

INSTITUTE
OF ECONOMICS



Scuola Superiore
Sant'Anna

LEM | Laboratory of Economics and Management

Institute of Economics
Scuola Superiore Sant'Anna

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

LEM

WORKING PAPER SERIES

On the impact of serial dependence on penalized regression methods

Simone Tonini ^a
Francesca Chiaromonte ^{a,b}
Alessandro Giovannelli ^c

^a Institute of Economics & EMbeDS, Sant'Anna School of Advanced Studies, Pisa, Italy.

^b Department of Statistics, Penn State University, USA.

^c Department of Information Engineering, Computer Science and Mathematics,
University of L'Aquila, Italy.

2022/21

July 2022

ISSN(ONLINE) 2284-0400

On the impact of serial dependence on penalized regression methods

Simone Tonini ^{*1}, Francesca Chiaromonte^{2, 3}, and Alessandro Giovannelli⁴

¹Institute of Economics, Sant'Anna School of Advanced Studies, Pisa, Italy.

²Institute of Economics & EMbeDS, Sant'Anna School of Advanced Studies, Pisa, Italy.

³Department of Statistics, Penn State University, USA.

⁴University of L'Aquila, Department of Information Engineering, Computer Science and Mathematics, L'Aquila, Italy.

Abstract. This paper characterizes the impact of serial dependence on the non-asymptotic estimation error bound of penalized regressions (PRs). Focusing on the direct relationship between the degree of cross-correlation of covariates and the estimation error bound of PRs, we show that orthogonal or weakly cross-correlated stationary AR processes can exhibit high spurious cross-correlations caused by serial dependence. In this respect, we study analytically the density of sample cross-correlations in the simplest case of two orthogonal Gaussian AR(1) processes. Simulations show that our results can be extended to the general case of weakly cross-correlated non Gaussian AR processes of any autoregressive order. To improve the estimation performance of PRs in a time series regime, we propose an approach based on applying PRs to the residuals of ARMA models fit on the observed time series. We show that under mild assumptions the proposed approach allows us both to reduce the estimation error and to develop an effective forecasting strategy. The estimation accuracy of our proposal is numerically evaluated through simulations. To assess the effectiveness of the forecasting strategy, we provide the results of an empirical application to monthly macroeconomic data relative to the Euro Area economy.

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

JEL Classification: C13, C18, C22, C31, C46.

*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., 2013). In the context of high dimensional regression problems, with a number of predictors comparable to or larger than the sample size, coefficient estimates produced by ordinary least squares (OLS) are affected by unacceptable variance, or even numerically unstable or undetermined. Among the approaches proposed to tackle this issue are penalized regression methods (PRs) which reduce (or shrink) standard OLS regression coefficients towards zero. This reduces their variance at the price of introducing some bias, but if the penalty is properly tuned, overall, estimation error improves. In particular, we refer to the most commonly used PRs, namely, those based on ℓ_1 -penalty (Tibshirani, 1996; Zou, 2006), ℓ_2 -penalty (Hoerl and Kennard, 1970) and their combinations (Zou and Hastie, 2005). Depending on the form of the penalty, PRs can produce *dense solutions*, where coefficients may have small yet non-zero estimates, or *sparse solutions*, where less relevant predictors have coefficient estimates equal to zero. Thus, PRs that provide sparse solutions (sparse PRs) in effect select a subset of predictors relevant for the regression.

Zhao and Yu (2006); Bickel et al. (2009); Lounici et al. (2009); Negahban et al. (2009, 2012); Hastie et al. (2015); Zou and Zhang (2009); Xin et al. (2017) studied the estimation properties of PRs and showed that their non-asymptotic estimation error bound depends critically on the degree of cross-correlation between covariates. In particular, PRs perform best with orthogonal or weakly cross-correlated covariates, since the bound is inversely proportional to the minimum eigenvalue of the sample cross-correlation matrix of the covariates themselves. In this respect we can distinguish two different situations; that of multicollinear covariates, where cross-correlations exists at the population level, and that of spurious correlations, where covariates may be orthogonal or nearly orthogonal at the population level but other mechanisms in the way observations are collected generate cross-correlations in the sample. Unfortunately, spurious correlations become more prevalent the higher the dimension of the data (Fan and Zhou, 2016; Fan et al., 2018) and can lead to false scientific discoveries and wrong statistical inferences (Fan et al., 2013).

In the context of time series, and especially of economic and financial time series, the case of multicollinear covariates has been extensively studied. In particular, Stock and Watson (2002a,b); Forni et al. (2000, 2003, 2016, 2018) proposed approximated dynamic factor models (DFM) as a means

to produce effective forecasts with highly cross-correlated covariates. De Mol et al. (2008); Giannone et al. (2018) showed that multicollinearity induced by latent factors leads to an approximation of the factor space with sparse PRs, producing forecasts similar to those of DFMs, but noted that the selection of covariates is not stable over time. Fan et al. (2020) proposed a consistent strategy for sparse PRs when covariates can be effectively de-correlated via a few pervasive latent factors. Medeiros and Mendes (2012) studied the covariate selection consistency of sparse PRs in the presence of multicollinear time series. Finally, in an exhaustive simulation analysis, Smeekes and Wijler (2018) showed that sparse PRs improve forecasts over traditional approaches when applied to non-stationary and co-integrated data despite, again, poor performance in terms of covariate selection.

In this paper we focus on the issue of spurious correlations, as already described in Fan et al. (2013); Fan and Zhou (2016); Fan et al. (2018), extending its treatment to the time series framework. The econometric literature has dealt with the issue of autocorrelated errors, referred to as *spurious regression*, with seminal papers by Box and Newbold (1971) and Granger and Newbold (1974). Box and Newbold (1971) pointed out the risk of obtaining a spurious model if sufficient care is not placed on an appropriate formulation for the autocorrelation structure of the errors in the regression equation. Using simulations, Granger and Newbold (1974) studied the case of two independent drift-free random walks showing that, as a consequence of autocorrelated regression errors, the usual significance tests on the regression coefficients are invalid. In particular, they concluded that if the residuals are strongly autocorrelated (low Durbin-Watson value), then the regression equation is misspecified, whatever the observed coefficient of determination (R^2).

In time series spurious correlations are well known to occur with pairs of independent unit root processes, not least is the case of two stationary processes. Bartlett (1935) observed that the variance of the sample correlation between two orthogonal Gaussian AR(1) processes depends on both the sample size and the degree of serial dependence. In this regard, McGregor (1962) provided the approximate null distribution of the sample correlation when the series are stationary Markov type. Granger et al. (2001) explored the possible existence of spurious regression between a pair of independent stationary series through both theoretical and simulation results, and provided the limiting distribution of the t -statistic for the linear regression coefficient between two stationary Gaussian AR(1) processes, showing that the variance of such limiting distribution increases with the degree of serial dependence.

With most of the existing literature focused on the effects of serially dependent errors for the OLS estimator, our aim is to broaden the picture by showing that the serial dependence of covariates

also entails problems for estimating regression coefficients. Our main theoretical contribution is to show the effect of such serial dependence on the minimum eigenvalue of the sample correlation matrix of the covariates, which is one of the major determinants of the non-asymptotic estimation error bound of PRs. Specifically, we demonstrate how the probability of spurious cross-correlation between stationary Gaussian AR processes depends not only on the sample size, but also on the degree of serial dependence. To achieve this result we derived the density of the sample cross-correlation between orthogonal stationary Gaussian AR(1) processes, adapting the treatment in Anderson (2003) to time series. Since the theoretical upper bound of the minimum eigenvalue of the sample correlation matrix of the covariates decreases as the maximum absolute value of the diagonal grows, this finding for a single generic sample correlation can be leveraged in settings with more than two covariates.

Beyond our theoretical results, we use Monte Carlo simulations to study the impact of serial dependence when covariates are not Gaussian, when they do show cross-correlations at the population level, and when they are generated by more complex models than AR(1). We also consider the case where covariates are generated by an approximate factor model with an autocorrelated idiosyncratic component.

Our main results can be summarized as follows: (i) through our theoretical density, we show that, whenever the autocorrelation coefficients of the 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-correlation; (ii) through simulations, we show that the association between serial dependence and the probability of large spurious correlation holds much more generally, e.g., in cases where the processes are not Gaussian, weakly correlated or generated by models different than AR(1). These results highlight that a small minimum eigenvalue is more likely in finite realizations of serially dependent weakly 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.

To mitigate the adverse effects of covariates serial dependence, we propose to apply PRs on the residuals of ARMA models fitted on the observed time series. We show that, under mild assumptions, this produces better estimates of regression coefficients, as well as an improved selection of the relevant ones in sparse regimes. However, our procedure does involve a cost; namely, it hinders our ability to estimate regression parameters relative to past values of the target variable. To illustrate our rationale we provide simulations where we apply the standard OLS estimator to residuals of the ARMA processes, and compare it with some of the best known methods for addressing serial dependence in

regression. We then apply LASSO (that is the most representative ℓ_1 -penalty based PR) to ARMA residuals, evaluating our proposal both on simulated data and on monthly macroeconomic data relative to the Euro Area economy.

The remainder 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 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, i.e. the application of PRs to ARMA residuals, using simulated and actual econometric 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 $\mathbf{a} = (a_1, a_2, \dots, a_p)'$. $\text{Supp}(\mathbf{a})$ denotes the support of a vector; that is, $\{j \in \{1, 2, \dots, p\} : a_j \neq 0\}$, and $|\text{Supp}(\mathbf{a})|$ the support cardinality. The ℓ_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} . 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 concerning 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 (Ω, \mathcal{F}, P) . Further assume that y_t has finite mean and that \mathbf{x}_t has mean vector $\mathbf{0}$ and finite second order moments. Let $\mathbf{C}_k^x = E[\mathbf{x}_t \mathbf{x}'_{t-k}]$, $k \geq 0$ indicate the generic lag k covariance matrix of the covariates, and $\widehat{\mathbf{C}}_0^x = \frac{1}{T-1} \mathbf{X} \mathbf{X}'$ an estimate of the cross-covariance matrix, with generic element \widehat{c}_{ij}^x and eigenvalues $\widehat{\psi}_{max}^x \geq \dots \geq \widehat{\psi}_{min}^x$. Finally, assume that each \mathbf{x}_i has been 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 variable

$$\mathbf{y} = \mathbf{X}' \boldsymbol{\alpha} + \boldsymbol{\varepsilon}, \quad (1)$$

where $\boldsymbol{\alpha}$ is the $n \times 1$ unknown s -sparse vector of regression coefficients, i.e. $\|\boldsymbol{\alpha}\|_0 = s < n$, and $\boldsymbol{\varepsilon} \in \mathbb{R}^T$ is a random noise vector. Note that in equation (1), for simplicity of exposition, at each time y_t depends only on \mathbf{x}_t (in standard time series literature y_t is allowed to depend on lagged vector(s) \mathbf{x}_{t-l} , $l = 1, \dots$). If n is comparable to or larger than T , $\boldsymbol{\alpha}$ is estimated solving a convex optimization problem where a quadratic loss function is combined with a regularization penalty:

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

Here $\lambda > 0$ represents the weight of the penalty, and $\ell : \mathbb{R}^n \rightarrow \mathbb{R}^+$ is a norm. We note that this setup allows for misspecified models as well. The following definitions will be used in our developments.

Definition 1 (*Strong Convexity*): Given a differentiable function $\mathcal{L} : \mathbb{R}^n \rightarrow \mathbb{R}$ and the vector differential operator ∇ , we say that \mathcal{L} is strongly convex with parameter $\gamma > 0$ at $\mathbf{a} \in \mathbb{R}^n$ if the inequality

$$\mathcal{L}(\mathbf{b}) - \mathcal{L}(\mathbf{a}) \geq \nabla \mathcal{L}(\mathbf{a})'(\mathbf{b} - \mathbf{a}) + \frac{\gamma}{2} \|\mathbf{b} - \mathbf{a}\|_2^2$$

holds for all $\mathbf{b} \in \mathbb{R}^n$.

