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 $t = 2, \ldots, T$. The forecasting equation is $y_t^* = \text{Proj}(y_t^* | x_{1t-1}^*, x_{2t-1}^*)$, where $y_t^* = y_t - \hat{\phi}_e y_{t-1}$, $x_{it-1}^* = x_{it-1} - \hat{\phi}_e x_{it-2}$, and $\hat{\phi}_e$ is the CO estimate of $\phi_e$.

- DynReg: the dynamic regression method \cite{Baillie: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 $x_{it-1}$, $i = 1, \ldots, n$. 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.

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

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 $||\hat{\alpha} - \alpha^*||_2$ 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 $\hat{\alpha}$-OLS have smaller and similar coefficient estimation errors. However, in terms of $R^2$, CO is outperformed by DynReg and $\hat{\alpha}$-OLS, which both include $y_{t-1}$ as predictor in their forecasting equation. We note that, while DynReg and $\hat{\alpha}$-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, $\hat{\alpha}$-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 $||\hat{\alpha} - \alpha^*||_2$ | Stat. | (a) $T = 100$ | (b) $T = 1000$ |
|-----|---------------------------------|-------|---------------|---------------|
|     | ave.                           | s.d.  | NW            | CO            | DynReg | $\hat{\alpha}$-OLS | NW | CO | DynReg | $\hat{\alpha}$-OLS |
| 1   |                                |       | NW            | CO            | DynReg | $\hat{\alpha}$-OLS | NW | CO | DynReg | $\hat{\alpha}$-OLS |
|     | 0.317                          | 0.132 | 0.747         | 0.066         | 0.351  | 0.474             | 0.317| 0.132| 0.747| 0.066| 0.351| 0.474 | 0.341| 0.040| 0.040| 0.040 | 0.040|
|     | 0.128                          | 0.069 | 0.682         | 0.056         | 0.124  | 0.126             | 0.128| 0.069| 0.682| 0.056| 0.124| 0.126 | 0.040| 0.020| 0.020| 0.020 | 0.020 |
|     | 0.129                          | 0.060 | 0.824         | 0.045         | 0.132  | 0.134             | 0.129| 0.060| 0.824| 0.045| 0.132| 0.134 | 0.037| 0.037| 0.039| 0.039 | 0.040 |
|     |                                |       |               |               | 0.747  | 0.704             | 0.071| 0.071| 0.071| 0.071| 0.747| 0.704 | 0.238| 0.020| 0.020| 0.020 | 0.021 |
|     | 0.068                          | 0.069 | 0.817         | 0.059         | 0.066  | 0.057             | 0.074| 0.066| 0.070| 0.078| 0.072| 0.054 | 0.791| 0.701| 0.888| 0.846 | 0.789| 0.701| 0.888| 0.846 |
|     |                                |       |               |               | 0.068  | 0.069             | 0.078| 0.072| 0.070| 0.078| 0.072| 0.054 | 0.791| 0.701| 0.888| 0.846 | 0.789| 0.701| 0.888| 0.846 |

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 ($\hat{\alpha}$-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 ($\hat{u}$-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 $\phi$ spurious regression occurs frequently. Note that these two results are in line with those in Granger et al. (2001). (iii) CO, DynReg and $\hat{u}$-OLS (Table 5 panel (c)-(e)) solve the problem of spurious regression due to serial dependence, by making the variance of $b$ independent of $\phi$. However, $\hat{u}$-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       |
| 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       |
| 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        |
| 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     | 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        |
| 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        |
| 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 $t_{\text{ols}}^b$ depends only on the degree of serial dependence of the processes.

**Proposition G.1:** Let $S_{\text{ols}}^2 = \frac{\hat{\sigma}_e^2}{\sum_{t=1}^T (x_{1t} - \bar{x}_t)^2}$, where $\hat{\sigma}_e^2$ is the estimated variance of the residual of model (20). Then

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

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

Note that the result in Proposition G.1 has been also derived in Granger et al. (2001). This result show that the misspecification of $t_{\text{ols}}^b$ 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 $|t_{\text{ols}}^b|$ increases with the degree of serial dependence $\phi$, but stay quite constant regardless of the sample size.
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

$$
y_t = \sum_{i=1}^{q} \alpha^*_i (\phi x_{it-2} + \phi x_{it-3} + u_{it-1}) + \sum_{j=1}^{(n-q)} \alpha^*_j (\phi x_{jt-2} + \theta u_{jt-2} + u_{jt-1}) + \phi \epsilon_{t-1} + \omega_{it}
$$

