Economics Bulletin

Volume 31, Issue 1

The "spurious regression problem" in the classical regression model framework

Gueorgui I. Kolev
EDHEC Business School

Abstract

I analyse the "spurious regression problem" from the Classical Regression Model (CRM) point of view. Simulations show that the autocorrelation corrections suggested by the CRM, e.g., feasible generalised least squares, solve the problem. Estimators are unbiased, consistent, efficient and deliver correctly sized tests. Conversely, first differencing the data results in inefficiencies when the autoregressive parameter in the error process is less than one. I offer practical recommendations for handling cases suspected to be in the "spurious regression" class.

Address for correspondence: 24 avenue Gustave Delory, CS 50411, F-59057 ROUBAIX CEDEX 1, France. Email: gueorgui.kolev@edhec.edu. I would like to thank John P. Conley (the Editor) and an anonymous referee for comments and suggestions. Remaining errors are mine. This research was supported by the grant SEJ2006-13537 from the Spanish Ministry of Science and Technology.
Citation: Gueorgui I. Kolev, (2011) "The "spurious regression problem" in the classical regression model framework", *Economics Bulletin*, Vol. 31 no.1 pp. 925-937.
Submitted: Mar 18 2010. **Published:** March 21, 2011.

1. Introduction

In their seminal paper Granger and Newbold (1974) showed by simulation the consequences of applying Ordinary Least Squares (OLS) to two independent random walks. Regressing one variable y_t on another x_t, where y_t and x_t follow mutually independent random walks, results in a) inconsistent estimate of the slope and b) even asymptotically incorrect inference. (As the usual t-ratio is divergent, the true null hypothesis will be rejected more often than the stated nominal size of the t-test, and the over-rejection problem gets worse as the sample size increases, see Phillips, 1986.) The two random walks would seem to be related, as judged by standard criteria as the t-statistic and R^2, where in fact they are not. Granger and Newbold (1974) also coined the term "spurious regression" (borrowing from Pearson, 1987 and Yule, 1987, 1986) and spurred a whole generation of research in econometrics of unit roots and asymptotic theory specially designed to handle non-stationarity.

Here I take a simple and somewhat unconventional look at the "spurious regression" problem – I analyse the issue from the (Generalised) Classical Regression Model (CRM) point of view. To anticipate the discussion

1. the "spurious regression" problem is simply a problem of severe autocorrelation in the error process
2. severe autocorrelation is trivial to spot in any empirical application
3. there are straightforward remedies suggested by the CRM, involving transformations of the structural equation that deliver an estimating equation that satisfies all the requirements of the CRM
4. I show by simulation that when these remedies are implemented, the "spurious regression" problem does not arise – for the true generalised least squares estimator the t-statistics follow *exactly* the Student's t-distribution in any sample size, and the actually observed rejection rate of the correct null hypothesis is equal to the nominal stated size of the test
5. for the feasible generalised least squares versions the performance is as good as the true generalised least squares estimator for moderately large sample sizes (say 70 observations)
6. conversely, even when "spurious regression" is *not* present in the data generating process, however autocorrelation *is* present, OLS inference fails quite spectacularly with observed rejection rates many times larger than the nominal stated size of the test.

"Spurious regression" is neither necessary, nor sufficient condition for severe autocorrelation related pathologies in OLS inference to occur. It is not clear how helpful it is to think of the issue as a "spurious regression" problem and how much this mode of thinking helps us in finding a satisfactory remedy.

What is notable about the CRM suggested remedies to the problem, is that no reference is made to asymptotic theory and to the time series properties of the variables involved (i.e., the remedies work perfectly fine for the unit root autoregressive case). When the data generating processes for both y_t and x_t are assumed to be known to have unit roots, the Generalised Least Squares (GLS) solution amounts to what asymptotic theory suggests – apply first difference transformation to the data. More interestingly, even when the true data generating process is not assumed to be known, the feasible estimated versions of the transformations work fine for moderately large sample sizes, in the sense of delivering standard correct inference, which is as efficient as the true GLS.

The "spurious regression" problem has cropped up in various contexts and for very diverse data generating processes (for a recent survey see Ventosa-Santaularia, 2009). I am not able to cover all these data generating processes here. I do extend slightly the original Granger and Newbold (1974) set-up with which most practitioners and students of econometrics are well familiar to cover cases where we regress an I(0) or I(1) dependent variable on a I(1) or I(0) or I(2) regressor.

