

A SIMPLE SOLUTION OF THE SPURIOUS REGRESSION PROBLEM

Wang, S. and C. Hafner

REPRINT | 2018 / 44

Cindy Shin-Huei Wang¹ / Christian M. Hafner²

A simple solution of the spurious regression problem

¹ National Tsing Hua University, Department of Quantitative Finance, Hsinchu City, Taiwan² Institut de statistique, biostatistique et sciences actuarielles, and CORE, Université catholique de Louvain, Voie du Roman Pays, 20, 1348 Louvain-la-Neuve, Belgium, E-mail: christian.hafner@uclouvain.be

Abstract:

This paper develops a new estimator for cointegrating and spurious regressions by applying a two-stage generalized Cochrane-Orcutt transformation based on an autoregressive approximation framework, even though the exact form of the error term is unknown in practice. We prove that our estimator is consistent for a wide class of regressions. We further show that a convergent usual t -statistic based on our new estimator can be constructed for the spurious regression cases analyzed by (Granger, C. W. J., and P. Newbold. 1974. "Spurious Regressions in Econometrics." *Journal of Econometrics* 74: 111–120) and (Granger, C. W. J., N. Hyung, and H. Jeon. 2001. "Spurious Regressions with Stationary Series." *Applied Economics* 33: 899–904). The implementation of our estimator is easy since it does not necessitate estimation of the long-run variance. Simulation results indicate the good statistical properties of the new estimator in small and medium samples, and also consider a more general framework including multiple regressors and endogeneity.

Keywords: autoregressive approximation, cointegration, generalized Cochrane-Orcutt estimation, spurious regression

JEL classification: C22, C53

DOI: 10.1515/snde-2015-0040

1 Introduction

The long-run relation between economic time series often plays an important role in macroeconomics and finance. Additionally, many macroeconomic and financial models imply that certain variables are cointegrated as defined by Engle and Granger (1987). Empirical tests, however, often fail to reject the null hypothesis of no cointegration, even for moderate sample sizes. One possible explanation of these test results is that the error term has a unit-root. For instance, the error term may contain a unit-root because of a nonstationary measurement error in one variable or nonstationary omitted variables (see Choi, Hu & Ogaki, 2008). In these cases, when the error term is a non-stationary $I(1)$ process but structural parameters can be recovered, the regression is called a structural spurious regression. Another related issue that is pervasive in the time series literature is the danger of obtaining spurious correlation findings. A spurious correlation occurs when a pair of independent series, each of them nonstationary or strongly autoregressive, are found apparently to be related according to standard inference in an OLS regression.

Since the seminal contribution by Yule (1926) on nonsense correlations between time series, it has been shown that spurious regression results may occur not only for pairs of independent unit root processes (see Granger and Newbold 1974, for a simulation study and Phillips 1986, for a theoretical explanation) but also for other persistent processes, such as $I(2)$ (Haldrup 1994), or even positively autocorrelated (stationary) autoregressive series (Granger, Hyung & Jeon, 2001). Some important applications of spurious regressions in economics and finance, although this list is by no means exhaustive, include Plosser, Schwert, and White (1982), Plosser and Schwert (1978), Ferson, Sarkissian, and Simin (2003), Hendry (1980), and Valkanov (2003).

The purpose of this paper is to propose a new estimator for cointegrating and spurious regressions that can successfully address both problems described above. More specifically, we propose a two-stage generalized Cochrane-Orcutt transformation estimator based on an autoregressive $AR(k)$ approximation framework (henceforth CO-AR estimator) that can cover a wide class of regression models. The order k of the AR approximation is assumed to grow with the sample size but at a slower rate. In practice, it can be chosen according to classical selection criteria such as AIC or BIC. This generalization of existing Cochrane-Orcutt transformations

in this framework, in particular that of Choi, Hu, and Ogaki (2008), is one of the main contributions of our paper. The CO-AR estimator turns out to be a robust procedure with respect to error specifications of unknown form even when the regressors and regressand are highly persistent (or possibly unit-root) processes.

Several papers have studied the problem of spurious regressions. Two recent contributions analyze an approach to correcting spurious regressions involving nonstationary regressors and error terms. Choi, Hu, and Ogaki (2008) propose the CHO-FGLS estimator that is consistent not just when the regression error is stationary but also when it is unit-root nonstationary.

Our estimator generalizes CHO-FGLS by approximating the unknown error term using an AR(k) model rather than AR(1). This implies that the CO-AR estimator can be successfully applied to a much wider class of regression models. In particular, unlike CHO-FGLS, the CO-AR estimator provides consistent and accurate estimation results when the error term follows a stationary (but potentially highly persistent) ARMA(p, q) process as well as a nonstationary ARIMA($p, 1, q$) process. The fact that the CO-AR estimator can cover general regression models is especially appealing in empirical work where the determination of the exact form of the error terms is often unknown, and the errors rarely follow a simple AR(1) process.

Our estimator differs from the CHO-FGLS estimator is that we use first differencing in the first step. This implies that the slope coefficients of the preliminary step are recovered using the GLS corrected estimator rather than the standard OLS regression. By doing so, the CO-AR estimator remains consistent, but is more robust than the CHO-FGLS when the innovations are highly persistent.

Deng (2014) considers the spurious regression issue when the regressors follow an ARMA(1,1) process where both the AR root and the MA root are close to unity, and derives a t test with a well defined distribution under the null by employing the fixed-bandwidth long run variance framework. Since the limiting distribution of Deng's t statistic critically depends on the long-run variance estimator, its implementation requires the selection of a suitable bandwidth parameter and kernel function. On the other hand, the CO-AR version of the t -statistic asymptotically follows the standard normal distribution and does not rely on a long-run variance estimator. Therefore the CO-AR t test can be easily implemented.

We now briefly describe the implementation of the two-stage CO-AR estimator. First, we start by taking the first difference of both the dependent and explanatory variables of our regression model, and use standard OLS to estimate the slope coefficients. Then, we compute residuals as the difference between the level of the dependent variable and the level of the explanatory variables multiplied by the estimated slope coefficients. Second, we fit an AR(k) model to these residuals, where the order k of the approximation is determined by minimizing the value of an information criterion. Berk (1974), Shibata (1980), and Bauer and Wagner (2008) show that an AR model can well approximate an $I(0)$ and an $I(1)$ process with increasing k . By doing so, we can filter the error term by an AR(k) model without the prior knowledge of the forms of the error term, and even if some $I(0)$ or $I(1)$ omitted variables are included in the error term. Third, we conduct a generalized Cochrane-Orcutt transformation (of order k) of both the dependent and independent variables. Finally, we run a standard OLS regression on the transformed variables.

The main findings of the paper can be summarized as follows. First, we derive the consistency and the limit distribution of the CO-AR estimator when a) the dependent, the explanatory variables and the innovations are stationary, but potentially highly persistent, b) the dependent, the explanatory variables and the innovations are non-stationary $I(1)$ series and c) the dependent and the explanatory variables are unit-root non-stationary series and the innovations are stationary. As a corollary, we show that the usual t -statistic of the slope coefficients constructed by our CO-AR estimation (CO-AR t -test) is convergent and follows the standardized Gaussian $N(0, 1)$ distribution, even when the regressors and regressand are highly persistent or possibly non-stationary processes. In sum, our proposed CO-AR estimator represents a new solution to the spurious regression problem, as examined in the influential articles of Granger and Newbold (1974) and Granger, Hyung, and Jeon (2001).