Strong Convexity (see Negahban et al. 2009, 2012) guarantees a small coefficient estimation error. In particular, when \mathcal{L} (the loss function) is “sharply curved” around its optimum $\hat{\boldsymbol{\alpha}}$, a small $|\mathcal{L}(\boldsymbol{\alpha}) - \mathcal{L}(\hat{\boldsymbol{\alpha}})|$ guarantees that $\|\boldsymbol{\alpha} - \hat{\boldsymbol{\alpha}}\|_2$ is also small. The parameter γ governs the strength of convexity; when \mathcal{L} is twice differentiable, strong convexity requires the minimum eigenvalue of the Hessian $\nabla^2 \mathcal{L}(\boldsymbol{\alpha})$ to be at least γ for all $\boldsymbol{\alpha}$ in a neighborhood of $\hat{\boldsymbol{\alpha}}$. Thus, since its Hessian is $\nabla^2 \mathcal{L}(\boldsymbol{\alpha}) = \widehat{\mathbf{C}}_0^x$, the quadratic loss $\mathcal{L}(\boldsymbol{\alpha}) = \frac{1}{2T} \|\mathbf{y} - \mathbf{X}'\boldsymbol{\alpha}\|_2^2$ is strongly convex with parameter γ if and only if $\hat{\psi}_{\min} \geq \gamma$ (see Hastie et al. 2015, p. 293). Consequently, in this case $\|\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}\|_2$ depends on $\hat{\psi}_{\min}$. It is relevant to note that when $n > T$ the quadratic loss cannot be strongly convex since $\widehat{\mathbf{C}}_0^x$ 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 $\widehat{\mathbf{C}}_0^x$ as a function of the degree of sparsity, s . The Restricted Eigenvalue Condition allows for strong convexity (Definition 1) to hold in the singular case, and we refer to this as *Restricted Strong Convexity* (see Negahban et al. 2012; we provide more details in Supplement A).

Definition 2 (*Dual Norm and Subspace Compatibility Constant*): Given a norm ℓ and the inner product $\langle \cdot, \cdot \rangle$, we define the dual norm of ℓ as

$$\ell^*(\mathbf{v}) := \sup_{\mathbf{u} \in \mathbb{R}^n \setminus \{0\}} \frac{\langle \mathbf{u}, \mathbf{v} \rangle}{\ell(\mathbf{u})}.$$

For any subspace \mathcal{A} of \mathbb{R}^n that captures the constraints underlying (2), we define the subspace compatibility constant with respect to the pair $(\ell, \|\cdot\|_2)$ as

$$\Psi(\mathcal{A}) := \sup_{\mathbf{u} \in \mathcal{A}: \mathbf{u} \neq 0} \frac{\ell(\mathbf{u})}{\|\mathbf{u}\|_2}.$$

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

Proposition 1 *Consider the convex optimization problem in (2). If the penalty parameter λ is strictly positive and $\geq 2\ell^* (\frac{1}{T}\mathbf{X}'\boldsymbol{\varepsilon})$, and strong convexity holds with parameter $\gamma > 0$. Then, any optimal solution $\hat{\boldsymbol{\alpha}}$ satisfies the bound*

$$\|\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}\|_2 \leq 3\frac{\lambda}{\gamma}\Psi(\mathcal{A}). \quad (3)$$

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

The coefficient estimation error bound in Proposition 1 increases with the penalty parameter λ , which must be strictly positive and satisfy the lower bound $2\ell^* (\frac{1}{T}\mathbf{X}'\boldsymbol{\varepsilon})$; increases with the subspace compatibility constant $\Psi(\mathcal{A})$, which in turn increases with the size of the model subspace \mathcal{A} ; and decreases with the convexity parameter γ . Negahban et al. (2009, 2012) derive the bound for PRs in the case of independent observations (no serial dependence). To this end, the authors compute the probability that $\lambda \geq 2\ell^* (\frac{1}{T}\mathbf{X}'\boldsymbol{\varepsilon})$ under the assumption that the entries of \mathbf{X} and $\boldsymbol{\varepsilon}$ are sub-Gaussian, and assume that strong convexity (or restricted strong convexity) holds with parameter γ , i.e. that $\hat{\psi}_{min}^x \geq \gamma$ (see Corollary 2 in Negahban et al. 2012 for an example of sparse PR estimation error bound, and Corollary 6 in Negahban et al. 2009 for an example of dense PR estimation error bound).

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 high sample cross-correlations correspond to small $\hat{\psi}_{min}^x$. As mentioned in the Introduction, high sample cross-correlations may be due to the presence of true, population-level multicollinearities. 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 (see Supplement B for more details). However, high 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 serial dependence on $\widehat{\psi}_{min}^x$, which determines the “strength” of strong convexity (see Definition 1), that is one of the main components of the PRs error bound presented in Proposition 1. In this respect, we relax the assumption that strong convexity (or restricted strong convexity) holds with parameter $\gamma > 0$, and show that the probability of getting $\widehat{\psi}_{min}^x \leq \gamma$ increases as the degree of serial dependence grows. Note that in order to focus on $\widehat{\psi}_{min}^x$ we are assuming that $\widehat{\mathbf{C}}_0^x$ is positive definite. This choice is motivated by the wish to simplify our treatment. If $n > T$ and the matrix $\widehat{\mathbf{C}}_0^x$ is singular, we can replicate our arguments considering the probability that the restricted eigenvalue is $\leq \gamma$ (see Bickel et al. 2009).

The logical structure behind our theoretical contribution is as follows. Given $\gamma = 1 - \tau$, $\tau \in [0, 1)$, and the upper bound $\widehat{\psi}_{min}^x \leq 1 - \max_{i \neq j} |\widehat{c}_{ij}^x|$, we emphasize the role of a generic off-diagonal element of the matrix $\widehat{\mathbf{C}}_0^x$ in determining the probability that $\widehat{\psi}_{min}^x \leq \gamma$ through the inequalities

$$\Pr \left\{ \widehat{\psi}_{min}^x \leq 1 - \tau \right\} \geq \Pr \left\{ 1 - \max_{i \neq j} |\widehat{c}_{ij}^x| \leq 1 - \tau \right\} \geq \Pr \left\{ 1 - |\widehat{c}_{i \neq j}^x| \leq 1 - \tau \right\} = \Pr \left\{ |\widehat{c}_{i \neq j}^x| \geq \tau \right\}. \quad (4)$$

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

To better illustrate our reasoning, we introduce a toy example where we show numerically the impact of serial dependence on $\max_{i \neq j} |\widehat{c}_{ij}^x|$ and $\widehat{\psi}_{min}^x$. We generate 10 processes from the model $\mathbf{x}_t = D_\phi \mathbf{x}_{t-1} + \mathbf{u}_t$, $t = 1, \dots, 100$, where D_ϕ is a 10×10 diagonal matrix with the same autocorrelation coefficient ϕ in all positions along the main diagonal, and $\mathbf{u}_t \sim N(\mathbf{0}_{10}, \mathbf{I}_{10})$. Note that for these AR(1) processes the the degree of serial dependence is determined by $|\phi|$ and, since the processes are orthogonal, the minimum eigenvalue of the population cross-correlation matrix \mathbf{C}_0^x is $\psi_{min}^x = 1$. We consider five values for ϕ , namely 0.0, 0.3, 0.6, 0.9, 0.95, and for each we calculate the average and the standard deviation of both $\max_{i \neq j} |\widehat{c}_{ij}^x|$ and $\widehat{\psi}_{min}^x$ on 5000 Monte Carlo replications. Results are reported in Figure 1. We see that the stronger the persistence of the process (ϕ closer to 1) the higher is the probability of a large spurious sample correlation (orange circle), which in turn leads to a small minimum eigenvalue of the sample cross-correlation matrix (blue triangle), as a consequence of (4).

In light of these results, our next task is to derive the finite sample density of \widehat{c}_{ij}^x for the purpose

of formalizing the impact of serial dependence on $\Pr\{|\widehat{c}_{i \neq j}^x| \geq \tau\}$. It is noteworthy that when the covariates have a Factor-based DGP (as in Supplement B) high spurious cross-correlation 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).

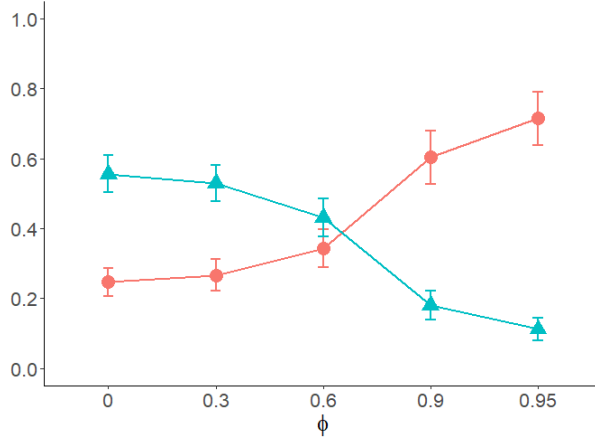


Figure 1: Results from a numerical toy example. The orange circles and bars represent means and standard deviations of $\max_{i \neq j} |\widehat{c}_{ij}^x|$ for various values of the autocorrelation ϕ , as obtained from 5000 Monte Carlo replications. Similarly, blue triangles and bars represent means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

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

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

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

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

Therefore $\mathbf{x}_t \sim N(\mathbf{0}_2, \mathbf{C}_0^x)$ with $(\mathbf{C}_0^x)_{ii} = \frac{1}{1-\phi_i^2}$, and $(\mathbf{C}_0^x)_{12} = c_{12}^x = 0$. In this setting we focus on the density of the sample correlation coefficient defined as

$$\widehat{c}_{12}^x = \frac{a_{12}}{\sqrt{a_{11}}\sqrt{a_{22}}}, \quad (5)$$

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 $\bar{x}_i = 0$, $i, j = 1, 2$. We generalize the approach of Anderson (2003) to our time series context. In particular, when $c_{12}^u = 0$, $b = a_{21}/a_{11}$ and $v = a_{22} - a_{21}^2/a_{11}$, Anderson (2003, p. 119) shows that

$$\frac{\sqrt{a_{11}} b}{\sqrt{v/(T-2)}} = \sqrt{T-2} \frac{a_{12}/\sqrt{a_{11}a_{22}}}{\sqrt{1 - a_{12}^2/(a_{11}a_{22})}} = \sqrt{T-2} \frac{\widehat{c}_{12}^x}{\sqrt{1 - (\widehat{c}_{12}^x)^2}}. \quad (6)$$

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

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

Sample Distribution of b . We start by deriving the sample distribution of b , the OLS regression coefficient for x_2 on x_1 . This is an arbitrary choice, in fact regressing x_1 on x_2 would not change the distribution we will eventually obtain for the sample correlation coefficient.

Proposition 2 *Under Assumption 1 the sample distribution of b is approximately*

$$N\left(0, \frac{(1 - \phi_1^2 \phi_2^2)(1 - \phi_1^2)}{(T-1)(1 - \phi_2^2)(1 - \phi_1 \phi_2)^2}\right).$$

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

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

where $\widehat{Cov}(u_{i[-l]}, u_j) = \sum_{t=l+1}^T (u_{it-l} - \bar{u}_i)(u_{jt} - \bar{u}_j)/(T-l-1)$, for $i \neq j = 1, 2$, $\bar{u}_i = \frac{1}{T-l-1} \sum_{t=l+1}^T u_{it-l}$ 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 C)

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

Moreover, the quantity

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

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

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

and

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

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

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

■

Remark 2 For $T < \infty$ the quantity η_{12} has a variance that increases with the degree of serial dependence. This quantity strongly affects the impact of the degree of serial dependence on the variance of a_{12} and, as a consequence, on the variance of both $\widehat{Cov}(x_1, x_2)$ and b .

Proposition 2 shows that the OLS estimate b is normally distributed with a variance that strongly depends on the degrees of 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, it has been recently shown that 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.

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 η_{12} that impacts the sample covariance between orthogonal $AR(1)$ processes.

Sample Distribution of v . 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.

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: Consider a $(T-1) \times (T-1)$ orthogonal matrix M with first row $\mathbf{x}'_1/\sqrt{a_{11}}$ and let $s_t = \sum_{h=1}^{T-1} d_{th}x_{2h}$, $t = 1, \dots, T$, $h = 1, \dots, T$. 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

$$v = \sum_{t=2}^{T-1} s_t^2 \approx \sum_{t=2}^{T-1} \frac{z_t^2}{1-\phi_2^2}. \quad (7)$$

The right side of (7) is a Gamma distribution with shape parameter $\frac{T-2}{2}$ and rate parameter $\frac{2}{1-\phi_2^2}$. ■

Sample Density of \widehat{c}_{12}^x . Note that b and v are independent. Using Propositions 2 and 3 and equation (6) we can now derive the density associated to the sample distribution of \widehat{c}_{12}^x .