This paper is closely related to a contemporary work by McCallum (2010). He shows by simulation that no "spurious regression" problem occurs when (the suggested by the heavy autocorrelation in the OLS residuals) Cochrane-Orcutt estimator is applied. Sun (2004) derived a convergent t-statistic for "spurious regression" set-up. His t-ratio divides by a Heteroskedasticity and Autocorrelation Consistent (HAC) estimate of the variance, where to take into account the heavy and not dying out autocorrelation one does not truncate the series, contrary to what is done in the conventional HAC variance estimate.

The common theme here is that "spurious regression" relationships are accompanied by easy to spot severe autocorrelation in the residuals of the estimated by OLS equation. Estimation methods taking account of the potential autocorrelation, be it through feasible GLS, another transformation that eliminates the autocorrelation, or by fixing the covariance matrix as in Sun (2004), eliminates the spuriously detected relationships.

This paper differs in placing emphasis on the fact that all this is to be expected once one looks at the "spurious regression problem" from the Classical Regression Model point of view.

2. The Classical Regression Model and how "spurious regression" fits in it

Consider the following model

$$y_t = \beta_0 + \beta_1 x_t + u_t \quad \text{defined over} \quad t = 1,2,..,T \tag{1}$$

$$u_t = \rho u_{t-1} + e_t, \quad u_0 = e_0 \tag{2}$$

$$x_t = \theta x_{t-1} + w_t, \quad x_0 = w_0 \tag{3}$$

$$\text{and each } w_t \text{ is iid} \sim (0, \sigma_w^2) \text{ and is independent of each } e_t \sim \text{iid}\,\text{Normal}(0, \sigma_e^2). \tag{4}$$

Setting $\beta_0 = \beta_1 = 0$, $\rho = 1$, and $\theta = 1$ we obtain as a special case the Granger and Newbold (1974) classical example of "spurious regression" of two independent random walks.

Observe that eq. (4) imply that each u_t is independent of each x_t, which in turn implies the weaker condition known as *strict exogeneity* in the CRM context

$$\text{E}(u_t | x_1, x_2, .., x_T) = 0 \quad \text{for each} \quad t. \tag{5}$$

Define $y \equiv [y_T\, y_{T-1}\, ..\, y_1]'$, $X \equiv [[1\,1, .., 1]'[x_T\, x_{T-1}\, ..x_1]']$, $\beta \equiv [\beta_0\, \beta_1]'$, and $u \equiv [u_T\, u_{T-1}\, ..\, u_1]'$. Then eq. (1) can be written as $y = X\beta + u$ and the OLS estimator is

$$\hat{\beta} = (X'X)^{-1}X'y \tag{6}$$

and eq. (5) can be written as $\text{E}(u|X) = 0$. Observe that eq. (1) and eq. (5) together imply that

$$\text{E}(\hat{\beta}|X) = \text{E}[(X'X)^{-1}X'y|X] = \text{E}[(X'X)^{-1}X'(X\beta + u)|X] = \beta + \text{E}(u|X) = \beta$$

Result 1: $\text{E}(\hat{\beta}|X) = \beta$ and by the law of iterated expectations $\text{E}\hat{\beta} = \beta$ holds.

One often hears both econometricians and practitioners alike reciting a version of the statement "When you regress one random walk on another you obtain a spurious regression, a regression that does not make *any sense*." As Result 1 shows, applying OLS results in estimates that make perfect sense in one respect – they are *unbiased* for the true parameter value.

Conditional on X the error term in eq. (1) follows an autoregression and so it violates the assumption of spherical disturbance in the CRM

$$\text{E}(uu'|X) \neq \sigma_u I.$$

In particular when $\rho = 1$ the autocorrelation in u_t does not die out. The solution when ρ is known is to transform the model so that the new error term satisfies the CRM assumptions.