Second, we investigate the finite sample performance of our CO-AR estimator in a simulation exercise. The Monte Carlo experiments confirm our theoretical findings. We find that in spurious regressions the conventional significance tests based on standard least squares estimation and inference are seriously biased towards rejection of the null hypothesis of no relationship (and hence acceptance of a spurious relationship), even when the series are generated as statistically independent. The t -test constructed by our CO-AR estimator, however, remains approximately distributed as a standard Normal distribution without using the long-run variance framework, even in samples as small as 50 observations. Hence, the simulation results indicate the size control and the power pattern of our methodology are excellent even in small samples. On the other hand, our CO-AR t -test can be viewed as a new test to avoid the over-rejection of the null hypothesis and solve the spurious regression problem. We also find that in cointegrating relationships the CO-AR is as good as the other three conventional estimators, dynamic OLS (Saikkonen, 1991; Stock & Watson, 1993), GLS Corrected, and CHO-FGLS (Choi, Hu & Ogaki, 2008) when the innovation follows either a white noise or an AR(1) process, and significantly outperforms the FGLS when the innovation follows a stationary ARMA process or nonstationary ARIMA dynamics.

The rest of the paper is organized as follows. Section 2 starts by establishing the asymptotic properties of the new CO-AR estimator for cointegrating and spurious regressions, and constructing a convergent usual t -statistic in spurious regressions. Section 3 investigates the finite sample performance of the CO-AR estimator through Monte Carlo experiments. Section 4 contains some concluding comments. An Appendix provides proofs of the results given in the paper.

2 The statistics and main results

The objective of this section is to establish the asymptotic properties of the CO-AR estimator in both spurious and cointegrating regressions, and to construct a convergent usual t -statistic in spurious regressions as analyzed by Granger and Newbold (1974) and Granger, Hyung, and Jeon (2001).

2.1 Stationary regressions

Consider the regression Model I,

$$y_t = \beta x_t + u_t$$

where the regressor x_t and the error term u_t are independent and follow ARMA processes as defined by the following Assumption 1. For the sake of notational simplicity we do not include a constant in the regression and assume that x_t is univariate. The independence assumption is strong as it implies, among other things, absence of endogeneity. We will discuss ways to generalize our framework to multiple regressors and endogeneity in Section 2.4.

Assumption 1

The processes x_t and u_t follow independent ARMA(p, q) processes of the form (z_t is either x_t or u_t)

$$\phi(L)z_t = \theta(L)e_t$$

where (i) s the autoregressive (AR) and moving average (MA) polynomials $\phi(\cdot)$ and $\theta(\cdot)$ in the lag operator L are assumed to have all roots outside the unit circle; (ii) $\phi(\cdot)$ and $\theta(\cdot)$ have no common roots; (iii) e_t is an i.i.d. process with $E[e_t] = 0$, $E[e_t^2] = \sigma_e^2$, and $E[e_t^4] < \infty$.

Note that the lag polynomials of the models for x_t and u_t can be different and, in particular, can be of different order. Assumption 1 guarantees that the conditions of Theorem 2 of Berk (1974) hold and allows to represent the ARMA process u_t as

$$u_t = \sum_{j=1}^{\infty} b_j u_{t-j} + e_t$$

where $\sum_{j=1}^{\infty} |j|^s |b_j| < \infty$, $s > 1$, and to approximate u_t by a fitted AR(k) process as

$$u_t = \sum_{j=1}^k \hat{b}_j u_{t-j} + \hat{e}_{tk} + o_p(1)$$

with $k \rightarrow \infty$. The fact that we use an AR(k) approximation of increasing order is the key feature of our procedure and allows us to approximate a much wider class of serially correlated error processes than for example the AR(1) approximation used by Choi, Hu, and Ogaki (2008). As a consequence, we expect the new procedure to provide more efficient parameter estimates as well as parameter tests that are more robust with respect to wider patterns of error autocorrelation. Section 3 will investigate this idea via a detailed simulation study.

The following assumption gives conditions on the choice of k as the sample size increases.

Assumption 2

The order $k = k_T$ is chosen such that

- a. $k \rightarrow \infty$ as $T \rightarrow \infty$,
- b. $k = o(T^{1/3})$, and
- c. $\sqrt{T} \sum_{j=k+1}^{\infty} |b_j| \rightarrow 0$, as $T \rightarrow \infty$.

Assumption 2(a) requires that the order of the fitted AR(k) model increases as the sample size increases. Assumptions 2(b) and 2(c) give upper and lower bounds on the rate at which k increases, based on the conditions of Berk (1974). The condition 2(c) is a lower bound since k has to increase sufficiently fast such that $\sum_{j=k+1}^{\infty} |b_j| = o(T^{-1/2})$. Shibata (1980) and Ing and Wei (2003) show that a model which minimizes the mean squared prediction error is of order $k^* = O(\log T)$, $1 \leq k^* \leq k_T$.

We are now in the position to introduce our two-stage Cochrane-Orcutt GLS estimator as follows:

Stage 1. We start by taking the first difference of the dependent and explanatory variables, and compute the standard OLS estimate of the coefficients β , denoted as $\hat{\beta}_{GLSC}$, of the following model,

$$\Delta y_t = \beta \Delta x_t + \Delta u_t$$

where $\Delta = 1 - L$ denotes the difference operator. In the case of I(1) processes y_t and x_t , to be discussed in the following sections, this procedure can be viewed as a GLS corrected estimation as described by Choi, Hu, and Ogaki (2008). Note that when x_t and u_t are uncorrelated, the estimator $\hat{\beta}_{GLSC}$ is consistent.

Stage 2. Construct the series of fitted residuals $\hat{u}_t = y_t - \hat{\beta}_{GLSC} x_t$. Then, approximate the errors $\hat{u}_t$ by a finite order AR(k) model, i.e. $\hat{u}_t = \sum_{j=1}^k \hat{b}_j \hat{u}_{t-j} + \hat{\varepsilon}_{tk}$. Subsequently, conduct the following Cochrane-Orcutt transformation of the variables x_t and y_t :

$$\tilde{y}_t = y_t - \sum_{j=1}^k \hat{b}_j y_{t-j}, \quad \tilde{x}_t = x_t - \sum_{j=1}^k \hat{b}_j x_{t-j}.$$

Consider OLS estimation of the regression

$$\tilde{y}_t = \beta \tilde{x}_t + \tilde{u}_t$$

where $\tilde{u}_t = \hat{\varepsilon}_{tk} + o_p(1)$ is an asymptotically uncorrelated error term. The OLS estimator is computed as

$$\hat{\beta}_{CO-AR} = \left(\sum_{t=k+1}^T \tilde{x}_t^2 \right)^{-1} \sum_{t=k+1}^T \tilde{x}_t \tilde{y}_t.$$

Under our condition that x_t and u_t are independent ARMA processes, the following theorem gives the asymptotic properties of the CO-AR estimator of β .

Theorem 1

If the data generating process satisfies Model 1, x_t and u_t are independent ARMA processes satisfying Assumption 1, then, under Assumption 2,

$$\hat{\beta}_{CO-AR} = \beta + O_p(T^{-1/2}).$$

Theorem 1 includes the particular case $\beta = 0$, in which case x_t and y_t are independent. We obtain the following corollary.