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

$$\mathcal{D}_{\widehat{c}_{12}^x} = \frac{\Gamma\left(\frac{T-1}{2}\right)(1-\phi_{12})}{\Gamma\left(\frac{T-2}{2}\right)\sqrt{\pi}} (1 - (\widehat{c}_{12}^x)^2)^{\frac{T-4}{2}} (1 - \phi_{12}^2)^{\frac{T-2}{2}} \left(\frac{1}{1 - \phi_{12}^2 + 2(\widehat{c}_{12}^x)^2\phi_{12}(\phi_{12} - 1)} \right)^{\frac{T-1}{2}}. \quad (8)$$

Proof: Because of Proposition 2, $\sqrt{a_{11}}b$ is approximately $N\left(0, \frac{1-\phi_1^2\phi_2^2}{(1-\phi_2^2)(1-\phi_1\phi_2)^2}\right)$. Let $\delta^2 = \frac{1-\phi_1^2\phi_2^2}{(1-\phi_2^2)(1-\phi_1\phi_2)^2}$, $\theta^2 = \frac{1}{1-\phi_2^2}$ and $t = \frac{\sqrt{a_{11}}b}{\sqrt{v/(T-2)}}$. In the remainder 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

$$g(\sqrt{a_{11}}b) = \frac{1}{\delta\sqrt{2\pi}} e^{-\frac{a_{11}b^2}{2\delta^2}}, \quad (9)$$

$$h(v) = \frac{1}{(2\theta^2)^{\frac{T-2}{2}} \Gamma\left(\frac{T-2}{2}\right)} v^{\frac{T-2}{2}-1} e^{-\frac{v}{2\theta^2}}. \quad (10)$$

We focus on

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

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

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

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

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

Thus we obtain

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

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

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

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

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

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

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

Thus, the (finite) sample density of \widehat{c}_{12}^x , taking the densities in (9) and (10) as exact, is

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

Remark 3 $\mathcal{D}_{\widehat{c}_{12}^x}$ is the density of the sample correlation coefficient (5) based on a finite T , with serial dependence expressed by ϕ_1 and ϕ_2 , and under the assumption of orthogonal Gaussian AR(1) processes.

Remark 4 From (8) we see that ϕ_{12} determines the density of \widehat{c}_{12}^x 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 cross-correlation between orthogonal Gaussian Autoregressive processes heavily depends on the degree of serial dependence. This has important consequences on the non-asymptotic performance of PRs for the reasons that we pointed out at the beginning of Section 2.2, related to the role of $\Pr\{|\widehat{c}_{12}^x| \geq \tau\}$ (see inequality (4), Definition 1 and Proposition 1). The implication of Theorem 1 for such probability can be summarized in the following remark.

Remark 5 *Because of Theorem 1*

$$\Pr \{|\widehat{c}_{12}^x| \geq \tau\} \approx \int_{-1}^{-\tau} \mathcal{D}_{\widehat{c}_{12}^x} d\widehat{c}_{12}^x + \int_{\tau}^1 \mathcal{D}_{\widehat{c}_{12}^x} d\widehat{c}_{12}^x \quad (11)$$

depends on the degrees of serial dependence of the processes.

4 Monte Carlo Experiments

Here we first conduct Monte Carlo experiments to assess numerically the approximation of the density of \widehat{c}_{12}^x described in the previous Section. Next, we expand the theoretical results in more generic contexts, relaxing the assumption that the covariates are orthogonal Gaussian AR(1) processes. To simplify matters, henceforth we indicate the density of \widehat{c}_{12}^x obtained by simulations as $d_s(\widehat{c}_{12}^x)$.

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

We generate data from the bivariate process $\mathbf{x}_t = D_\phi \mathbf{x}_{t-1} + \mathbf{u}_t$ for $t = 1, \dots, T$, where D_ϕ is a 2×2 diagonal matrix with same autocorrelation coefficient ϕ in both position along the diagonal, and $u_t \sim N(\mathbf{0}_2, \mathbf{I}_2)$. We consider $T = 50, 100, 250$ and $\phi = 0.3, 0.6, 0.9, 0.95$ – thus, the parameter ϕ_{12} in $\mathcal{D}_{\widehat{c}_{12}^x}$, here equal to ϕ^2 , takes values 0.09, 0.36, 0.81, 0.90. The left panels of Figures 2-4 show, for various values of T and ϕ_{12} , the density $d_s(\widehat{c}_{12}^x)$ generated through 5000 Monte Carlo replications. The right panels of Figures 2-4 (b) show the corresponding $\mathcal{D}_{\widehat{c}_{12}^x}$. These were plotted using 5000 values of the argument starting at -1 and increasing by steps of size 0.0004 until 1. Unsurprisingly, we observe that the approximation of $d_s(\widehat{c}_{12}^x)$ to $\mathcal{D}_{\widehat{c}_{12}^x}$ increases as T grows and/or ϕ_{12} decreases. In particular, in Figure 2, where $T = 50$, $\mathcal{D}_{\widehat{c}_{12}^x}$ approximates well $d_s(\widehat{c}_{12}^x)$ for a low-to-intermediate degree of serial dependence ($\phi_{12} \leq 0.36$, i.e. $\phi \leq 0.6$). In contrast, in cases with high degree of serial dependence ($\phi_{12} \geq 0.81$, i.e. $\phi \geq 0.9$), $\mathcal{D}_{\widehat{c}_{12}^x}$ has larger tails compared to $d_s(\widehat{c}_{12}^x)$; that is, it over-estimates the probability of large spurious correlations. However, it is noteworthy that the difference between the two densities is negligible for $T \geq 100$ (Figures 3 and 4), also with high degree of serial dependence ($\phi_{12} = 0.90$, i.e. $\phi = 0.95$). These numerical experiments corroborate the fact that the sample cross-correlation between orthogonal Gaussian AR(1) processes is affected by the degree of serial dependence in a way that is well approximated by $\mathcal{D}_{\widehat{c}_{12}^x}$. In fact, for a sufficiently large finite T , we observe that $\Pr \{|\widehat{c}_{12}^x| \geq \tau\}$, $\tau > 0$, increases with ϕ_{12} in a similar way for $d_s(\widehat{c}_{12}^x)$ and $\mathcal{D}_{\widehat{c}_{12}^x}$.

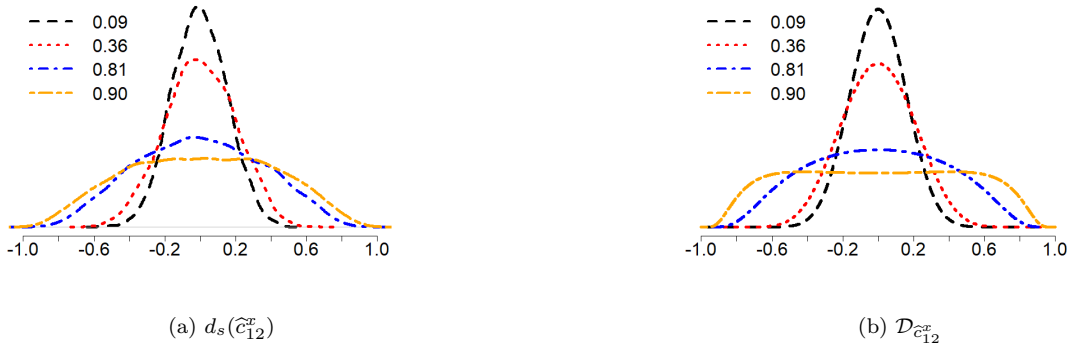


Figure 2: Densities of \hat{c}_{12}^x (a), and corresponding $\mathcal{D}_{\hat{c}_{12}^x}$ (b), for $T=50$ and various values of ϕ_{12} .

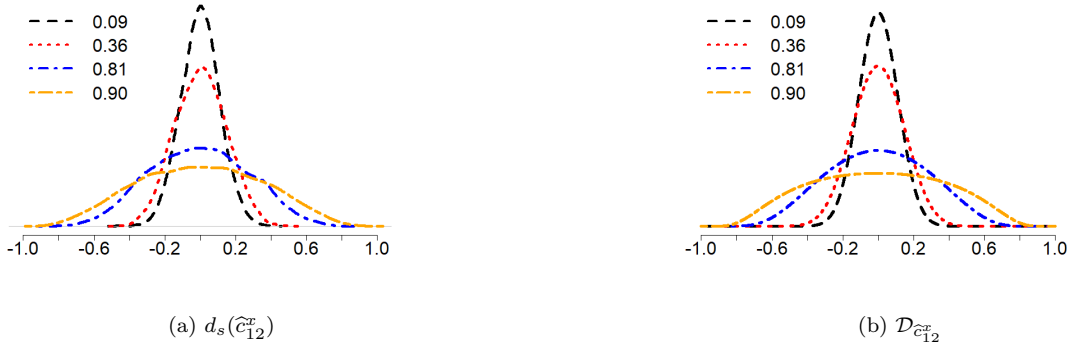


Figure 3: Densities of \hat{c}_{12}^x (a), and corresponding $\mathcal{D}_{\hat{c}_{12}^x}$ (b), for $T=100$ and various values of ϕ_{12} .



Figure 4: Densities of \hat{c}_{12}^x (a), and corresponding $\mathcal{D}_{\hat{c}_{12}^x}$ (b), for $T=250$ and various values of ϕ_{12} .

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

In Remark 4 we pointed out that the impact of ϕ_{12} on $\mathcal{D}_{\hat{c}_{12}^x}$ depends on $Sign(\phi_{12})$. In particular, when $-1 < \phi_{12} < 0$ an increment on $|\phi_{12}|$ makes the density of \hat{c}_{12}^x more concentrated around 0. In order to validate this result numerically, we run simulations with $T = 100$ and different values for the second element of the diagonal of D_ϕ ; namely, $-0.3, -0.6, -0.9, -0.95$. Results are shown in Figure 5. Indeed, we see that when $Sign(\phi_1) \neq Sign(\phi_2)$ and $|\phi_{12}|$ increases, $d_s(\hat{c}_{12}^x)$ increases its concentration around 0 in a way that is, again, well approximated by $\mathcal{D}_{\hat{c}_{12}^x}$.

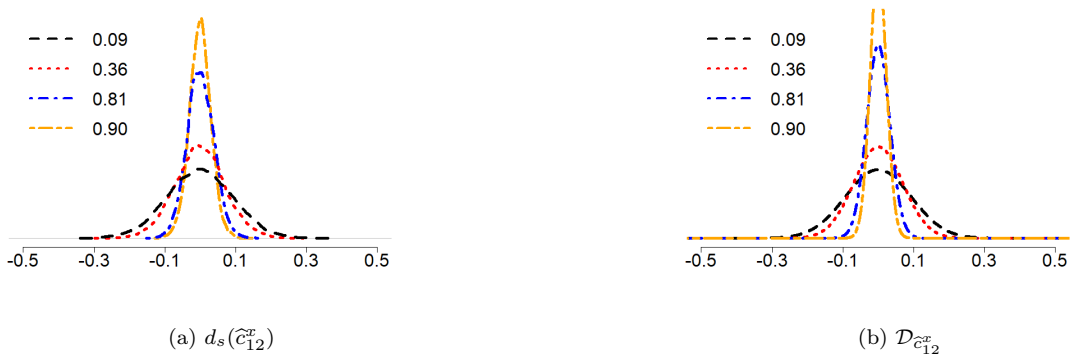


Figure 5: Densities of \hat{c}_{12}^x (a), and corresponding $\mathcal{D}_{\hat{c}_{12}^x}$ (b), for $T=100$ and various (negative) values of ϕ_{12} .

4.2 General Case

To generalize our findings to the case of non-Gaussian weakly correlated ARMA and AR processes, we generate covariates according to the following DGPs:

$$x_{1t} = (\phi + 0.1)x_{1t-1} + (\phi + 0.1)x_{1t-2} - 0.2x_{1t-3} + u_{1t} \quad (12)$$

$$x_{2t} = \phi x_{2t-1} + \phi x_{2t-2} + u_{2t} + 0.8u_{2t-1} \quad (13)$$

where $t = 1, \dots, 100$ and $\phi = 0.15, 0.3, 0.45, 0.475$. Moreover, we generate u_{1t} and u_{2t} from a bivariate Laplace distribution with means 0, variances 1, and $c_{12}^u = 0.2$. In these more general cases, we do not know an approximate theoretical density for \hat{c}_{12}^x . Therefore, we rely entirely on simulations to show the effect of serial dependence on $\Pr\{|\hat{c}_{12}^x| \geq \tau\}$. Figure 6 shows $d_s(\hat{c}_{12}^x)$ obtained from 5000 Monte Carlo replications for the different values of ϕ . In short, also in the more general cases where covariates are non Gaussian, weakly correlated AR(3) and ARMA(2,1) processes, the probability of getting large sample cross-correlations depends on the degree of serial dependence. More simulation

results are provided in Supplement E.