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 $\epsilon_t = \phi \epsilon_{t-1} + \phi \epsilon_{t-2} + \omega_t$ with $2|\phi| < 1$. Model (7) becomes

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

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

![Heatmaps of the variables selected by LASSO (left column) and \(\hat{u}\)-LASSO (right column) when tuning parameter is selected with CV, categorized by groups.](image)

Figure 10: Heatmaps of the variables selected by LASSO (left column) and \(\hat{u}\)-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) \(\Delta x_t\);
(3) $\Delta^2 x_t$; (4) $\log(x_t)$; (5) $\Delta \log(x_t)$; (6) $\Delta^2 \log(x_t)$. (7) $\Delta(x_t/x_{t-1} - 1.0)$.

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     |
| 9  | Ind Svy: Export Order Book Position              | EA   | ICS    | 1     |

Continued on next page
| ID | Description                                                                 | Area | Sector | Tcode |
|----|------------------------------------------------------------------------------|------|--------|-------|
| 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     |
| 35 | Svy: Mfg Of Trsp Eq - Finished Goods Inventories                              | FR   | ICS    | 1     |

Continued on next page
| ID | Description                                              | Area | Sector | Tcode |
|----|----------------------------------------------------------|------|--------|-------|
| 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     |
| 61 | Mfg: Bus Sit                                            | DE   | ICS    | 1     |

Continued on next page
| ID | Description                                      | Area | Sector | Tcode |
|----|--------------------------------------------------|------|--------|-------|
| 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     |
| 87 | Gfk Cons. Climate Svy - Bus. Cycle Expectations | DE   | CCI    | 1     |

Continued on next page
| ID | Description                                                                 | Area | Sector | Tcode |
|----|------------------------------------------------------------------------------|------|--------|-------|
| 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     |
| 113| Indl. Prod. - Manuf. Of Motor Vehicles, Trailers, Semitrailers              | FR   | IP     | 5     |

Continued on next page
| ID  | Description                                      | Area | Sector | Tcode |
|-----|------------------------------------------------|------|--------|-------|
| 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     |
| 139 | Indl. Prod. - Constr.                          | DE   | IP     | 5     |

Continued on next page
| ID  | Description                                                                 | Area | Sector | Tcode |
|-----|------------------------------------------------------------------------------|------|--------|-------|
| 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. - Cmpt, 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. - Cmpt, 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     |
| 165 | Indl. Prod. - Ind Excl Fbt                                                     | DE   | IP     | 5     |
| ID  | Description                                          | Area | Sector | Tcode |
|-----|------------------------------------------------------|------|--------|-------|
| 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     |
| 191 | Mfi Loans To Resident Private Sector                 | FR   | M&IR   | 5     |

Continued on next page
| ID  | Description                                                                 | Area | Sector  | Tcode |
|-----|------------------------------------------------------------------------------|------|---------|-------|
| 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     |
| 217 | All-Items Hicp                                                               | ES   | HCPI    | 6     |

Continued on next page
| ID  | Description                                      | Area | Sector | Tcode |
|-----|--------------------------------------------------|------|--------|-------|
| 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     |
| 243 | Ppi: Int. Goods                                  | EA   | PPI    | 6     |

Continued on next page
| ID  | Description                                                                 | Area | Sector | Tcode |
|-----|-----------------------------------------------------------------------------|------|--------|-------|
| 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, Househd & 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     |
| 269 | Fr Hur Hommes (Ages 15-24)                                                  | FR   | HUR    | 5     |

Continued on next page
| ID | Description                        | Area | Sector | Tcode |
|----|-----------------------------------|------|--------|-------|
| 270| Fr Hur Hommes (All Ages)          | FR   | HUR    | 5     |
| 271| Fr Hur All Persons (Ages 15-24)   | FR   | HUR    | 5     |
| 272| Fr Hur All 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 Hur All 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 Hur All 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     |
| 295| Es Hur All Persons (Ages 25 And Over) | ES | HUR   | 5     |

Continued on next page
| ID  | Description                                      | Area | Sector | Tcode |
|-----|--------------------------------------------------|------|--------|-------|
| 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 $N(0, \frac{1}{T-1})$. 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 $\frac{K_0(|xy|)}{\pi y}$, where $K_0(|xy|)$ is the Bessel function of the second kind.
Figure 11: Density of $\hat{\text{Cov}}(u_1, u_2)$ between two uncorrelated standard Normal variables for $T = 10$ (a) and $T = 100$ (b).