Corollary 1

If x_t and y_t are independent ARMA processes satisfying Assumption 1, then under Assumption 2, $\hat{\beta}_{CO-AR} = O_p(T^{-1/2})$

Theorem 1 shows that the CO-AR estimator of β is consistent and converges at rate $T^{1/2}$ when the data generating process satisfies Model I. The corollary indicates that when an ARMA process is regressed on another, independent ARMA process, the CO-AR process remains consistent at rate $T^{1/2}$. This asymptotic result improves the previous studies of Deng (2014) where $\hat{\beta} = O_p(1)$. Simply speaking, the spirit of our methodology

follows the suggestion of Granger, Hyung, and Jeon (2001) saying that the proper reaction to having a possible spurious relationship is to add lagged dependent and independent variables until the errors appear to be white noise.

We now construct the usual t -statistic to test the hypothesis $H_0: \beta = 0$ by using the CO-AR estimate of β and the corresponding standard error $S_{\hat{\beta}} = (\hat{\sigma}_{\tilde{u}}^2 / \sum_{t=k+1}^T \tilde{x}_t^2)^{1/2}$, where $\hat{\sigma}_{\tilde{u}}^2 = (T - k)^{-1} \sum_{t=k+1}^T \tilde{u}_t^2$ and $\tilde{u}_t = \tilde{y}_t - \hat{\beta}_{\text{CO-AR}} \tilde{x}_t$. The asymptotic properties of the t -statistic are given in the following theorem.

Theorem 2

If x_t and y_t are independent ARMA processes satisfying Assumption 1, then, under Assumption 2,

$$t_{\beta} = \frac{\hat{\beta}_{\text{CO-AR}}}{S_{\hat{\beta}}} \rightarrow_{\mathcal{L}} N(0, 1).$$

The important implication of this theorem is that we can build a convergent t -statistic t_{β} that, under the null hypothesis, is asymptotically normally distributed without having to estimate the long-run variance. The associate finite sample performance of the statistic will be analyzed in Section 3.

2.2 Spurious regressions with $I(1)$ processes

Many macroeconomics models imply that certain variables are cointegrated as defined by Engle and Granger (1987). However, cointegration tests often fail to reject the null hypothesis of no cointegration for these variables. Choi, Hu, and Ogaki (2008) pointed out that one possible explanation for these empirical results is that the error is unit-root non-stationary due to a non-stationary measurement error in one variable or non-stationary omitted variables. Recall that a regression with nonstationary stochastic errors is defined “spurious” in the time series literature. In this section, we therefore consider the general case where both x_t and u_t are integrated of order one.

Consider the following Model (II):

$$y_t = \beta x_t + u_t$$

where x_t and u_t are independent $I(1)$ processes. Again, we note that the independence assumption is strong, but refer to Section 2.4 for possible extensions. We make the following assumption about the dynamics of x_t and u_t .

Assumption 3

The processes x_t and u_t follow independent ARIMA($p, 1, q$) processes of the form (z_t is either x_t or u_t)

$$\phi(L)(1 - L)z_t = \theta(L)e_t,$$

where (i) the autoregressive (AR)-and moving average (MA)-polynomials $\phi(\cdot)$ and $\theta(\cdot)$ are assumed to have all roots outside the unit circle; (ii) $\phi(\cdot)$ and $\theta(\cdot)$ have no common roots; (iii) e_t is an i.i.d. process with $E(e_t) = 0$, $E(e_t^2) = \sigma_e^2$ and $E(e_t^4) < \infty$.

Again, the lag polynomials can be different for x_t and u_t . By applying the CO-AR estimating procedure, one obtains a $\sqrt{T}$ -consistent first stage estimator $\hat{\beta}_{\text{GLSC}}$ and residuals $\hat{u}_t = y_t - \hat{\beta}_{\text{GLSC}} x_t$. Then, the CO-AR estimator of β is the OLS estimator of the regression $\tilde{y}_t = \beta \tilde{x}_t + \tilde{u}_t$. We summarize the asymptotic properties of the CO-AR estimate of β in Model (II) in the next theorem.

Theorem 3

If the data generating processes satisfy Model (II) and Assumption 3, then, under Assumption 2, $\hat{\beta}_{\text{CO-AR}} - \beta = O_p(T^{-1/2})$.

Theorem 3 not only implies that the CO-AR estimator is consistent (while the OLS estimator is inconsistent), but also is more general when compared to the corrected GLS and CHO-FGLS estimators (see Choi, Hu & Ogaki, 2008). On the one hand, the corrected GLS is the ideal solution for the spurious regression problem when the error term follows a random walk process. However if the error term does not follow a pure random walk process, differencing the data can result in a misspecified model. On the other hand, the CHO-FGLS estimator has promising performance for regressions with a stationary AR(1) error term. However, in empirical work the

exact autocorrelation structure of the error terms is often unknown, and the errors rarely follow a simple AR(1) process. Therefore in practice both the GLS and the CHO-FGLS estimators should be used with care. Instead, the performance of the CO-AR estimator is very good not only for stationary AR(1) or random walk error terms but also for more general innovations, such as stationary ARMA(p, q) processes as well as nonstationary ARIMA($p, 1, q$) processes.

Similar to the stationary case, Theorem 3 also includes the particular case $\beta = 0$, in which case x_t and y_t are independent. We obtain the following Corollary.

Corollary 2

If x_t and y_t are independent $I(1)$ processes satisfying Assumption 3, then, under Assumption 2, $\beta_{\text{CO-AR}} = O_p(T^{-1/2})$.

Thus, $\beta_{\text{CO-AR}}$ is consistent in the important case of spurious regressions with integrated processes. Based on this result, we next show that a convergent t -statistic for β can be constructed for the case of spurious regressions.

Since the Monte Carlo study by Granger and Newbold (1974) and the asymptotic theory developed by Phillips (1986) it is well known that the usual t -statistic for a regression between independent $I(1)$ processes does not have a limiting distribution but diverges at the rate of $T^{1/2}$ as the sample size T increases. In this section, in a more general framework, we establish a convergent standard t -statistics by using the CO-AR estimation for a regression between two independent $I(1)$ processes. In effect, we allow y_t and x_t to be rather general integrated processes whose first differences are weakly dependent and possibly heterogeneously distributed innovations. This includes a wide variety of data-generating mechanisms, such as the ARIMA ($p, 1, q$) model. The following theorem shows that by using the CO-AR estimator, a convergent t -statistic can be constructed and is asymptotically normal distributed for a regression between two independent $I(1)$ processes.

Theorem 4

Assume that x_t and y_t are independent $I(1)$ processes satisfying Assumption 3, then, under Assumption 2, $t_\beta = \frac{\hat{\beta}_{\text{CO-AR}}}{s_\beta} \rightarrow_{\mathcal{L}} N(0, 1)$.

Note also that using the CO-AR methodology to construct a convergent standard t -statistic avoids the difficult issues of choosing a suitable bandwidth parameter and a kernel function for long run variance estimation.

2.3 Cointegrating regressions

In this section we consider the asymptotic distribution of the CO-AR estimator under the assumption of cointegration, i.e. the error term u_t in Model (II) is an $I(0)$ process.

Theorem 5

If the data generating processes x_t and y_t follow Model (II) but u_t is an $I(0)$ process satisfying Assumption 1, x_t and u_t are independent processes, then, under Assumption 2, $\hat{\beta}_{\text{CO-AR}} - \beta = O_p(T^{-1})$ and

$$T(\hat{\beta} - \beta) \rightarrow_d \left(\int_0^1 V(r)^2 dr \right)^{-1} \int_0^1 V(r) dW(r)$$

where $V(r)$ is a Brownian motion with variance $\sum_{j=-\infty}^{\infty} E[v_t v_{t-j}]$, $v_t = \Delta x_t$, and $W(r)$ is a Brownian motion with variance $\sum_{j=-\infty}^{\infty} E[u_t u_{t-j}]$.