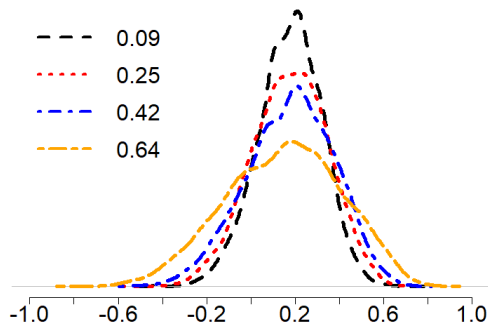


Figure 6: $d_s(\hat{c}_{12}^x)$ in the case of non Gaussian weakly correlated AR(3) and ARMA(2,1) processes, for various values of ϕ .

5 A Remedy for Serial Dependence-Induced Spurious Correlation

In this section we propose a solution to 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 at the first step instead of the original covariates. More precisely, let x_{it} (the i -th time series at time t) be generated by the model

$$x_{it} = \sum_{l=1}^{p_i} \phi_{il} x_{it-l} + \sum_{k=1}^{q_i} \theta_{ik} u_{it-k} + u_{it} \quad (14)$$

where $i = 1, \dots, n, t = 1, \dots, T$. This describes an ARMA(p_i, q_i) process where p_i is the order of autocorrelation, which determines the order of the weighted moving average over past values of the covariate, and q_i is the order of the weighted moving average over past errors. Note that the AR (i.e. $p_i \geq 1, q_i = 0$) and MA (i.e. $p_i = 0, q_i \geq 1$) models are special cases of (14). For notational simplicity let $x_{it|t-1} = \sum_{l=1}^{p_i} \phi_{il} x_{it-l} + \sum_{k=1}^{q_i} \theta_{ik} u_{it-k}$ 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 predictors are generated by the following DGPs:

$$\begin{aligned} y_t &= \alpha_1 x_{1t-1} + \alpha_2 x_{2t-1} + \varepsilon_t \\ x_{it} &= \phi_i x_{it-1} + u_{it} \\ \varepsilon_t &= \phi_\varepsilon \varepsilon_{t-1} + \omega_t, \end{aligned} \tag{15}$$

where $i = 1, 2$, $t = 1, \dots, T$, $|\phi_i| < 1$, $|\phi_\varepsilon| < 1$, and u_{it} and ω_t are the *i.i.d.* random errors of the processes. 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 .

In this context we can apply a standard OLS estimator to estimate α_1 and α_2 . However, we remark that the problems caused by serial dependence for the non-asymptotic estimation error bound of the PRs also apply to the OLS estimator (see Corollary 6 in Negahban et al. 2009). For the sake of the argument, temporarily assume that the u_{it-1} 's are observable, so that we do not need to estimate them through the $(x_{it-1} - \widehat{x}_{it-1|t-2})$'s. If we could observe the errors, our proposal would consist of estimating the following *working model*:

$$y_t = \alpha_1 u_{1t-1} + \alpha_2 u_{2t-1} + \phi_y y_{t-1} + \omega_t. \tag{16}$$

We refer to OLS applied to this as *u*-OLS, and we illustrate it in Examples 1– 4.

Example 1 (*Equal degrees of serial dependence*). Suppose $\phi_1 = \phi_2 = \phi_\varepsilon = \phi$. Then, model (15) can be rewritten as

$$\begin{aligned} y_t &= \alpha_1(\phi x_{1t-2} + u_{1t-1}) + \alpha_2(\phi x_{2t-2} + u_{2t-1}) + \phi \varepsilon_{t-1} + \omega_{it} \\ &= \alpha_1 u_{1t-1} + \alpha_2 u_{2t-1} + \phi y_{t-1} + \omega_t. \end{aligned}$$

Thus, in an “ideal regime” in terms of degree of serial dependence (also known as “common factor restriction”), the working model (16) is equivalent to the true model (15) because of the decomposition of the AR(1) processes x_{1t-1} , x_{2t-1} and ε_t .

Example 2 (*Different degrees of serial dependence*). Suppose $\phi_1 \neq \phi_2 \neq \phi_\varepsilon$. Then, with some simple steps, model (15) can be rewritten as

$$\begin{aligned} y_t &= \alpha_1(\phi_1 x_{1t-2} + u_{1t-1}) + \alpha_2(\phi_2 x_{2t-2} + u_{2t-1}) + \phi_\varepsilon \varepsilon_{t-1} + \omega_{it} \\ &= \alpha_1 u_{1t-1} + \alpha_2 u_{2t-1} + \alpha_1 \phi_1 x_{1t-2} + \alpha_2 \phi_2 x_{2t-2} + \phi_\varepsilon \varepsilon_{t-1} + \omega_t . \end{aligned}$$

Thus, in this perhaps more realistic regime, the working model (16) is not equivalent to the true model (15) since the predictors and the error do not have the same degree of serial dependence, and therefore the use of y_{t-1} does not allow us to summarize the serial dependence of y_t .

Example 3 (*Equal degrees of serial dependence and different models for the predictors*). Consider x_{1t} and x_{2t} generated as

$$\begin{aligned} x_{1t} &= \phi x_{1t-1} + \phi x_{1t-2} + u_{1t} \\ x_{2t} &= \phi x_{2t-1} + \theta u_{2t-1} + u_{2t} \end{aligned}$$

where $2|\phi| < 1$. Model (15) can be rewritten as

$$\begin{aligned} y_t &= \alpha_1(\phi x_{1t-2} + \phi x_{1t-3} + u_{1t-1}) + \alpha_2(\phi x_{2t-2} + \theta u_{2t-2} + u_{2t-1}) + \phi_\varepsilon \varepsilon_{t-1} + \omega_{it} \\ &= \alpha_1 u_{1t-1} + \alpha_2 u_{2t-1} + \phi y_{t-1} + \phi \alpha_1 x_{1t-3} + \alpha_2 \theta u_{2t-2} + \omega_t . \end{aligned}$$

Thus, if we have an “ideal regime” in terms of degree of serial dependence, but different models for the predictors, the working model (16) is not equivalent to the true model (15). Here, the difference between true and working model is due to the differences between the mechanisms generating $x_{1t|t-1}$ and $x_{2t|t-1}$. Again, this makes y_{t-1} not suitable for summarizing the serial dependence of y_t .

Example 4 (*Equal degrees of serial dependence and different model for the error*). Consider now the case where

$$\varepsilon_t = \phi \varepsilon_{t-1} + \phi \varepsilon_{t-2} + \omega_t$$

with $2|\phi| < 1$. Model (15) can be rewritten as

$$\begin{aligned} y_t &= \alpha_1(\phi x_{1t-2} + u_{1t-1}) + \alpha_2(\phi x_{2t-2} + u_{2t-1}) + \phi \varepsilon_{t-1} + \phi \varepsilon_{t-2} + \omega_{it} \\ &= \alpha_1 u_{1t-1} + \alpha_2 u_{2t-1} + \phi y_{t-1} + \phi \varepsilon_{t-2} + \omega_t . \end{aligned}$$

Thus, if we have an “ideal regime” in terms of degree of serial dependence, but a different model for the error, the working model (16) is not equivalent to the true model (15). 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.

It is crucial to note that in Examples 1-4 the working model allows us to estimate the true α_1 and α_2 through u_{1t-1} and u_{2t-1} , regardless of the possible issues in estimating the serial dependence of y_t . This is possible because $u_{it} \perp x_{it|t-1}$ for any specification of $x_{it|t-1}$, which is a consequence of Assumptions 2 and 3. For this reason, even if we omit the autoregressive and/or the moving average component(s) from the working model, this does not lead to an omitted-variables bias of the estimated α 's. However, we get a reduction in the explained variance of y , which is mitigated by including the lags of y_t among the regressors of the working model.

Thus, even when true and working models do not match, as in Examples 2-4, u -OLS moves us from estimating coefficients in a context characterized by high spurious cross-correlation, to one characterized by very weak (or absent) spurious cross-correlations. Of course the parameters we estimate about the past of y_t change, but we can still formulate an effective forecasting strategy. Specifically, when $\varepsilon_t = \sum_{j=1}^{p_\varepsilon} \phi_{\varepsilon j} \varepsilon_{t-j} + \omega_t$, our conjecture is that (even in cases such as Examples 2-4), the variability introduced by a misspecification of the serial dependence of y_t through the estimation of the working model is less than that introduced by estimating the model directly on the \mathbf{x} 's.

Regarding the error term, since it is not correlated with the regressors included in the model, serial dependence does not violate the assumption of exogeneity and the OLS estimator remains unbiased and consistent. However, one strategy to prevent autocorrelation in standard errors is to increase the number of lags of y_t considered in the working model, as to come as close as possible to a “white noise” residual. Therefore, considering lagged dependent variables can help us cope with the existence of autocorrelation in the model.

Of course, in practice, the u_{it-1} 's 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} -OLS or, in the case of PRs, as \hat{u} -PRs. In the following Sections we show its potential. First, we present simulation experiments where we compare \hat{u} -OLS with methods customarily applied to correct for serial dependence. Then, we demonstrate the estimation and forecasting performance of \hat{u} -LASSO (LASSO applied on ARMA residuals) through both simulations and an empirical application. We note here that an in-depth study of the advantages in forecasting and variable selection offered by sparse \hat{u} -PRs will be the topic of future studies.

5.2 \hat{u} -OLS: Coefficients Estimation and Prediction Accuracy in Low Dimension

We simulate the data using (15); that is

$$y_t = \alpha_1 x_{1t-1} + \alpha_2 x_{2t-1} + \varepsilon_t$$

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

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

for $i = 1, 2$, $t = 1, \dots, T$. Here we take the *i.i.d.* errors u_{it} and ω_t to be standard Normal random variables for which Assumptions 2 and 3 hold, and we consider three different scenarios:

1. *Equal degrees of serial dependence*, with $\phi_1 = \phi_2 = \phi_\varepsilon = 0.7$ (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. Moreover, in this common factor restriction regime, the true model is those estimated through the CO and DynReg methods; see below (McGuirk and Spanos, 2002; Baillie et al., 2022).
2. *Different degrees of serial dependence*, with $\phi_1 = 0.75$, $\phi_2 = 0.6$, and $\phi_\varepsilon = 0.9$ (as in Example 2). Here the common factor restriction does not hold.
3. *Different models for predictors and error*, with

$$x_{1t} = 0.6x_{1t-1} + u_{1t} + 0.5u_{1t-1}$$

$$x_{2t} = 0.75x_{2t-1} + u_{2t}$$

$$\varepsilon_t = 0.6\varepsilon_{t-1} + 0.3\varepsilon_{t-2} + \omega_t .$$

Here x_{1t} , x_{2t} and ε_t are ARMA(1,1), AR(1) and AR(2) processes, respectively.

We compare coefficients estimation and forecasting performance of the following methods:

- NW: the Heteroskedasticity and Autocorrelation Consistent (HAC) Newey-West estimator (Newey and West, 1987), which accommodates autocorrelation and heteroskedasticity of the error terms in model (15). The forecasting equation, in terms of the projection of y_t on the hyperplane spanned by the covariates, is $y_t^{(x)} = \text{Proj}(y_t | y_{t-1}, x_{1t-1}, x_{2t-1})$.
- CO: the Cochrane-Orcutt generalized least squares (GLS) estimator (Cochrane and Orcutt, 1949), which adjusts a linear model for serial correlation in the error terms iterating two steps,

one to estimate the first order autocorrelation on OLS residuals, and one to transform the variables to eliminate serial dependence in the errors, until a certain criterion is satisfied (e.g., the estimated autocorrelation has converged); transformations are applied from the second observation onward, i.e. for $t = 2, \dots, T$. The forecasting equation is $y_t^{*(x)} = \text{Proj}(y_t^* | x_{1t-1}^*, x_{2t-1}^*)$, where $y_t^* = y_t - \hat{\phi}_\varepsilon^* y_{t-1}$, $x_{it-1}^* = x_{it-1} - \hat{\phi}_\varepsilon^* x_{it-2}$, and $\hat{\phi}_\varepsilon^*$ is the CO estimate of ϕ_ε .