To that end define $\tilde{y}_t = y - \rho y_{t-1}$, $\tilde{x}_t = x_t - \rho x_{t-1}$ and $e_t = u_t - \rho u_{t-1}$. Therefore we observe that the quasidifferenced model

$$\tilde{y} = \tilde{X}\beta + e \tag{7}$$

satisfies all the CRM assumptions as

$$\mathrm{E}(ee'|\tilde{X}) = \sigma_e I \text{ and } e|\tilde{X} \sim \mathrm{Normal}(0, \sigma_e I).^2 \tag{8}$$

Therefore applying OLS on eq. (7) results in the best unbiased estimator known as the generalised least squares estimator

$$\tilde{\beta} = (\tilde{X}'\tilde{X})^{-1}\tilde{X}'\tilde{y}. \tag{9}$$

Result 2: When ρ is known the generalised least squares estimator $\tilde{\beta}$ in eq. (9) is efficient and unbiased for the population parameter β. Moreover the associated with the slope parameter t-ratio is *exactly* Student's t-distributed in *any* sample size.[3]

Result 3: When it is known that $\rho = 1$ and $\theta = 1$, the CRM recommendation is the same as the one coming from asymptotic stationarity considerations – take first differences of the data and estimate the model in differences.

In practical applications the exact value of ρ will rarely be known, so we have two major ways of implementing the above idea. Following Cochrane and Orcutt (1949) we can apply the feasible generalised least squares procedure. In the first step we estimate by OLS the original (untrasformed) model eq. (1) and from an auxiliary regression of the residual on the lagged residual we compute $\hat{\rho}$, the OLS estimator of ρ. In the second step we apply OLS on the quasi-differenced data $\tilde{y} = y_t - \hat{\rho} y_{t-1}$, $\tilde{x} = x_t - \hat{\rho} x_{t-1}$. Alternatively we can estimate both ρ and β in one step in the following regression

$$y_t = \beta_0(1-\rho) + \beta_1 x_t + \rho y_{t-1} + \beta_1(-\rho)x_{t-1} + e_t \tag{10}$$

where this equation results from shifting all variables multiplied by unknown parameters in eq. (7) to the right hand side (Durbin, 1970).

In the simulations that follow I study the performance of five estimators

1. the OLS estimator on data in levels
2. the generalised least squares estimator, i.e., OLS on quasidifferenced data using the correct population ρ
3. the OLS estimator on first differenced data
4. the Cochrane and Orcutt (1949) feasible generalised least squares
5. the one step estimator of β in eq. (10) that results from including one lag of the dependent variable and the regressor.

In the process I vary the autoregressive parameter of the u_t series ($\rho = 1, 0.8, 0.6$), and the autoregressive parameter of the x_t series ($\theta = 1, 0.8$). The case where $\rho = 1$ and $\theta = 1$ corresponds to the original Granger and Newbold (1974) "spurious regression" case.

3. Simulation

I generate 30,000 replications from the model

$$y_t = \beta_0 + \beta_1 x_t + u_t, \qquad [\beta_0\, \beta_1] = [0\, 0]$$

$$u_t = \rho u_{t-1} + e_t, \quad u_0 = e_0, \qquad \rho = [1, 0.8, 0.6]$$

[2]Under strict exogeneity, conditioning on $\tilde{X}$ is the same as conditioning on X – as the former is a linear transformation of the latter, the two information sets contain the same information.

[3]Proofs of these facts can be found in any textbook on econometrics, e.g., Hayashi (2000, Chapter 1).

$$x_t = \theta x_{t-1} + w_t, \quad x_0 = w_0, \qquad \theta = [1, 0.8]$$

and each $w_t \sim$ iid Normal$(0,1)$ is independent of each $e_t \sim$ iid Normal$(0,1)$

for sample sizes of $T = [15, 30, 70, 100, 200, 500, 1000, 9000]$.

For each replication round I generate $1000 + T$ observations and discard the first 1000, using only the last T observations for estimation.

3.1 Means for the estimators and t-statistic rejection rates for the true null hypothesis

Each entry in Table I and Table II that follow contains the *mean* of a quantity computed over the 30,000 replication rounds. The mean of the estimate is followed by the mean of an indicator variable taking the value of 1 if the t-statistic associated with the given estimate and with the (true) null hypothesis Ho: $\beta_1 = 0$ exceeds in absolute value the appropriate critical value for Student's t random variable. In other words we have a column of the means of the estimates followed by the rejection rate under the null hypothesis. The test of the null hypothesis is always carried out at the 5% significance level. Therefore,