Theorem 5 shows that the CO-AR estimation method is also useful for cointegration analysis. For instance, in a cointegration model, the OLS estimator is consistent at the convergence rate T , but if the error term follows a stationary ARMA (p, q) process, the OLS estimator may cause inaccurate estimation results, similar to the stationary case investigated by Granger, Hyung, and Jeon (2001). In large samples, CO-AR and OLS behave similarly under cointegration, and the corresponding asymptotic distributions are identical and given by Theorem 5. However, the lag selection procedure of the CO-AR estimator offers additional flexibility in small samples.

Again, our results use the strong assumption of independence between u_t and x_t . We now turn to possible generalizations in the next section.

2.4 Multiple regressors and endogeneity

We consider two extensions of the proposed procedure, the case of multiple explanatory variables, and that of possible endogeneity.

Consider the following regression

$$y_t = \beta' x_t + u_t$$

where x_t is now a $(p \times 1)$ vector of explanatory variables, and $\beta = (\beta_1, \dots, \beta_p)$ a vector of coefficients.

Suppose now that the error term u_t is not necessarily uncorrelated with each of the explanatory variables, and that we have a set of instruments $w_t = (w_{1t}, \dots, w_{pt})'$ which are uncorrelated with u_t but correlated with the respective explanatory variable. If a particular x_{it} is uncorrelated with u_t we simply set $w_{it} = x_{it}$. Define the $(T \times p)$ matrices $X = (x_1, \dots, x_T)'$ and $W = (w_1, \dots, w_T)'$, the $(T-1 \times p)$ matrices $\Delta X = (\Delta x_2, \dots, \Delta x_T)'$ and $\Delta W = (\Delta w_2, \dots, \Delta w_T)'$, the $(T \times 1)$ vector $Y = (y_1, \dots, y_T)'$ and the $(T-1 \times 1)$ vector $\Delta Y = (\Delta y_2, \dots, \Delta y_T)'$.

In the first stage of the CO-AR procedure above, the regression $\Delta y_t = \beta' \Delta x_t + \Delta u_t$ is now estimated by the instrumental variable estimator $\hat{\beta}_{IV} = (\Delta W' \Delta X)^{-1} \Delta W' \Delta Y$, which remains consistent unlike the usual OLS or GLS estimators.

In the second stage, as above we obtain fitted residuals after the first stage, then fit an $AR(k)$ model and apply a Cochrane-Orcutt transformation to the series x_t and y_t to obtain the transformed $\tilde{x}_t$ and $\tilde{y}_t$. The regression

$$\tilde{y}_t = \beta' \tilde{x}_t + \tilde{u}_t$$

is then estimated by the IV estimator

$$\hat{\beta}_{IV\text{-COAR}} = (\tilde{W}' \tilde{X})^{-1} \tilde{W}' \tilde{Y}$$

where $\tilde{X} = (\tilde{x}_1, \dots, \tilde{x}_T)'$, $\tilde{W} = (\tilde{w}_1, \dots, \tilde{w}_T)'$ and $\tilde{Y} = (\tilde{y}_1, \dots, \tilde{y}_T)'$.

In the spurious regression case, the asymptotic covariance matrix of $\hat{\beta}_{IV\text{-COAR}}$ can be consistently estimated using

$$S = \hat{\sigma}_{\tilde{u}}^2 (\tilde{W}' \tilde{X})^{-1} (\tilde{W}' \tilde{W}) (\tilde{X}' \tilde{W})^{-1}$$

where $\hat{\sigma}_{\tilde{u}}^2 = T^{-1} \sum_t (y_t - \hat{\beta}_{IV\text{-COAR}}' x_t)^2$, and we can construct usual Wald-type statistics to test the hypothesis $H_0: \beta = \beta_0$ as

$$(\hat{\beta}_{IV\text{-COAR}} - \beta_0)' S^{-1} (\hat{\beta}_{IV\text{-COAR}} - \beta_0),$$

which, under H_0 , has an asymptotic χ^2 distribution with p degrees of freedom. As before, we can also construct individual tests of hypotheses $H_0: \beta_j = \beta_{j0}$, $j = 1, \dots, p$, using the t -statistic

$$t_{\beta} = \frac{\hat{\beta}_{j, IV\text{-COAR}} - \beta_{j0}}{\sqrt{S_{jj}}}$$

where S_{jj} is the j -th element on the diagonal of S . Under H_0 this statistic has an asymptotic standard normal distribution. We will provide additional simulation evidence in the next section to investigate the performance of our estimator in this more general setting.

3 Simulation results

In this section we use simulations to show, for several typical cases, that the test-rejection probabilities of the CO-AR estimator are close to the nominal levels in small and medium samples. Moreover, we analyze the finite-sample properties of our CO-AR estimator in cointegrating relationships with serially correlated errors and spurious regressions, and compare its performance to the other estimators (OLS, GLSC and CHO-FGLS) discussed in Choi, Hu, and Ogaki (2008).

3.1 Stationary and spurious regressions

We study the finite sample performance of the CO-AR t -statistic compared to the t -statistic of the standard least squares estimation and inference (OLS). By adopting the same experimental design as Granger, Hyung, and Jeon (2001), suppose that X_t and Y_t are generated by the following independent processes:

- DGP 1: $x_t = e_t$
- DGP 2: $x_t = 0.95x_{t-1} + e_t$
- DGP 3: $x_t = 0.99x_{t-1} + e_t$
- DGP 4: $(1 - 0.9L)(1 - L)x_t = (1 + 0.5L)e_t$
- DGP 5: $(1 - L)^2x_t = e_t$

where x_t stands for either X_t or Y_t , and e_t is drawn from independent $N(0, 1)$ populations. In words, X_t or Y_t are generated respectively by a white noise process (DGP 1), a strongly positively autocorrelated autoregressive series (DGP 2 and 3), an ARIMA(1,1,1) process (DGP 4) and a nonstationary I(2) process (DGP 5). The number of iterations in each simulation is 20,000. To avoid the problem of fixing X_0 and Y_0 , in each replication, the first 100 observations are discarded, and $X_{-100} = Y_{-100} = 0$. We consider sample sizes of 50, 100 and 500 observations. The lag length of the AR(k) approximation of the error term is determined by applying the Bayesian information criterion (BIC) to each replication with k ranging from 0 to some maximal order k_{max} , which is set respectively at 3 (when the sample consists of 50 observations), 4 (when the sample is 100), and 8 (when the sample is 500).

Table 1 shows the percentage of rejection of the null hypothesis of no linear relationship between Y_t and X_t at the 5% critical value, i.e. absolute value of the t -statistic greater than 1.96, using both OLS (left) and the CO-AR estimator (right). The series X_t and Y_t are generated independently employing DGP 1 to 5. First, the t -test based on the OLS estimate is well-behaved as long as either X_t or Y_t are white noise processes. Second, as soon as both X_t and Y_t display serial correlation over time, the OLS t -test is spuriously biased towards rejection. This becomes very serious when both series have strong temporal dependence. Moreover, the percentage of spurious relationships tends to increase with the sample size. Third, the empirical size of the t -test based on the CO-AR estimator is very close to the theoretical size of 5%. Put differently, our estimator can control almost perfectly for the spurious regression problem, especially for moderate and large sample sizes.