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

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

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

Table 1: Coefficient estimation error and coefficient of determination (R^2) of Newey-West-style HAC estimator (NW), Cochrane-Orcutt GLS estimator (CO), Dynamic Regression (DynReg) and OLS applied on $\hat{\mathbf{u}}$'s (\hat{u} -OLS) across the three simulation scenarios (DGPs). Panel (a) $T = 100$, panel (b) $T = 1000$. Results are obtained on 1000 Monte Carlo replications.

5.3 \hat{u} -LASSO

5.3.1 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

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

where $\varepsilon_t = \phi \varepsilon_{t-1} + \omega_t$, and $\omega_t \sim N(0, \sigma_\omega^2)$. The coefficient vector $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_n)'$ is sparse with $\|\boldsymbol{\alpha}\|_0 = 10$. The active covariates are the first 10, followed by $n - 10$ inactive ones, and $\alpha_1 = \dots = \alpha_{10} = 1$. We generate the n covariates as $x_{it} = \phi x_{it-1} + u_{it}$, $i = 1, \dots, n$, $t = 1, \dots, 100$, where $u_{it} \sim N(0, 1)$ and $c_{ij}^u = 0.3^{|i-j|}$. We consider $n = 50, 150$ and $\phi = 0.3, 0.6, 0.9, 0.95$. Panels (a) and (b) of Figure 7 display mean and standard deviation of the ratio between $\hat{\psi}_{min}^{\hat{u}}$ and $\hat{\psi}_{min}^x$ obtained from 1000 Monte Carlo simulations run with $n = 50$ and 150 , respectively; for $n = 150$ we consider the minimum eigenvalue of the correlation matrix restricted to the 10 relevant variables. As expected, the

correlation matrix of the $\hat{\mathbf{u}}$'s does not suffer from spurious correlation induced by serial dependence, and this leads to an increment of $\frac{\hat{\psi}_{min}^{\hat{\mathbf{u}}}}{\psi_{min}^x}$ as ϕ increases. To observe how this result translates into coefficients estimation accuracy, we compare the coefficient estimation error of LASSO ($\|\hat{\boldsymbol{\alpha}}_x - \boldsymbol{\alpha}\|_2$) with that of $\hat{\mathbf{u}}$ -LASSO ($\|\hat{\boldsymbol{\alpha}}_{\hat{\mathbf{u}}} - \boldsymbol{\alpha}\|_2$), where the tuning parameter λ is selected by BIC. Figure 8 shows how the mean and standard deviation of $\frac{\|\hat{\boldsymbol{\alpha}}_{\hat{\mathbf{u}}} - \boldsymbol{\alpha}\|_2}{\|\hat{\boldsymbol{\alpha}}_x - \boldsymbol{\alpha}\|_2}$ vary as a function of ϕ . Also here, as expected, the application of LASSO on serially uncorrelated data reduces the coefficient estimation error, with a gain in estimation accuracy that increases with ϕ . To summarize, results shown in Figures 7 and 8 corroborate the theoretical analysis according to which an increase in the degree of serial dependence leads to an increase in the probability of large spurious correlations, which in turn increases the probability of a small minimum eigenvalue for the sample correlation matrix. This negatively affects the estimation accuracy of PRs (see Proposition 1).

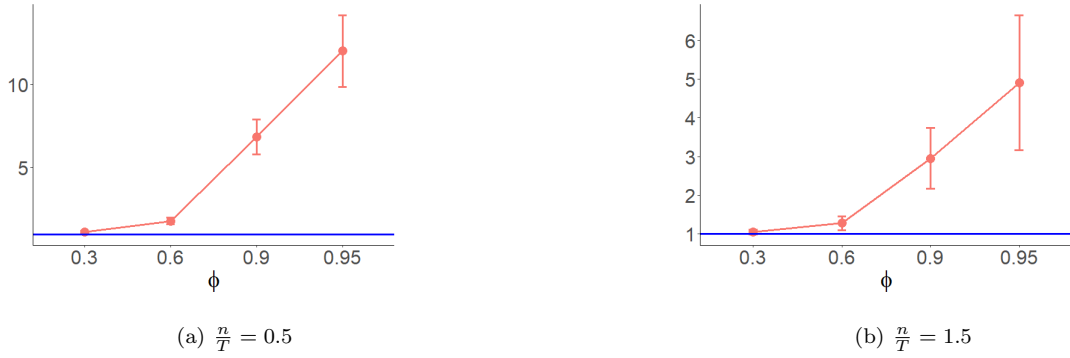


Figure 7: (a) Average and (b) standard deviation of $\frac{\hat{\psi}_{min}^{\hat{\mathbf{u}}}}{\psi_{min}^x}$ across 1000 Monte Carlo replications, for several degrees of serial dependence (ϕ). The blue horizontal line marks a ratio value of 1.



Figure 8: (a) Average and (b) standard deviation of $\frac{\|\hat{\boldsymbol{\alpha}}_{\hat{\mathbf{u}}} - \boldsymbol{\alpha}\|_2}{\|\hat{\boldsymbol{\alpha}}_x - \boldsymbol{\alpha}\|_2}$ across 1000 Monte Carlo replications, for several degree of serial dependence (ϕ). The blue horizontal line marks a ratio value of 1.

5.3.2 Empirical Application

We consider Euro Area data obtained from Proietti and Giovannelli (2021), composed by 309 monthly macroeconomic time series spanning the period between January 1997 and December 2018. The series were all transformed to achieve stationarity by taking first or second differences, logarithms or first or second differences of logarithms (full detail on the transformations applied to each series is given in Supplement H). However, no treatment for outliers was applied.

The target variable is the Consumer Price Index (CPI), which is transformed as I(2), i.e. integrated of order 2, following Stock and Watson (2002b):

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

where $y_t = 1200\log(CPI_t/CPI_{t-1}) - 1200\log(CPI_{t-1}/CPI_{t-2})$, and h is the forecasting horizon. We compute forecasts of y_{t+h} at horizon $h = 24$ using a rolling ω -year window $[t - \omega, t + 1]$; the models are re-estimated at each t , adding one observation on the right of the window and removing one observation on the left. In particular, we fix $\omega = 130$, using data from Feb:1997 to Dec:2007 for the first estimation. The last available date is Dec:2018. The methods employed for our empirical exercises are:

- *Univariate AR(p)*: the autoregressive forecasting model based on p lagged values of the target variable, i.e. $\hat{y}_{t+h} = \hat{\alpha}_0 + \sum_{i=1}^p \hat{\phi}_i y_{t-i+1}$.
- *LASSO*: the classical LASSO (Tibshirani, 1996). Forecasts are obtained from the equation $\hat{y}_{t+h}^x = \hat{\alpha}_0 + \hat{\alpha}'_x \mathbf{x}_t + \sum_{i=1}^p \hat{\phi}_i y_{t-i+1}$, where $\hat{\alpha}_x$ is the sparse vector of penalized regression coefficients estimated by the LASSO on the original time series.
- *\hat{u} -LASSO*: our proposal, where LASSO is applied to the residuals of the estimated serial-dependent processes. Forecasts are obtained from the equation $\hat{y}_{t+h}^{\hat{u}} = \hat{\alpha}_0^{\hat{u}} + \hat{\alpha}'_{\hat{u}} \hat{\mathbf{u}}_t + \sum_{i=1}^p \hat{\phi}_i y_{t-i+1}$, where $\hat{\alpha}_{\hat{u}}$ is the sparse vector of penalized regression coefficients estimated by the LASSO on the estimated residuals.

For the AR(p) benchmark, coefficients are estimated using the the R package *lm* and 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) using the R package *auto.arima*, and p_i and q_i are selected by BIC within $0 \leq p_i \leq 12$, $0 \leq q_i \leq 12$, $i = 1, \dots, n$. The shrinkage parameter λ of LASSO and \hat{u} -LASSO is selected with BIC by using the R package *HDeconometrics*.

Forecasting accuracy for all three methods is evaluated using the root mean square forecast error (RMSFE), defined as

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

where T_0 and T_1 are the first and last point in time 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 RMSFE’s between pairs of methods, as well as significance of the corresponding Diebold-Mariano test (Diebold and Mariano, 1995) for the alternative hypothesis that the second method is less accurate. There are two relevant findings. First, \hat{u} -LASSO produces significantly better forecasts than both the classical LASSO and the AR(p). Second, \hat{u} -LASSO provides a parsimonious model with respect to the LASSO. In fact, the ratio between the average number of selected variable with \hat{u} -LASSO over the average number of selected variable with LASSO is 0.133, and the ratio between the standard deviations is 0.289 (note that average and standard deviation of the selected variables by LASSO are 61 and 20, respectively). 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 – 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.

Method 1	Method 2	RMSFE (ratio)
\hat{u} -LASSO	LASSO	0.82**
\hat{u} -LASSO	AR(p)	0.88**
LASSO	AR(p)	1.08

Table 2: Ratio of the RMSFE of the employed methods. We report the p-value of the Diebold-Mariano test for the alternative hypothesis that the second method is less accurate in forecasting. In particular, the p-value are indicated as follows: 0 ****, 0.001 ***, 0.01 **, 0.05 *, 0.1’.

6 Concluding Remarks

In this paper we demonstrated that the probability of spurious cross-correlations between stationary orthogonal or weakly cross-correlated processes depends not only on the sample size, but also on the degree of 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 on the non-asymptotic estimation error bounds of PRs, our findings highlight the limitations of these methods in a time series context. In order to tackle such limitations, we proposed a two-step procedure based on the application of PRs on the residuals obtained by filtering each time series with an ARMA process.

We showed that in a low dimensional scenario, where the degree of serial dependence is the same for covariates and error, our procedure estimates a working model equivalent to the real one. We also showed that beyond such simple scenario, our procedure still provides a valid estimation and forecasting strategy. We assessed the performance of our proposal through Monte Carlo simulations and an empirical analysis of 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 cross-correlation, performing better than classical LASSO in coefficients estimation. Through the empirical analysis 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

- 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). On robust inference in time series regression.
- Bartlett, M. S. (1935). Some aspects of the time-correlation problem in regard to tests of significance. *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.
- Box, G. and P. Newbold (1971). Some comments on a paper of coen, gomme and kendall. *Journal of the Royal Statistical Society, Series A (General)* *134*(2), 229–240.
- 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 (2013). Challenges of big data analysis. *National Science Review* *1*.
- 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., 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., A. Giovannelli, M. Lippi, and S. Soccorsi (2018). Dynamic factor model with infinite-dimensional factor space: Forecasting. *Journal of Applied Econometrics* 33(5), 625–642.
- 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.
- Forni, M., M. Hallin, M. Lippi, and L. Reichlin (2003). The generalized dynamic factor model. one-sided estimation and forecasting.
- Forni, M., M. Hallin, M. Lippi, and P. Zaffaroni (2016). Dynamic Factor Models with Infinite-Dimensional Factor Space. Asymptotic Analysis. Technical report.
- Giannone, D., M. Lenza, and G. E. Primiceri (2018, August). Economic Predictions with Big Data: The Illusion Of Sparsity. CEPR Discussion Papers 12256, C.E.P.R. Discussion Papers.
- 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., R. Tibshirani, and M. Wainwright (2015). Statistical learning with sparsity : the lasso and generalizations. pp. 291 – 294.
- Hoerl, A. and W. Kennard (1970). Ridge regression: Biased estimation for nonorthogonal problems. pp. 55–67.
- Lounici, K., M. Pontil, A. B. Tsybakov, and S. van de Geer (2009). Taking advantage of sparsity in multi-task learning.
- McGregor, J. R. (1962, 12). The approximate distribution of the correlation between two stationary linear Markov series. *Biometrika* 49(3-4), 379–388.
- McGuirk, A. M. and A. Spanos (2002). The Linear Regression Model With Autocorrelated Errors: Just Say No To Error Autocorrelation. Technical report.

- Medeiros, M. C. and E. F. Mendes (2012). Estimating high-dimensional time series models. Texto para discuss textasciitildeao 602, Rio de Janeiro.
- Negahban, S., B. Yu, M. J. Wainwright, and P. Ravikumar (2009). A unified framework for high-dimensional analysis of m -estimators with decomposable regularizers. *22*.
- 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.
- Smeekes, S. and E. Wijler (2018). Macroeconomic Forecasting Using Penalized Regression Methods.
- 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. Boston, Mass. ; London: 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. *J. Mach. Learn. Res.* 7, 2541–2563.
- Zou, H. (2006). The adaptive lasso and its oracle properties. *Journal of the American Statistical Association* 101, 1418–1429.
- Zou, H. and T. Hastie (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society, Series B* 67, 301–320.
- Zou, H. and H. H. Zhang (2009). On the adaptive elastic-net with a diverging number of parameters. *The Annals of Statistics* 37(4), 1733 – 1751.