1. $\hat{\beta}_{\text{OLS}}$ is the OLS estimator on levels data, i.e., the slope estimated from a regression of the form $y_t = a + bx_t +$ error, and the indicator $[|t_{\hat{\beta}_{\text{OLS}}}| > t_c]$ takes the value of 1 if the absolute value of the t-statistic associated with the null hypothesis that b is 0 exceeds the appropriate critical value from the Student's t distribution, and takes the value of 0 otherwise[4]
2. $\tilde{\beta}_{\text{GLS}}$ is the generalised least squares estimator (i.e., OLS on quasidifferenced data using the correct population ρ, i.e., the slope estimated from a regression of the form $(y_t - \rho y_{t-1}) = a + b(x_t - \rho x_{t-1}) +$ error, and the indicator $[|t_{\tilde{\beta}_{\text{GLS}}}| > t_c]$ takes the value of 1 if the absolute value of the t-statistic associated with the null hypothesis that b is 0 exceeds the appropriate critical value from the Student's t distribution
3. $\hat{\beta}_{\text{DOLS}}$ is the OLS estimator on first differenced data, i.e., the slope estimated from a regression of the form $(y_t - y_{t-1}) = a + b(x_t - x_{t-1}) +$ error and $[|t_{\hat{\beta}_{\text{DOLS}}}| > t_c]$ is the associated indicator for the null hypothesis that $b = 0$ rejection
4. $\hat{\beta}_{\text{CO}}$ is the Cochrane and Orcutt (1949) feasible generalised least squares, i.e., the slope estimated from a regression of the form $(y_t - \hat{\rho} y_{t-1}) = a + b(x_t - \hat{\rho} x_{t-1}) +$ error and $[|t_{\hat{\beta}_{\text{CO}}}| > t_c]$ is the associated indicator for the null hypothesis that $b = 0$ rejection
5. $\hat{\beta}_{\text{OS}}$ is the one step estimator of β_1 in eq. (10) that results from including one lag of the dependent variable and the regressor, i.e., the first slope parameter estimated from a regression of the form $y_t = a + bx_t + ry_{t-1} + cx_{t-1} +$ error and $[|t_{\hat{\beta}_{\text{OS}}}| > t_c]$ is the associated $b = 0$ null hypothesis rejection indicator.

In Table I and Table II below all five estimators, including the OLS estimator are *unbiased*. Hence statements of the sort that OLS in levels in the spurious regression model delivers results that "do not make any sense" are unnecessary and incorrect over-generalisations.

3.1.1 Means for the estimators and t-statistic rejection rates for the true null hypothesis when x_t is I(1)

Looking through the entries of the Table I, we observe the following.

1. The OLS estimator on levels results in extremely misleading inference whether or not we have "spurious regression."[5]

[4] The Iverson bracket [•], named after Kenneth E. Iverson, is used to denote a number that takes the value of 1 if the condition in square brackets is true, and takes the value of 0 otherwise.

[5] The tests of the true null hypothesis from OLS on levels are grossly oversized even when the true $\rho = 0.6$ – very far

Table I: The parameter ρ in eq. (2) is set to 1, 0.8 and 0.6, and the parameter θ in eq. (3) is set to 1. To conserve space, results for selected *sample sizes* of $T = [70, 200, 500, 1000]$ observations are reported in the table.

T	$\hat{\beta}_{\text{OLS}}$	$[\lvert t_{\hat{\beta}_{\text{OLS}}}\rvert > t_c]$	$\tilde{\beta}_{\text{GLS}}$	$[\lvert t_{\tilde{\beta}_{\text{GLS}}}\rvert > t_c]$	$\hat{\beta}_{\text{DOLS}}$	$[\lvert t_{\hat{\beta}_{\text{DOLS}}}\rvert > t_c]$	$\hat{\beta}_{\text{CO}}$	$[\lvert t_{\hat{\beta}_{\text{CO}}}\rvert > t_c]$	$\hat{\beta}_{\text{OS}}$	$[\lvert t_{\hat{\beta}_{\text{OS}}}\rvert > t_c]$
	Means of the statistics for $\rho = 1$									
70	-0.0004	0.7199	-0.0018	0.0511	-0.0018	0.0511	-0.0019	0.0928	-0.0019	0.0560
200	-0.0021	0.8291	0.0005	0.0507	0.0005	0.0507	0.0003	0.0640	0.0005	0.0505
500	-0.0020	0.8936	0.0001	0.0501	0.0001	0.0501	0.0000	0.0555	0.0001	0.0509
1000	-0.0007	0.9230	0.0004	0.0518	0.0004	0.0518	0.0004	0.0557	0.0004	0.0520
	Means of the statistics for $\rho = 0.8$									
70	0.0009	0.4615	0.0002	0.0496	-0.0001	0.0482	0.0001	0.0953	-0.0001	0.0505
200	0.0003	0.4959	0.0002	0.0508	-0.0002	0.0490	0.0002	0.0726	-0.0000	0.0493
500	-0.0001	0.5108	-0.0000	0.0499	0.0001	0.0481	0.0000	0.0629	0.0001	0.0496
1000	0.0001	0.5049	0.0000	0.0476	0.0000	0.0527	0.0000	0.0551	0.0001	0.0527
	Means of the statistics for $\rho = 0.6$									
70	-0.0006	0.2979	-0.0003	0.0491	0.0004	0.0512	-0.0004	0.0836	0.0001	0.0527
200	-0.0002	0.3162	-0.0002	0.0492	0.0005	0.0487	-0.0001	0.0627	0.0004	0.0515
500	0.0000	0.3227	0.0000	0.0480	-0.0003	0.0508	0.0000	0.0544	-0.0002	0.0511
1000	-0.0001	0.3294	-0.0001	0.0517	-0.0001	0.0475	-0.0001	0.0542	-0.0001	0.0486