Table 1: Spurious regression with Normal distributions. The table displays the percentage of rejection, i.e. absolute value of t -value greater than 1.96. The order of the AR approximation is selected using BIC. The maximum number of lags to be used in the selection procedure is set respectively at 3 (when $n = 50$), 4 (when $n = 100$), 8 (when $n = 500$).

			Percentage of $ t > 1.96$				
T = 50	OLS	X_1	Y_1	Y_2	Y_3	Y_4	Y_5
		X_2	5.2	5.4	5.7	5.5	5.8
T = 50	CO-AR	X_3	5.7	56.5	60.5	61.6	61.8
		X_4	5.8	60.1	65.0	65.5	66.3
		X_5	5.5	60.2	65.5	67.2	67.5
		X_1	5.5	62.5	67.8	68.1	68.4
		X_2	5.3	6.3	6.8	6.9	7.0
		X_3	5.8	6.3	7.0	6.6	6.6
		X_4	5.7	6.3	6.0	6.7	6.6
		X_5	5.8	6.1	6.0	6.3	6.4
		X_1	5.9	6.2	6.1	6.2	6.0
		X_2	5.2	5.2	5.5	5.5	5.6
T = 100	CO-AR	X_3	5.4	62.5	66.8	61.6	62.6
		X_4	5.3	67.3	72.9	65.5	67.8
		X_5	5.4	67.3	75.3	67.2	68.2
		X_1	5.4	68.5	76.1	68.1	69.4
		X_2	5.1	5.5	5.6	6.1	6.2
		X_3	4.9	5.9	5.1	6.0	6.1
		X_4	5.2	5.7	5.0	6.0	6.2
		X_5	5.6	5.6	5.0	5.6	5.5
		X_1	5.7	5.7	5.1	5.4	5.1
		X_2	4.8	5.1	5.0	5.0	5.0
T = 500	OLS	X_3	5.2	65.5	72.4	73.8	74.1
		X_4	4.9	72.2	84.4	85.8	86.4
		X_5					

CO-AR	X_4	4.8	73.7	85.5	89.6	89.3
	X_5	5.0	74.2	85.7	89.9	90.1
	X_1	4.9	4.9	3.9	4.1	4.0
	X_2	5.1	5.0	5.5	5.6	5.2
	X_3	4.7	5.2	5.4	5.1	5.3
	X_4	4.9	4.8	4.9	5.2	5.4
	X_5	5.0	4.9	5.1	5.3	5.2

We have compared the size performance of the CO-AR t -test with that of setting the AR order $k = 1$ in all regressions. In the vast majority of cases, except for those where the true error process was AR(1), the size of the CO-AR test was closer to the nominal size. The CHO-FGLS test happens to be severely biased in sufficiently large samples in situations where the serial correlation structure of the error term is more complicated than AR(1). Detailed results are available upon request.

To assess the robustness of the previous conclusions, we also use different distributions for the error terms, in particular a Student- t -with five degrees of freedom and a Laplace distribution. All gave very similar results compared to Table 1.

We further consider the power performance of the CO-AR t -test by considering two cross-correlated time series $\{Y_t\}$ and $\{X_t\}$. We assume that x_t and y_t are generated with $\rho_{xy}(j) = 0.2$ for $j = 0$ and $\rho_{xy}(j) = 0$ for $j \neq 0$ where $\rho_{xy}(j)$ denotes the cross correlation function of x_t and y_t at lag j . All power estimates are above 60% at a nominal size of 5%. Detailed results are not reported but available upon request.

3.2 Finite sample performance of the four estimators

Consider the following regression model:

$$y_t = \beta x_t + \gamma' v_t + e_t \quad (1)$$

where $v_t = (\Delta x_{t-k}, \dots, \Delta x_t)'$, and $\gamma = (\gamma_0, \dots, \gamma_k)'$, and x_t is integrated of order 1, i.e. I(1). Note that the inference procedure about the coefficients β substantially differs according to the different assumptions on the error term e_t . When the error term e_t is stationary, the equation (1) is a cointegrating regression with potentially serially correlated error. When the error term e_t is a unit-root process, the equation (1) is a spurious regression.

Choi, Hu, and Ogaki (2008) discuss three methods to estimate the structural parameters β :

- (1) Dynamic Ordinary Least Squares (DOLS): Regress y_t on x_t and v_t to get $\hat{\beta}_{\text{DOLS}}$.
- (2) Corrected Generalized Least Squares (GLSC): Regress Δy_t on Δx_t and Δv_t to get $\hat{\beta}_{\text{GLSC}}$.
- (3) Feasible Generalized Least Squares (FGLS): Let the residual from the DOLS regression be denoted by $\hat{e}_t$.

Then run the following AR(1) regression $\hat{e}_t = \rho \hat{e}_{t-1} + u_t$, and compute the OLS coefficient $\hat{\rho}$. Apply the Cochrane-Orcutt transformation to the data:

$$\hat{y}_t = y_t - \hat{\rho} y_{t-1}, \quad \hat{x}_t = x_t - \hat{\rho} x_{t-1}, \quad \hat{v}_t = v_t - \hat{\rho} v_{t-1}$$

Finally, regress $\hat{y}_t$ on $\hat{x}_t$ and $\hat{v}_t$ to get $\hat{\beta}_{\text{CHO-FGLS}}$.

This section compares the finite sample properties of $\hat{\beta}_{\text{CO-AR}}$ with those of $\hat{\beta}_{\text{DOLS}}$, $\hat{\beta}_{\text{GLSC}}$ and $\hat{\beta}_{\text{CHO-FGLS}}$. For the ease of comparability, we adopt the same experimental design as Choi, Hu, and Ogaki (2008). More specifically, in the simulation we generate v_t and u_t from two independent standard normal distributions. The structural parameter is set to $\beta = 2$ and $\gamma' v_t = 0.5 v_t$. The error process e_t is specified as follows:

- DGP a: $e_t = u_t$
- DGP b: $e_t = 0.95 e_{t-1} + u_t$
- DGP c: $e_t = e_{t-1} + u_t$
- DGP d: $e_t = 0.5 e_{t-1} + 0.5 e_{t-2} + u_t$
- DGP e: $\Delta e_t = 0.95 \Delta e_{t-1} + u_t$

where $u_t \sim N(0, 1)$. In words, the error term is a white noise (DGP a), an AR(1) process with autoregressive coefficient set to 0.95 (DGP b), a random walk (DGP c), an AR(2) that has a unit root (DGP d), and an ARIMA(1,1,0) (DGP e). Note that the first two cases correspond to a cointegrating relationship. The number of iterations in each simulation is 5000, and in each replication 100 + n observations are generated ($n = 50, 100$, and 500), of which the first 100 observations are discarded. The lag length of the AR(k) approximation of the error term u_t is determined by applying the Bayesian information criterion to each replication. The maximum number of lags to be used in selection procedure is set respectively at 3 (when $n = 50$), 4 (when $n = 100$), 8 (when $n = 500$). Table 2 shows the bias and the root mean square error (RMSE) of all four estimators. As we would expect, the DOLS estimator is the best one when the error process is DGP a, while the GLS Corrected is the best estimator when the error process is DGP c. As shown by Choi, Hu, and Ogaki (2008), the CHO-FGLS estimator is almost as good as the DOLS estimator in cointegrating relationships and significantly outperforms the DOLS estimator in spurious regressions. Hence, the FGLS estimation is a robust procedure with respect to error specifications.