Supplement

A Restricted Eigenvalue

In the specific case of $n > T$ the loss function $\mathcal{L}(\boldsymbol{\alpha})$ cannot be strongly convex since $\mathbf{X}\mathbf{X}'/T$ 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 $\Delta = |\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}| \in \mathbb{R}^n$, named *Restricted Eigenvalue Condition*. In particular, for any subset $S \subseteq \{1, 2, \dots, n\}$ with cardinality s , let $\Delta_S \in \mathbb{R}^S$ and $\Delta_{S^c} \in \mathbb{R}^{S^c}$. The restricted eigenvalue condition requires that there exists a positive number ν such that

$$\min_{\Delta \in \mathbb{R}^n: \Delta \neq 0} \frac{\|\mathbf{X}\Delta\|_2}{\sqrt{T} \|\Delta_S\|_2} \geq \nu.$$

Such condition is essentially a restriction on the eigenvalues of $\mathbf{X}\mathbf{X}'/T$ as a function of sparsity, which allows for the strong convexity (Definition 1) 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

We consider the case where the covariates in model (1) are generated as follows

$$x_{it} = \lambda_i F_t + u_{it}, \tag{17}$$

with $i = 1, \dots, n$, $t = 1, \dots, T$. F_t represents a common factor that introduce population cross-correlation between covariates, λ_i is the factor loading relative to x_i , and u_{it} is the idiosyncratic component relative to x_i at time t .

In this case, Fan et al. (2020) propose a method to reduce the cross-correlation between covariates in order to improve the estimation performance 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 λ_i , F_t and u_{it} . Then the vector $\hat{\boldsymbol{\alpha}}$ is obtained by replacing the design matrix \mathbf{X} in equation (2) with the $T \times (1+n)$ design matrix containing $\hat{\mathbf{F}}$, $\hat{\mathbf{u}}_1, \dots, \hat{\mathbf{u}}_n$. Note that the penalty is applied on the estimated residuals only. Hence, when covariates are generated by model (17), the procedure proposed by Fan et al. (2020) allows us

to deal with the problem from PRs estimation with highly cross-correlated covariates $\mathbf{x}_1, \dots, \mathbf{x}_n$ to PRs estimation with weakly or orthogonal covariates $\hat{\mathbf{u}}_1, \dots, \hat{\mathbf{u}}_n$.

However, we want to stress that if the idiosyncratic components are orthogonal Gaussian AR(1) processes with equal signs of the autoregressive coefficients, then the density of their sample cross-correlation coincides with that observed in Figure 2-4, and therefore the methodology proposed by Fan et al. (2020) would not solve the problem of the high spurious cross-correlation caused by serial dependence.

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

In Figure 9 we report the density of $\widehat{Cov}(u_1, u_2)$ when u_1 and u_2 are standard Normal in the cases of $T = 10$ and 100. Red line shows the density of $N\left(0, \frac{1}{T-1}\right)$. Observations are obtained on 5000 Monte Carlo replications. We observe that the approximation of $\widehat{Cov}(u_1, u_2)$ to $N\left(0, \frac{1}{T-1}\right)$ holds also when T is small (see Figure 9 (a) relative to $T=10$). This analysis corroborate numerically the results in Glen et al. (2004), which show that if x and y are $N(0, 1)$, then the probability density function of xy is $\frac{K_0(|xy|)}{\pi}$, where $K_0(|xy|)$ is the Bessel function of the second kind.

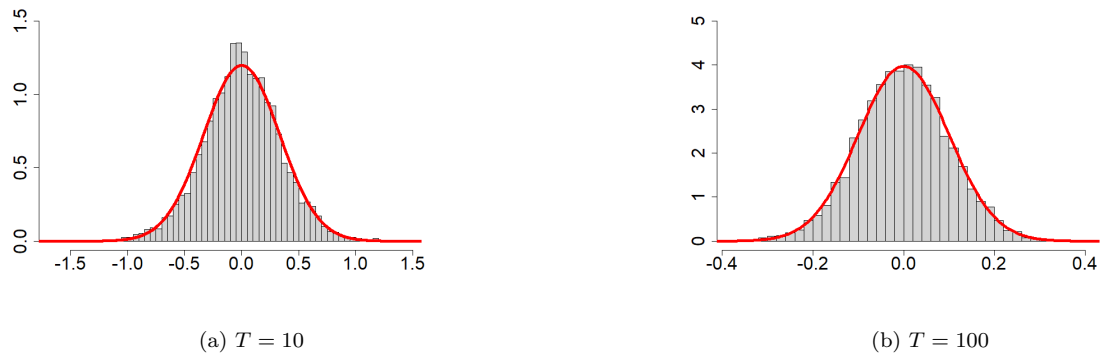


Figure 9: Density of $\widehat{Cov}(u_1, u_2)$ between two uncorrelated standard Normal variables for $T = 10$ (a) and $T = 100$ (b).

D Distribution of b

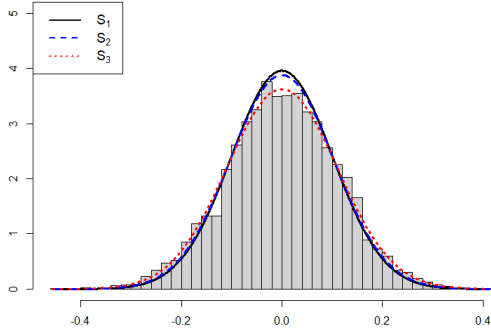
Consider two orthogonal Gaussian AR(1) processes generated according to the model $x_{it} = \phi_i x_{it-1} + u_{it}$, where $u_{it} \sim N(0, 1)$, $i = 1, 2$, $t = 1, \dots, 100$ and $\phi_1 = \phi_2 = \phi$. In this simulation exercise we run the model

$$x_{2t} = \beta x_{1t} + e_t, \quad (18)$$

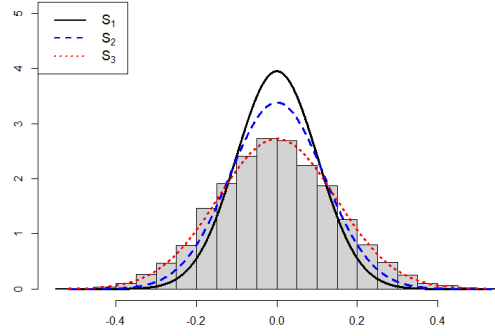
where $e_t \sim N(0, \sigma_e^2)$, and study the distribution of the OLS estimator b of β in the following four cases in terms of degrees of serial dependence: $\phi = 0.3, 0.6, 0.9, 0.95$. Figure 10 reports the density of b across the ϕ values obtained on 5000 Monte Carlo replications. We compare this density with that of three zero-mean Normal variables where the variances are respectively:

- $S_1^2 = \frac{\hat{\sigma}_e^2}{\sum_{t=1}^T (x_{1t} - \bar{x}_1)^2}$, where $\hat{\sigma}_e^2$ is the sample variance of the estimated residual $\hat{e}_t = x_{2t} - bx_{1t}$. This is the OLS estimator for the variance of β .
- $S_2^2 = \frac{1}{T} \frac{\frac{1}{T-2} \sum_{t=1}^T (x_{1t} - \bar{x}_1)^2 \hat{e}_t^2}{[\frac{1}{T} \sum_{t=1}^T (x_{1t} - \bar{x}_1)^2]^2} \hat{f}_t$, is the Newey-West (NW) HAC estimator (Newey and West, 1987; Stock and Watson, 2008), 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 $\hat{\rho}_j = \frac{\sum_{t=j+1}^T \hat{v}_t \hat{v}_{t-j}}{\sum_{t=1}^T \hat{v}_t^2}$, with $\hat{v}_t = (x_{1t} - \bar{x}_1) \hat{e}_t$. A rule of thumb for choosing m is $m = [0.75T^{1/3}]$.
- $S_3^2 = \frac{(1-\phi_1^2\phi_2^2)(1-\phi_1^2)}{(T-1)(1-\phi_2^2)(1-\phi_1\phi_2)^2}$, is the theoretical variance of b obtained in Proposition 2.

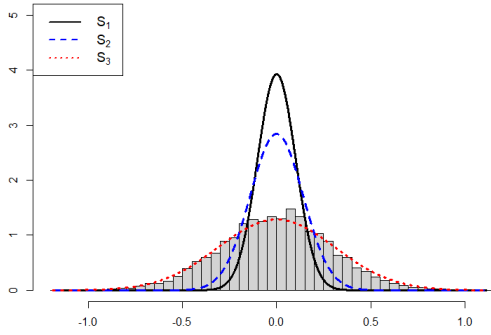
From Figure 10 we observe that the variance of b increases with the degree of serial dependence (ϕ) in a way that is well approximated by the distribution derived in Proposition 2 (see dotted red line). On the contrary, OLS (solid line) and NW (dashed blue line), are highly sub-optimal in the presence of strong serial dependence, underestimating the variability of b as the serial dependence increases.



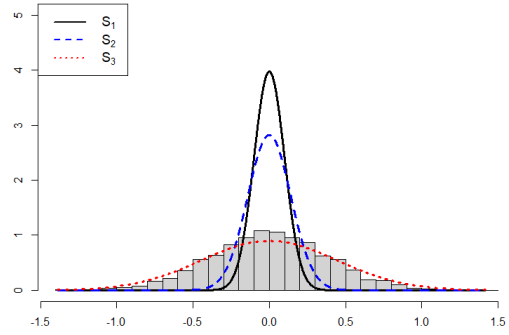
(a) $\phi = 0.3, T = 100$



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



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



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

Figure 10: Density of b between uncorrelated AR(1) Gaussian processes. Black line indicates the approximated density obtained by using the classical OLS estimator, dashed blue line indicates the approximated density obtained by using the NW estimator, and, finally, dotted red line shows the theoretical approximated density obtained in Proposition 2.

E More General Cases

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

The Impact of non-Gaussianity

The theoretical contribution reported in Section 3 requires the Gaussianity of u_1 and u_2 . With the following simulation experiments we show that the impact of ϕ_{12} on the density of \widehat{c}_{12}^x is relevant also when u_{1t} and u_{2t} are not 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 mixed case where u_1 is generated by a t -student with 1 degree of freedom and u_2 by a Uniform in the interval $(-4,4)$ (case (C)). Figure 11 reports the results of the simulation experiment. All densities in Figure 11 gave very similar result, with the exception of the Cauchy (Figure 11 (b)) where the effect of ϕ_{12} declines. However, we can state that regardless the distribution of the processes, whenever $Sign(\phi_1) = Sign(\phi_2)$, the probability of large values of \widehat{c}_{12}^x increases with ϕ_{12} .

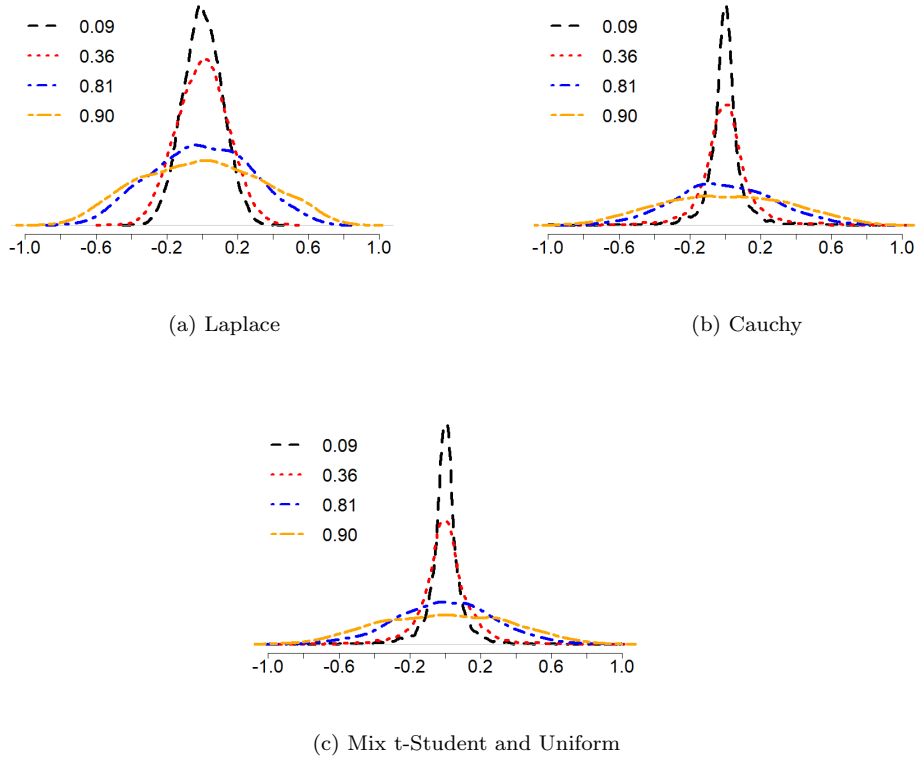


Figure 11: Simulated density of \widehat{c}_{12}^x in various scenarios in terms of ϕ_{12} for not Gaussian processes.