2. The true generalised least squares estimator delivers correctly sized t-test for the true null hypothesis – the observed rejection rate under the null hypothesis is equal to the nominal stated size of the test, and this is so for *any* sample size.
3. When we incorrectly impose $\rho = 1$ by first differencing the data, the estimator on first differences is still unbiased and delivers correctly sized tests even when $\rho = 0.8$ or $\rho = 0.6$.
4. The Cochrane-Orcutt feasible GLS and the one step estimator in eq. (10) perform reasonably well for moderately large sample sizes and any values of ρ.
5. The Cochrane-Orcutt estimator does exhibit size distortions for small sample sizes of 70 observations or less. Note however that the size distortion is present for any ρ, whether we have "'spurious regression" or not.

3.1.2 *Means for the estimators and t-statistic rejection rates for the true null hypothesis when x_t is I(0)*

Looking through the entries of the Table II, we draw conclusions very similar to the ones coming from our inspection of Table I. OLS is misleading, the true GLS delivers correctly sized t-statistics for any sample size, and incorrectly first differencing the variables does not lead to any harm in terms of test size or bias. The one step estimator in eq. (10) exhibits negligible size distortion, and only so for very small sample sizes. (E.g., not reported in the table, but the rejection rate is 6.56% at nominal stated size of the test of 5% when the sample size is 15 observations.) The Cochrane-Orcutt feasible GLS exhibits non-negligible size distortions for very small sample sizes of 15 or 30 observations, but achieves correct size somewhat faster than in the simulation design of Table I, and the size distortion for very small sample sizes of 15 or 30 is somewhat smaller in the simulation design of Table II (not

away from a unit root in the y_t series! Hence it is doubtful how useful is to think of spurious regression as a non-stationarity phenomenon. Seems more useful to think of it as something that arises whenever heavy autocorrelation in the residual is present.

Table II: The parameter ρ in eq. (2) is set to 1, 0.8 and 0.6, and the parameter θ in eq. (3) is set to 0.8. To conserve space, results for selected *sample sizes* of T = [70, 200, 500, 1000] observations are reported in the table.