Table 2: The bias and RMSE of the four estimators. The order of the AR approximation is selected using BIC. The maximum number of lags to be used in the selection procedure is set respectively at 3 (when $n = 50$), 4 (when $n = 100$), 8 (when $n = 500$). The number of replications is 5000.

	T	DOLS		GLSC		CHO-FGLS		CO-AR	
		Bias	rmse	Bias	rmse	Bias	rmse	Bias	rmse
(a)	50	0.0005	0.047	−0.001	0.210	0.0007	0.048	−0.0002	0.083
	100	0	0.023	−0.003	0.144	0.0001	0.023	−0.0008	0.045
	500	0	0.004	−0.0003	0.063	0	0.004	0	0.012
(b)	50	−0.003	0.512	−0.004	0.180	−0.004	0.232	−0.003	0.181
	100	0.001	0.316	−0.004	0.128	−0.004	0.142	−0.004	0.124
	500	−0.000	0.083	−0.000	0.061	−0.000	0.048	−0.000	0.049
(c)	50	0.012	2.175	−0.004	0.180	−0.009	0.639	−0.004	0.185
	100	0.007	1.674	−0.005	0.127	−0.006	0.366	−0.0051	0.130
	500	−0.023	1.129	−0.000	0.061	−0.002	0.113	−0.000	0.061
(d)	50	0.008	1.446	−0.003	0.170	−0.008	0.500	−0.006	0.207
	100	0.005	1.114	−0.004	0.116	−0.003	0.323	−0.004	0.108
	500	−0.016	0.752	−0.000	0.051	−0.003	0.138	−0.000	0.047
(e)	50	0.305	38.92	−0.006	0.743	−0.064	10.47	−0.002	0.198
	100	0.124	30.72	−0.0112	0.553	−0.015	5.608	−0.001	0.135
	500	−0.459	21.91	−0.0059	0.267	−0.023	1.088	−0.000	0.062

The CO-AR estimator is robust with respect to the order of integration of the error terms. If the error follows a white noise process, the RMSE of CO-AR is similar to the CHO-FGLS and DOLS estimators. If the error follows a highly persistent AR(1) process (DGP b) or a unit-root process (DGP c), the RMSE of the CO-AR tends to be slightly smaller than the CHO-FGLS and GLS corrected estimators, and much smaller than the DOLS procedure. Not surprisingly, when the error process follows complex unit-root dynamics (cases DGP d and DGP e), the CO-AR estimator performs very well compared to all the three other estimators proposed by Choi, Hu, and Ogaki (2008), especially for large sample sizes.

The adequacy of an approximate model for the DGP of the error term depends on the choice of the order of the AR approximation. In particular, different lag selection methods for the order of the AR approximation may potentially lead to drastically different conclusions regarding the finite sample performance of the CO-AR estimator. Therefore, we also investigate the sensitivity of our Monte Carlo results to the choice of the lag selection criterion by applying the Akaike information criterion (AIC), the Hannan-Quinn information criterion (HQIC), and by using the “true” order of the approximation (i.e. zero lags for the error process specified in DGP a, one lag for DGP b and DGP c, and two lags for DGP d and DGP e). To economize on space we do not report the results in detail, but rather discuss some important aspects. Overall we find that the results are broadly in line with the findings displayed in Table 2. The performance of the CO-AR estimator is remarkably robust with respect to the different information criteria. However, when the error term follows a white noise process, the RMSE of the CO-AR estimator using the “true” lag length is halved compared to the BIC lag selection. Moreover, the performance of the BIC is broadly similar to the HQIC, and both criteria slightly dominate the AIC lag selection in terms of the RMSE metric.

3.3 Multiple regressors and endogeneity

We finally consider the case of two regressors, one of which is correlated with the error term, to see how our procedure performs under endogeneity. We generate 5000 replications of the following processes:

$$\begin{aligned}y_t &= 0.6x_{1t} + 0.6x_{2t} + u_t \\x_{1t} &= 0.95x_{1,t-1} + 0.5\varepsilon_{1,t-1} + \varepsilon_{1t} \\x_{2t} &= 0.7x_{2,t-1} + 0.3\varepsilon_{2,t-1} + \varepsilon_{2t} \\u_t &= 0.8u_{t-1} + 0.2\varepsilon_{3,t-1} + \varepsilon_{3t}.\end{aligned}$$

The iid error terms ε_{it} are drawn from a standard normal distribution, but where the correlation between ε_{1t} and ε_{3t} is 0.5. We assume the presence of an instrumental variable w , also iid and drawn from a standard normal distribution, which has a correlation of 0.8 with ε_{1t} . To summarize, the vector $(\varepsilon_{1t}, \varepsilon_{2t}, \varepsilon_{3t}, w_t)'$ is drawn from a Gaussian distribution with mean zero and covariance matrix Σ given by

$$\Sigma = \begin{pmatrix} 1 & 0 & 0.5 & 0.8 \\ 0 & 1 & 0 & 0 \\ 0.5 & 0 & 1 & 0 \\ 0.8 & 0 & 0 & 1 \end{pmatrix}$$

In Table 3 we report the results of the t -statistic for testing $H_0: \beta_1 = 0.6$ and $H_0: \beta_2 = 0.6$. While the test is slightly over-sized in small samples, it performs well in moderate and large samples, and the empirical size is not significantly different from the nominal size.

Table 3: Size of the t -tests and RMSE of the estimation error in the case of two regressors and partial endogeneity. The lag order k was chosen according to the AIC criterion. The number of replications is 5000.

	T	50	100	200	500	1000
β_1	size	7.56%	6.74%	5.78%	5.54 %	5.22 %
	RMSE	9.27%	2.02 %	0.9%	0.33 %	0.16%
β_2	size	10.68%	8.74%	6.36%	5.80%	5.02%
	RMSE	4.70%	1.20%	0.54%	0.21%	0.10%

To sum up, this paper develops a new robust estimator for structural parameters in dynamic regressions. The simulation study indicates that the CO-AR estimator is particularly useful in most applied situations in macroeconomics and finance where the researcher does not know the exact form of the error term.

4 Concluding remarks

This paper proposes a new estimation method, namely CO-AR estimation, for cointegrating and spurious regressions by applying a two-stage generalized Cochrane-Orcutt transformation based on the autoregressive approximation framework developed by Berk (1974). We prove that our CO-AR estimator is consistent. We further show that a convergent usual t -statistic based on the CO-AR estimator (the CO-AR t test), asymptotically distributed as $N(0, 1)$, can be constructed for the spurious regression cases analyzed by Granger and Newbold (1974) and Granger, Hyung, and Jeon (2001). More importantly, the CO-AR t test does not rely on the long-run variance estimator, and therefore it can be easily implemented. Moreover, the CO-AR estimation method turns out to be a robust procedure with respect to error specifications even when the regressors and regressand are highly persistent (or possibly unit-root) processes. The fact that the CO-AR estimator can cover general regression models is especially appealing in empirical work where the determination of the exact form of the error terms is often unknown, and the errors rarely follow a simple AR(1) process. Finally, the simulation results indicate that the finite sample performance of the CO-AR methodology is promising even if the sample size is as small as 50 observations.