The Impact of Population Cross-Correlation

Since orthogonality is an unrealistic assumption for most economic applications, here we admit population cross-correlation. In Figure 12 we report $d_s(\widehat{c}_{12}^x)$ when the processes are weakly cross-correlated with $c_{12}^u = 0.2$, 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 ϕ_{12} on $d_s(\widehat{c}_{12}^x)$ depends on the degree of (population) cross-correlation as follows. In the case of weakly correlated processes, an increase in ϕ_{12} yields a high probability of observing large sample correlations in absolute value (the obtained density is similar to that in Figure 3 (a) with the obvious difference that here it is not centered at zero). In the case of multicollinear processes, on the other hand, an increase in ϕ_{12} leads to a high probability of underestimating the true population cross-correlation.

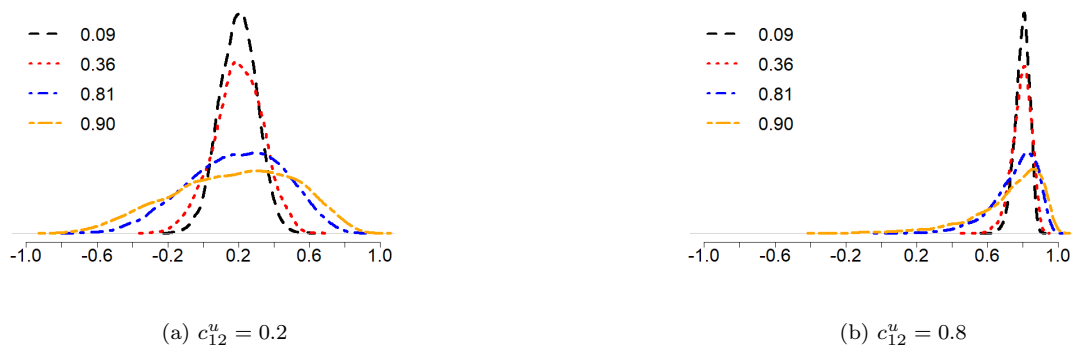


Figure 12: $d_s(\widehat{c}_{12}^x)$ obtained through simulations in the case of $c_{12}^x = 0.2$ (a) and $c_{12}^x = 0.8$ (b).

Density of \widehat{c}_{12}^x 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

$$\begin{aligned} x_{1t} &= \phi x_{1t-1} + \phi x_{1t-2} - \phi x_{1t-3} u_{1t} + 0.5 u_{1t-1}, \\ x_{2t} &= \phi x_{2t-1} + \phi x_{2t-2} + u_{2t} + 0.7 u_{2t-1} - 0.4 u_{3t-2}, \end{aligned}$$

where $t = 1, \dots, 100$ and $u_i \sim N(0, 1)$. In Figure 13 we report the density of \widehat{c}_{12}^x in the case of $T = 100$ and $\phi = 0.1, 0.2, 0.3, 0.33$. With no loss of generality we can observe that $d_s(\widehat{c}_{12}^x)$ gets larger as ϕ increases, that is $\Pr\{|\widehat{c}_{12}^x| \geq \tau\}$ increases with $|\phi|$.

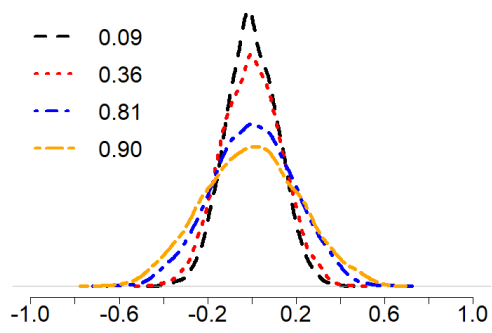


Figure 13: Densities of $d_s(\hat{\mathcal{C}}_{12}^r)$ between two uncorrelated ARMA Gaussian processes.

F Extended Results for \hat{u} -OLS

Metric	Stat.	(a)				(b)			
		NW	CO	DynReg	\hat{u} -OLS	NW	CO	DynReg	\hat{u} -OLS
$ \hat{\alpha}_1 - \alpha_1 $	ave.	0.241	0.104	0.105	0.104	0.244	0.031	0.031	0.032
	s.d.	0.244	0.124	0.124	0.123	0.145	0.038	0.038	0.038
	p -val. ave.	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	p -val. s.d.	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
$ \hat{\alpha}_2 - \alpha_2 $	ave.	0.244	0.101	0.103	0.102	0.243	0.032	0.032	0.032
	s.d.	0.248	0.122	0.123	0.121	0.143	0.038	0.038	0.038
	p -val. ave.	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	p -val. s.d.	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
$\hat{\phi}_y$	ave.	0.322	–	0.658	0.685	0.344	–	0.697	0.698
	s.d.	0.231	–	0.320	0.280	0.134	–	0.176	0.155
	p -val. ave.	0.002	–	0.000	0.000	0.000	–	0.000	0.000
	p -val. s.d.	0.013	–	0.000	0.000	0.000	–	0.000	0.000
R^2	ave.	0.747	0.682	0.824	0.817	0.747	0.668	0.829	0.829
	s.d.	0.066	0.056	0.045	0.046	0.021	0.017	0.014	0.014

Table 3: Estimation and in sample prediction performance 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) in the case of constant degree of serial correlation: $\phi_1 = \phi_2 = \phi_\varepsilon = 0.7$. Panel (a) $T = 100$, panel (b) $T = 1000$. Results are obtained on 1000 Monte Carlo replications.

Metric	Stat.	(a)				(b)			
		NW	CO	DynReg	\hat{u} -OLS	NW	CO	DynReg	\hat{u} -OLS
$ \hat{\alpha}_1 - \alpha_1 $	ave.	0.430	0.101	0.105	0.110	0.481	0.030	0.031	0.033
	s.d.	0.376	0.123	0.127	0.131	0.228	0.037	0.038	0.040
	p -val. ave.	0.003	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	p -val. s.d.	0.026	0.000	0.000	0.000	0.000	0.000	0.000	0.000
$ \hat{\alpha}_2 - \alpha_2 $	ave.	0.357	0.097	0.104	0.105	0.383	0.029	0.031	0.032
	s.d.	0.338	0.116	0.124	0.124	0.205	0.035	0.037	0.038
	p -val. ave.	0.002	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	p -val. s.d.	0.008	0.000	0.000	0.000	0.000	0.000	0.000	0.000
$\hat{\phi}_y$	ave.	0.573	–	0.854	0.782	0.642	–	0.895	0.807
	s.d.	0.355	–	0.313	0.303	0.216	–	0.159	0.169
	p -val. ave.	0.000	–	0.000	0.000	0.000	–	0.000	0.000
	p -val. s.d.	0.000	–	0.000	0.000	0.000	–	0.000	0.000
R^2	ave.	0.749	0.694	0.874	0.858	0.759	0.688	0.888	0.875
	s.d.	0.072	0.053	0.040	0.044	0.022	0.017	0.011	0.013

Table 4: Estimation and in sample prediction performance 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) in the case of different degree of serial correlation: $\phi_1 = 0.75$, $\phi_2 = 0.6$, and $\phi_\varepsilon = 0.9$. Panel (a) $T = 100$, panel (b) $T = 1000$. Results are obtained on 1000 Monte Carlo replications.

Metric	Stat.	(a)				(b)			
		NW	CO	DynReg	\hat{u} -OLS	NW	CO	DynReg	\hat{u} -OLS
$ \hat{\alpha}_1 - \alpha_1 $	ave.	0.358	0.090	0.104	0.111	0.407	0.027	0.032	0.034
	s.d.	0.335	0.109	0.127	0.137	0.209	0.032	0.037	0.042
	p -val. ave.	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	p -val. s.d.	0.003	0.000	0.000	0.000	0.000	0.000	0.000	0.000
$ \hat{\alpha}_2 - \alpha_2 $	ave.	0.342	0.110	0.111	0.121	0.375	0.033	0.032	0.036
	s.d.	0.342	0.129	0.130	0.143	0.210	0.040	0.038	0.043
	p -val. ave.	0.002	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	p -val. s.d.	0.016	0.000	0.000	0.000	0.000	0.000	0.000	0.000
$\hat{\phi}_y$	ave.	0.403	–	0.572	0.777	0.470	–	0.597	0.798
	s.d.	0.312	–	0.352	0.333	0.200	–	0.197	0.192
	p -val. ave.	0.004	–	0.000	0.000	0.000	–	0.000	0.000
	p -val. s.d.	0.021	–	0.002	0.000	0.000	–	0.000	0.000
R^2	ave.	0.762	0.698	0.874	0.831	0.765	0.683	0.887	0.853
	s.d.	0.067	0.052	0.038	0.050	0.022	0.016	0.012	0.015

Table 5: Estimation and in sample prediction performance 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) in the case of different DPs: $x_{1t} = 0.6x_{1t-1} + u_{1t} + 0.5u_{1t-1}$, $x_{2t} = 0.75x_{2t-1} + u_{2t}$, $\varepsilon_t = 0.6\varepsilon_{t-1} + 0.3\varepsilon_{t-2} + \omega_t$. 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 methods analyzed in the Section 5.2 to evaluate their ability in avoiding spurious regressions. In Table 6 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 6 panel (a)) suffers of spurious regressions for any $\phi > 0$, which occurs about %50 when $\phi = 0.9$. (ii) NW estimator (Table 6 panel (b)) reduces the problem, but for large value of ϕ spurious regression occurs frequently. Note that these two results are in line with those in Granger et al. (2001). (iii) CO, DynReg and \hat{u} -OLS (Table 6 panel (c)-(e)) solve the

problem of spurious regression due to serial dependence, by making the variance of b independent of ϕ . 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).

	T	$\phi = 0.0$	$\phi = 0.3$	$\phi = 0.6$	$\phi = 0.9$	$\phi = 0.95$
(a) $ t_b^{ols} > 1.96$	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
(b) $ t_b^{nw} > 1.96$	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
(c) $ t_b^{co} > 1.96$	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	5.14	4.80	4.86	5.56
(d) $ t_b^{dr} > 1.96$	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
(e) $ t_b^{\hat{u}-ols} > 1.96$	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

Table 6: Percentage of t -statistics over 1.96 in absolute value obtained on 1000 Monte Carlo replications.

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

Proposition 4 Let $S_{ols}^2 = \frac{\hat{\sigma}_\varepsilon^2}{\sum_{t=1}^T (x_{1t} - \bar{x}_1)^2}$, where $\hat{\sigma}_\varepsilon^2$ is the estimated variance of the residual of model

(18). Then

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

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

Note that the result in Proposition 4 has been also derived in Granger et al. (2001). This result show that the misspecification of t_b^{ols} is only due to the degree of serial dependence. To confirm this, look at the columns of Table 6 and consider that the value of $|t_b^{ols}|$ increases with the degree of serial dependence ϕ , but stay quite constant regardless of the sample size T .

H 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. As for the FRED data, the column tcode denotes the data transformation for a given series x_t : (1) no transformation; (2) Δx_t ; (3) $\Delta^2 x_t$; (4) $\log(x_t)$; (5) $\Delta \log(x_t)$; (6) $\Delta^2 \log(x_t)$. (7) $\Delta(x_t/x_{t-} - 1.0)$.