Of course, several important issues are not considered in this paper and deserve further study. For instance, in future work it will be important to investigate the choice of k , the order of the AR model for approximating I(1) processes, allowing alternative selection criteria, such as the modified AIC and BIC developed in Bauer

and Wagner (2008), or a different maximum number of lags to be used in the selection procedure. Moreover, we believe that the AR approximation framework could be applied to the issue of the spurious regressions involving other persistent processes, such as I(2) series.

Acknowledgement

We appreciate useful comments of Cheng Hsiao, Yong-Miao Hong, Wolfgang Härdle, Hashem Pesaran, Mark Watson and all participants at the Sir Clive Granger Memorial Conference, Econometrics Workshop at Humboldt University Berlin, and European Econometrics Meeting, Oslo.

A Appendix

A.1 Proof of Theorem 1

First note that $\hat{\beta}_{GLSC}$ is consistent under our assumptions, see e.g. Robinson (1998). Then $\hat{u}_t = y_t - \hat{\beta}_{GLSC}x_t = u_t - (\hat{\beta}_{GLSC} - \beta)x_t = u_t + o_p(1)$. By Theorem 2 of Berk (1974), we note that an ARMA (p, q) model represented by an infinite autoregression can be approximated by an AR(k) model, where $k = k_T$ and where the autoregressive coefficients satisfy mild summability constraints. Thus, $\hat{u}_t$ can be approximated by an AR(k) model, i.e. $\hat{u}_t = \sum_{j=1}^k \hat{b}_j \hat{u}_{t-j} + \hat{e}_{tk}$ and $\|\hat{b}(k) - b(k)\| = O_p(T^{-1/2})$ as $k = o(T^{1/3})$, where $\hat{b}(k) = (\hat{b}_1, \dots, \hat{b}_k)$ is the OLS estimate of $b(k)$, $\hat{e}_{tk} = e_t + o_p(1)$.

Recall the notation $\tilde{y}_t = y_t - \sum_{j=1}^k \hat{b}_j y_{t-j}$ and $\tilde{x}_t = x_t - \sum_{j=1}^k \hat{b}_j x_{t-j}$. Define $X_{tk} = (x_t, x_{t-1}, \dots, x_{t-k})$, $U_{tk} = (u_t, u_{t-1}, \dots, u_{t-k})$ and $\hat{b} = (1, -\hat{b}_1, \dots, -\hat{b}_k)'$ such that $\tilde{x}_t = \hat{b}'X_{tk}$ and $\tilde{u}_t = \hat{b}'U_{tk}$. Then we obtain

$$\begin{aligned}\hat{\beta}_{CO-AR} - \beta &= \left[\sum_{t=k+1}^T \tilde{x}_t^2 \right]^{-1} \left[\sum_{t=k+1}^T \tilde{x}_t \tilde{u}_t \right] \\ &= [\hat{b}'\hat{\Gamma}_k\hat{b}]^{-1} [\hat{b}'\hat{\Psi}_k\hat{b}]\end{aligned}$$

where $\hat{\Gamma}_k = \frac{1}{T} \sum_{t=k+1}^T X_{tk}X_{tk}'$ and $\hat{\Psi}_k = \frac{1}{T} \sum_{t=k+1}^T X_{tk}U_{tk}'$. Since $\hat{\Gamma}_0 > 0$ almost surely, it follows by Proposition 5.1.1 of Brockwell and Davis (1991) that, for $k \geq 1$, $\lambda_{\min}(\hat{\Gamma}_k) > 0$ almost surely. Furthermore, by Proposition 7.3.4 of Brockwell and Davis (1991), $\hat{\Gamma}_k - \Gamma_k = O_p(T^{-1/2})$, where $\Gamma_k = E[X_{tk}X_{tk}']$. Due to the independence of X_{tk} and U_{tk} it also follows by standard arguments that $\hat{\Psi}_k = O_p(T^{-1/2})$. This implies that $\hat{\beta}_{CO-AR} - \beta = O_p(T^{-1/2})$ as stated. $\square$

A.2 Proof of Theorem 2

Under $H_0: \beta = 0$, we have $\hat{\beta}_{CO-AR} = \left(T^{-1} \sum_{t=k+1}^T \tilde{x}_t^2 \right)^{-1} T^{-1} \sum_{t=k+1}^T \tilde{x}_t \tilde{u}_t$. First, using a result of Theorem 1, $T^{-1} \sum_{t=k+1}^T \tilde{x}_t^2 \rightarrow_p b'\Gamma_k b$. Since $\tilde{u}_t = \hat{e}_{tk} + o_p(1) = e_t + o_p(1)$, the terms $\tilde{x}_t \tilde{u}_t$ have finite variance and are asymptotically uncorrelated, such that a central limit theorem applies to the term $T^{-1/2} \sum_{t=k+1}^T \tilde{x}_t \tilde{u}_t \rightarrow_{\mathcal{L}} N(0, b'\Gamma_k b \sigma_u^2)$. Hence, under H_0 , $\sqrt{T} \hat{\beta}_{CO-AR} \rightarrow_{\mathcal{L}} N(0, \sigma_\beta^2)$, where $\sigma_\beta^2 = (b'\Gamma_k b)^{-1} \sigma_u^2$. Next, note that $\hat{\sigma}_u^2$ is a consistent estimator of σ_u^2 due to the assumption of finite fourth moments of e_t . This implies that $T\hat{\sigma}_\beta^2$ is a consistent estimator of σ_β^2 , which then implies an asymptotically standard normally distributed t -statistic. $\square$

A.3 Proof of Theorem 3

First, note that $\hat{u}_t = u_t - (\hat{\beta}_{GLSC} - \beta)x_t$, and that the second term does not converge to zero, unlike in the case where x_t is stationary. However, the AR(k) approximation of $\hat{u}_t$ can be equivalently represented as

$$\hat{u}_t - \hat{u}_{t-1} = \hat{\psi}_1 \hat{u}_{t-1} + \sum_{j=2}^k \hat{\psi}_j \Delta \hat{u}_{t-j+1} + \hat{e}_{tk}$$

where $\hat{\psi}_1 = \sum_{i=1}^k \hat{b}_i - 1$, $\hat{\psi}_j = -\sum_{i=j}^k \hat{b}_i$, $j = 2, \dots, k$. The leading term of $\hat{u}_t$ is integrated, while $\Delta \hat{u}_t = \Delta u_t - (\hat{\beta}_{GLSC} - \beta) \Delta x_t$ is stationary. Under our assumptions, $\hat{\psi}_1 = O_p(T^{-1})$ and $\hat{\psi}_j = \psi_j + O_p(T^{-1/2})$, $j = 2, \dots, k$. In a similar way, the filtered series $\tilde{x}_t$ and $\tilde{y}_t$ can be represented as

$$\begin{aligned}\tilde{x}_t &= x_t - (1 + \hat{\psi}_1)x_{t-1} - \sum_{j=2}^k \hat{\psi}_j \Delta x_{t-j} \\ \tilde{y}_t &= y_t - (1 + \hat{\psi}_1)y_{t-1} - \sum_{j=2}^k \hat{\psi}_j \Delta y_{t-j}\end{aligned}$$

Define the processes

$$x_t^* = \Delta x_t - \sum_{j=2}^k \hat{\psi}_j \Delta x_{t-j} \quad (2)$$

$$y_t^* = \Delta y_t - \sum_{j=2}^k \hat{\psi}_j \Delta y_{t-j} \quad (3)$$

which are stationary by construction. Since $\hat{\psi}_1 = O_p(T^{-1})$, we have $\tilde{x}_t = x_t^* - \hat{\psi}_1 x_{t-1} = x_t^* + O_p(T^{-1/2})$, and thus, $\{\tilde{x}_t\}$ converges in probability to a stationary process. The same holds for $\tilde{y}_t = y_t^* + O_p(T^{-1/2})$. This implies that the CO-AR estimator, i.e. the OLS estimator of β in the regression $\tilde{y}_t = \beta \tilde{x}_t + \eta_t$ is $\sqrt{T}$ -consistent as stated. $\square$

A.4 Proof of Theorem 4

We can write the regression $\tilde{y}_t = \beta \tilde{x}_t + \eta_t$ as $y_t^* = \beta x_t^* + \eta_t + O_p(T^{-1/2})$, where x_t^* and y_t^* are stationary processes given by (2) and (3). Then, the same arguments of the proof of Theorem 2 apply to the CO-AR estimator of β , and we obtain the required result. $\square$

A.5 Proof of Theorem 5

The first stage regression yields an estimator $\hat{\beta}$ with $\sqrt{T}(\hat{\beta} - \beta) \rightarrow_d N(0, Z)$. At the second stage, an AR(k) model is fitted to the residuals $\hat{u}_t$. By assumption, the stationary sequence $\{u_t\}$ has the representation $u_t = \sum_{j=1}^{\infty} b_j u_{t-j} + e_t$ with exponentially decaying coefficients, i.e. there are constants $C, \kappa > 0$ such that $b_j \leq C \exp(-\kappa j)$ for all $j \geq 1$. This implies that $\sum_{j=1}^{\infty} |j|^s |b_j| < \infty$, $s > 1$, so that we can use the results of Lemma 2.1 of Phillips and Ouliaris (1990).

Lemma 3 of Berk (1974) implies that $\hat{b}(k) \rightarrow_p b(k)$, where $\hat{b}(k) = (\hat{b}_1, \dots, \hat{b}_k)'$ is the coefficient vector of the approximating model. The CO-AR estimator can be written as

$$T(\hat{\beta} - \beta) = \left(\frac{1}{T^2} \sum_{t=1}^T \tilde{x}_t^2 \right)^{-1} \frac{1}{T} \sum_{t=1}^T \tilde{x}_t \tilde{u}_t.$$

First, by Lemma 2.1 of Phillips and Ouliaris (1990), we have

$$\frac{1}{\sqrt{T}} \tilde{x}_{[Tr]} \rightarrow_d B(1)V(r)$$

where $r \in (0, 1)$, $B(1) = 1 - \sum_{j=1}^{\infty} b_j$, $V(r)$ is a Brownian motion with variance $\sum_{j=-\infty}^{\infty} E[v_t v_{t-j}]$ and $v_t = \Delta x_t$. Hence,

$$\frac{1}{T^2} \sum_{t=1}^T \tilde{x}_t^2 \rightarrow_d \int_0^1 B(1)^2 V(r)^2 dr.$$

Second, we have

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{[Tr]} \tilde{u}_t \rightarrow_d B(1)W(r)$$

where $W(r)$ is a Brownian motion with variance $\sum_{j=-\infty}^{\infty} E[u_t u_{t-j}]$. Hence,

$$\frac{1}{T} \sum_{t=1}^T \tilde{x}_t \tilde{u}_t \rightarrow_d B(1)^2 \int_0^1 V(r) dW(r).$$

Therefore, we obtain $\hat{\beta} - \beta = O_p(T^{-1})$ as stated, and the asymptotic distribution of the CO-AR estimator is given by

$$T(\hat{\beta} - \beta) \rightarrow_d \left(\int_0^1 V(r)^2 dr \right)^{-1} \int_0^1 V(r) dW(r)$$

which is identical to the asymptotic distribution of the OLS estimator. $\square$

References

- Bauer, D., and M. Wagner. 2008. "Autoregressive Approximations of Multiple Frequency I(1) Processes." Working paper, Economics Series, Institute for Advanced Studies, Vienna.
- Berk, K. N. 1974. "Consistent Autoregressive Spectral Estimates." *The Annals of Statistics* 2: 489–502.
- Brockwell, P. J., and R. A. Davis. 1991. *Time Series: Theory and Methods*. 2nd ed. New York, NY: Springer-Verlag.
- Choi, C.-Y., L. Hu, and M. Ogaki. 2008. "Robust Estimation for Structural Spurious Regressions and a Hausman-Type Cointegration Test." *Journal of Econometrics* 142: 327–351.
- Deng, A. 2014. "Understanding Spurious Regression in Financial Economics." *Journal of Financial Econometrics* 12: 122–150.
- Engle, R. F., and C. W. J. Granger. 1987. "Cointegration and Error Correction: Representation, Estimation, Testing." *Econometrica* 55 (2): 251–276.
- Ferson, W. E., S. Sarkissian, and T. Simin. 2003. "Spurious Regressions in Financial Economics?" *Journal of Finance* 58 (4): 1393–1414.
- Granger, C. W. J. 2001. "Spurious Regressions." In *A Companion to Theoretical Econometrics*, edited by B. Baltagi. Oxford: Blackwell Publishers.
- Granger, C. W. J., and P. Newbold. 1974. "Spurious Regressions in Econometrics." *Journal of Econometrics* 74: 111–120.
- Granger, C. W. J., N. Hyung, and H. Jeon. 2001. "Spurious Regressions with Stationary Series." *Applied Economics* 33: 899–904.
- Haldrup, N. 1994. "The Asymptotics of Single-Equation Cointegration Regressions with I(1) and I(2) Variables." *Journal of Econometrics* 63: 153–81.
- Hendry, D. F. 1980. "Econometrics: Alchemy or Science?" *Economica* 47: 387–406.
- Ing, C.-K., and C.-Z. Wei. 2003. "On Same-Realization Prediction in an Infinite-Order Autoregressive Process." *Journal of Multivariate Analysis* 85: 130–155.
- Phillips, P. C. B. 1986. "Understanding Spurious Regressions in Econometrics." *Journal of Econometrics* 33: 311–340.
- Phillips, P. C. B., and S. Ouliaris. 1990. "Asymptotic Properties of Residual Based Tests for Cointegration." *Econometrica* 58: 165–193.
- Plosser, C. I., and G. W. Schwert. 1978. "Money Income and Sunspots: Measuring Economic Relationships and the Effects of Differencing." *Journal of Monetary Economics* 4: 637–660.
- Plosser, C. I., G. W. Schwert, and H. White. 1982. "Differencing as a Test of Specification." *International Economic Review* 23: 535–552.
- Saikkonen, P. 1991. "Asymptotically Efficient Estimation of Cointegration Regressions." *Econometric Theory* 7: 1–21.
- Shibata, R. 1980. "Asymptotically Efficient Selection of the Order of the Model for Estimating Parameters of Linear Processes." *Annals of Statistics* 8: 147–164.
- Stock, J. H., and M. W. Watson. 1993. "A Simple Estimator of Cointegrating Vectors in Higher Order Integrated Systems." *Econometrica* 61: 783–820.
- Valkanov, R. 2003. "Long-Horizon Regressions: Theoretical Results and Applications." *Journal of Financial Econometrics* 68: 201–232.
- Yule, G. U. 1926. "Why Do We Sometimes Get Nonsense Correlations Between Time Series? A Study in Sampling and the Nature of Time Series." *Journal of the Royal Statistical Society* 89: 1–64.

Supplemental Material: The online version of this article offers supplementary material (DOI: <https://doi.org/10.1515/snde-2015-0040>).