The acronyms for the sectors refer to:

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

Table 7: 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
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

Table 7 – continued from previous 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
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

Continued on next page

Table 7 – continued from previous page

ID	Description	Area	Sector	Tcode
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
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	Incl. Prod. - Excluding Constr.	EA	IP	5
106	Incl. Prod. - Cap. Goods	EA	IP	5
107	Incl. Prod. - Cons. Non-Durables	EA	IP	5

Continued on next page

Table 7 – continued from previous page

ID	Description	Area	Sector	Tcode
108	Incl. Prod. - Cons. Durables	EA	IP	5
109	Incl. Prod. - Cons. Goods	EA	IP	5
110	Incl. Prod.	FR	IP	5
111	Incl. Prod. - Mfg	FR	IP	5
112	Incl. Prod. - Mfg (2010=100)	FR	IP	5
113	Incl. Prod. - Manuf. Of Motor Vehicles, Trailers, Semitrailers	FR	IP	5
114	Incl. Prod. - Int. Goods	FR	IP	5
115	Incl. Prod. - Incl. Prod. - Constr.	FR	IP	5
116	Incl. Prod. - Manuf. Of Wood And Paper Products	FR	IP	5
117	Incl. Prod. - Manuf. Of Computer, Electronic And Optical Prod	FR	IP	5
118	Incl. Prod. - Manuf. Of Electrical Equipment	FR	IP	5
119	Incl. Prod. - Manuf. Of Machinery And Equipment	FR	IP	5
120	Incl. Prod. - Manuf. Of Transport Equipment	FR	IP	5
121	Incl. Prod. - Other Mfg	FR	IP	5
122	Incl. Prod. - Manuf. Of Chemicals And Chemical Products	FR	IP	5
123	Incl. Prod. - Manuf. Of Rubber And Plastics Products	FR	IP	5
124	Incl. Prod. - Investment Goods	IT	IP	5
125	Incl. Prod.	IT	IP	5
126	Incl. Prod.	IT	IP	5
127	Incl. Prod. - Cons. Goods - Durable	IT	IP	5
128	Incl. Prod. - Investment Goods	IT	IP	5
129	Incl. Prod. - Int. Goods	IT	IP	5
130	Incl. Prod. - Chemical Products & Synthetic Fibres	IT	IP	5
131	Incl. Prod. - Machines & Mechanical Apparatus	IT	IP	5
132	Incl. Prod. - Means Of Transport	IT	IP	5
133	Incl. Prod. - Metal & Metal Products	IT	IP	5
134	Incl. Prod. - Rubber Items & Plastic Materials	IT	IP	5
135	Incl. Prod. - Wood & Wood Products	IT	IP	5
136	Incl. Prod.	IT	IP	5
137	Incl. Prod. - Computer, Electronic And Optical Products	IT	IP	5
138	Incl. Prod. - Basic Pharmaceutical Products	IT	IP	5
139	Incl. Prod. - Constr.	DE	IP	5
140	Incl. Prod. - Ind Incl Cnstr	DE	IP	5
141	Incl. Prod. - Mfg	DE	IP	5
142	Incl. Prod. - Rebased To 1975=100	DE	IP	5
143	Incl. Prod. - Chems & Chem Prds	DE	IP	5

Continued on next page

Table 7 – continued from previous page

ID	Description	Area	Sector	Tcode
144	Incl. Prod. - Ind Excl Cnstr	DE	IP	5
145	Incl. Prod. - Ind Excl Energy & Cnstr	DE	IP	5
146	Incl. Prod. - Mining & Quar	DE	IP	5
147	Incl. Prod. - Cmptr, Elecccl & Opt Prds, Elecl Eqp	DE	IP	5
148	Incl. Prod. - Interm Goods	DE	IP	5
149	Incl. Prod. - Cap. Goods	DE	IP	5
150	Incl. Prod. - Durable Cons Goods	DE	IP	5
151	Incl. Prod. - Tex & Wearing Apparel	DE	IP	5
152	Incl. Prod. - Pulp, Paper&Prds, Pubshg&Print	DE	IP	5
153	Incl. Prod. - Chem Prds	DE	IP	5
154	Incl. Prod. - Rub&Plast Prds	DE	IP	5
155	Incl. Prod. - Basic Mtls	DE	IP	5
156	Incl. Prod. - Cmptr, Elecccl & Opt Prds, Elecl Eqp	DE	IP	5
157	Incl. Prod. - Motor Vehicles, Trailers&Semi Trail	DE	IP	5
158	Incl. Prod. - Tex & Wearing Apparel	DE	IP	5
159	Incl. Prod. - Paper & Prds, Print, Reprod Of Recrd Media	DE	IP	5
160	Incl. Prod. - Chems & Chem Prds	DE	IP	5
161	Incl. Prod. - Basic Mtls, Fab Mtl Prds, Excl Mach&Eqp	DE	IP	5
162	Incl. Prod. - Repair & Install Of Mach & Eqp	DE	IP	5
163	Incl. Prod. - Mfg Excl Cnstr & Fbt	DE	IP	5
164	Incl. Prod. - Mining & Ind Excl Fbt	DE	IP	5
165	Incl. Prod. - Ind Excl Fbt	DE	IP	5
166	Incl. Prod. - Interm & Cap. Goods	DE	IP	5
167	Incl. Prod. - Fab Mtl Prds Excl Mach & Eqp	ES	IP	5
168	Incl. Prod.	ES	IP	5
169	Incl. Prod. - Cons. Goods	ES	IP	5
170	Incl. Prod. - Cap. Goods	ES	IP	5
171	Incl. Prod. - Int. Goods	ES	IP	5
172	Incl. Prod. - Energy	ES	IP	5
173	Incl. Prod. - Cons. Goods, Non-Durables	ES	IP	5
174	Incl. Prod. - Mining	ES	IP	5
175	Incl. Prod. - Mfg Ind	ES	IP	5
176	Incl. Prod. - Other Mining & Quarrying	ES	IP	5
177	Incl. Prod. - Textile	ES	IP	5
178	Incl. Prod. - Chemicals & Chemical Products	ES	IP	5
179	Incl. Prod. - Plastic & Rubber Products	ES	IP	5

Continued on next page

Table 7 – continued from previous page

ID	Description	Area	Sector	Tcode
180	Incl. Prod. - Other Non-Metal Mineral Products	ES	IP	5
181	Incl. Prod. - Metal Processing Ind	ES	IP	5
182	Incl. Prod. - Metal Products Excl. Machinery	ES	IP	5
183	Incl. Prod. - Electrical Equipment	ES	IP	5
184	Incl. 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
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

Continued on next page

Table 7 – continued from previous page

ID	Description	Area	Sector	Tcode
216	All-Items Hicp	DE	HCPI	6
217	All-Items Hicp	ES	HCPI	6
218	All-Items Hicp	FR	HCPI	6
219	All-Items Hicp	IT	HCPI	6
220	Goods (Overall Index Excluding Services)	DE	HCPI	6
221	Goods (Overall Index Excluding Services)	FR	HCPI	6
222	Processed Food Including Alcohol And Tobacco	DE	HCPI	6
223	Processed Food Including Alcohol And Tobacco	ES	HCPI	6
224	Processed Food Including Alcohol And Tobacco	FR	HCPI	6
225	Processed Food Including Alcohol And Tobacco	IT	HCPI	6
226	Unprocessed Food	DE	HCPI	6
227	Unprocessed Food	ES	HCPI	6
228	Unprocessed Food	FR	HCPI	6
229	Unprocessed Food	IT	HCPI	6
230	Non-Energy Industrial Goods	DE	HCPI	6
231	Non-Energy Industrial Goods	FR	HCPI	6
232	Services (Overall Index Excluding Goods)	DE	HCPI	6
233	Services (Overall Index Excluding Goods)	FR	HCPI	6
234	Overall Index Excluding Tobacco	DE	HCPI	6
235	Overall Index Excluding Tobacco	FR	HCPI	6
236	Overall Index Excluding Energy	DE	HCPI	6
237	Overall Index Excluding Energy	FR	HCPI	6
238	Overall Index Excluding Energy And Unprocessed Food	DE	HCPI	6
239	Overall Index Excluding Energy And Unprocessed Food	FR	HCPI	6
240	Ppi: Ind Excluding Constr. & Energy	EA	PPI	6
241	Ppi: Cap. Goods	EA	PPI	6
242	Ppi: Non-Durable Cons. Goods	EA	PPI	6
243	Ppi: Int. Goods	EA	PPI	6
244	Ppi: Non Dom. - Mining, Mfg & Quarrying	EA	PPI	6
245	Ppi: Non Dom. Mfg	DE	PPI	6
246	Ppi: Int. Goods Excluding Energy	DE	PPI	6
247	Ppi: Cap. Goods	DE	PPI	6
248	Ppi: Cons. Goods	DE	PPI	6
249	Ppi: Fuel	DE	PPI	6
250	Ppi: Indl. Products (Excl. Energy)	DE	PPI	6
251	Ppi: Machinery	DE	PPI	6

Continued on next page

Table 7 – continued from previous page

ID	Description	Area	Sector	Tcode
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 , Motorcycles & Fuel	ES	T/O	5
258	T/O: Ret. Sale Of Clth & Leath Gds In Spcld Str	ES	T/O	5
259	T/O: Ret. Sale Of Non-Food Prds (Exc Fuel)	ES	T/O	5
260	T/O: Ret. Sale Of Info, Househld & Rec Eqp In Spcld Str	ES	T/O	5
261	Ek Unemployment: All	EA	HUR	5
262	Ek Unemployment: Persons Over 25 Years Old	EA	HUR	5
263	Ek Unemployment: Women Under 25 Years Old	EA	HUR	5
264	Ek Unemployment: Women Over 25 Years Old	EA	HUR	5
265	Ek Unemployment: Men Over 25 Years Old	EA	HUR	5
266	Fr Hur All Persons (All Ages)	FR	HUR	5
267	Fr Hur Femmes (Ages 15-24)	FR	HUR	5
268	Fr Hur Femmes (All Ages)	FR	HUR	5
269	Fr Hur Hommes (Ages 15-24)	FR	HUR	5
270	Fr Hur Hommes (All Ages)	FR	HUR	5
271	Fr Hur All Persons (Ages 15-24)	FR	HUR	5
272	Fr Hurall Persons(Ages 25 And Over)	FR	HUR	5
273	Fr Hur Females (Ages 25 And Over)	FR	HUR	5
274	Fr Hur Males (Ages 25 And Over)	FR	HUR	5
275	Bd Hur All Persons (All Ages)	DE	HUR	5
276	Bd Hur Femmes (Ages 15-24)	DE	HUR	5
277	Bd Hur Femmes (All Ages)	DE	HUR	5
278	Bd Hur Hommes (Ages 15-24)	DE	HUR	5
279	Bd Hur Hommes (All Ages)	DE	HUR	5
280	Bd Hur All Persons (Ages 15-24)	DE	HUR	5
281	Bd Hurall Persons(Ages 25 And Over)	DE	HUR	5
282	Bd Hur Females (Ages 25 And Over)	DE	HUR	5
283	Bd Hur Males (Ages 25 And Over)	DE	HUR	5
284	It Hur All Persons (All Ages)	IT	HUR	5
285	It Hur Femmes (All Ages)	IT	HUR	5
286	It Hur Hommes (All Ages)	IT	HUR	5
287	It Hur All Persons (Ages 15-24)	IT	HUR	5

Continued on next page

Table 7 – continued from previous page

ID	Description	Area	Sector	Tcode
288	It Hurall Persons(Ages 25 And Over)	IT	HUR	5
289	Es Hur All Persons (All Ages)	ES	HUR	5
290	Es Hur Femmes (Ages 16-24)	ES	HUR	5
291	Es Hur Femmes (All Ages)	ES	HUR	5
292	Es Hur Hommes (Ages 16-24)	ES	HUR	5
293	Es Hur Hommes (All Ages)	ES	HUR	5
294	Es Hur All Persons (Ages 16-24)	ES	HUR	5
295	Es Hurall Persons(Ages 25 And Over)	ES	HUR	5
296	Es Hur Females (Ages 25 And Over)	ES	HUR	5
297	Es Hur Males (Ages 25 And Over)	ES	HUR	5
298	De - Service Confidence Indicator	DE	SI	1
299	De Services - Buss. Dev. Past 3 Months	DE	SI	1
300	De Services - Evol. Demand Past 3 Months	DE	SI	1
301	De Services - Exp. Demand Next 3 Months	DE	SI	1
302	De Services - Evol. Employ. Past 3 Months	DE	SI	1
303	Fr - Service Confidence Indicator	FR	SI	1
304	Fr Services - Buss. Dev. Past 3 Months	FR	SI	1
305	Fr Services - Evol. Demand Past 3 Months	FR	SI	1
306	Fr Services - Exp. Demand Next 3 Months	FR	SI	1
307	Fr Services - Evol. Employ. Past 3 Months	FR	SI	1
308	Fr Services - Exp. Employ. Next 3 Months	FR	SI	1
309	Fr Services - Exp. Prices Next 3 Months	FR	SI	1