T	$\hat{\beta}_{\text{OLS}}$	$[\lvert t_{\hat{\beta}_{\text{OLS}}}\rvert > t_c]$	$\tilde{\beta}_{\text{GLS}}$	$[\lvert t_{\tilde{\beta}_{\text{GLS}}}\rvert > t_c]$	$\hat{\beta}_{\text{DOLS}}$	$[\lvert t_{\hat{\beta}_{\text{DOLS}}}\rvert > t_c]$	$\hat{\beta}_{\text{CO}}$	$[\lvert t_{\hat{\beta}_{\text{CO}}}\rvert > t_c]$	$\hat{\beta}_{\text{OS}}$	$[\lvert t_{\hat{\beta}_{\text{OS}}}\rvert > t_c]$
	Means of the statistics for $\rho = 1$									
70	-0.0009	0.4662	0.0003	0.0499	0.0003	0.0499	0.0002	0.0622	-0.0002	0.0520
200	0.0087	0.4940	0.0003	0.0513	0.0003	0.0513	0.0003	0.0492	0.0005	0.0510
500	-0.0041	0.5074	-0.0006	0.0515	-0.0006	0.0515	-0.0006	0.0493	-0.0006	0.0492
1000	-0.0015	0.5070	-0.0003	0.0510	-0.0003	0.0510	-0.0003	0.0506	-0.0004	0.0496
	Means of the statistics for $\rho = 0.8$									
70	-0.0018	0.3371	-0.0006	0.0499	-0.0004	0.0565	-0.0003	0.0689	-0.0003	0.0522
200	0.0007	0.3489	-0.0000	0.0494	-0.0001	0.0542	0.0000	0.0524	0.0000	0.0491
500	-0.0002	0.3535	-0.0006	0.0484	-0.0007	0.0544	-0.0006	0.0494	-0.0006	0.0481
1000	-0.0000	0.3538	0.0003	0.0491	0.0003	0.0552	0.0002	0.0501	0.0002	0.0486
	Means of the statistics for $\rho = 0.6$									
70	0.0000	0.2314	0.0003	0.0507	0.0005	0.0590	0.0002	0.0709	0.0005	0.0526
200	-0.0007	0.2486	-0.0005	0.0497	-0.0003	0.0576	-0.0004	0.0562	-0.0002	0.0474
500	-0.0002	0.2391	0.0002	0.0499	0.0005	0.0611	0.0002	0.0518	0.0003	0.0518
1000	0.0000	0.2442	0.0000	0.0493	0.0001	0.0598	0.0001	0.0509	0.0001	0.0495

reported in the tables but found in the full set of simulations). The Cochrane-Orcutt feasible GLS achieves correct size at about 70 to 100 observations.

3.2 Efficiency of various estimators

Each odd column in Table III and Table IV below contains the standard deviations of the estimators computed over the 30,000 replication runs, i.e., the true standard errors of the estimators. Each even column contains the skewness of the sampling distribution of the estimators, and they are all roughly symmetric.

3.2.1 Efficiency of various estimators when x_t is I(1)

The following remarks can be made regarding efficiency of the five estimators.

1. As it is well known, OLS on levels $\hat{\beta}_{\text{OLS}}$ in the "spurious regression" case with $\rho = 1$ is not consistent, and the standard deviation of its sampling distribution does not go to zero as the sample size increases. We observe this in the first column and first panel above. All the other four estimators are consistent.
2. More interestingly, even in the "spurious regression" case with $\rho = 1$ the two feasible estimators that do not assume that ρ is known and estimate it from the data ($\hat{\beta}_{\text{CO}}$ and $\hat{\beta}_{\text{OS}}$) are roughly as efficient as the true generalised least squares estimators for moderately large sample sizes of 70 or more observations. Thus, an econometrician in the year of 1950 who had never heard of "spurious regression" and unit root asymptotics, but had read Cochrane and Orcutt (1949) and was aware of the dangers of serial correlation in the error term, and had applied the suggested remedy, would have done just fine.
3. Looking in the second ($\rho = 0.8$) and the third ($\rho = 0.6$) panels, we see that always differencing the data is not such a good idea. The Cochrane-Orcutt feasible GLS has smaller standard errors compared the standard errors of the (mis-specified) OLS on first differenced data. For example when

Table III: The parameter ρ in eq. (2) is set to 1, 0.8 and 0.6, and the parameter θ in eq. (3) is set to 1. To conserve space, results for selected *sample sizes* of $T = [70, 200, 500, 1000]$ observation are reported in the table.

T	se $\hat{\beta}_{\text{OLS}}$	skew $\hat{\beta}_{\text{OLS}}$	se $\tilde{\beta}_{\text{GLS}}$	skew $\tilde{\beta}_{\text{GLS}}$	se $\hat{\beta}_{\text{DOLS}}$	skew $\hat{\beta}_{\text{DOLS}}$	se $\hat{\beta}_{\text{CO}}$	skew $\hat{\beta}_{\text{CO}}$	se $\hat{\beta}_{\text{OS}}$	skew $\hat{\beta}_{\text{OS}}$
	Standard Deviation and Skewness of the sampling distribution for $\rho = 1$									
70	0.6265	0.0065	0.1226	-0.0154	0.1226	-0.0154	0.1346	-0.0513	0.1267	-0.0184
200	0.6222	-0.0312	0.0714	-0.0107	0.0714	-0.0107	0.0740	-0.0196	0.0723	-0.0051
500	0.6268	0.0150	0.0449	-0.0081	0.0449	-0.0081	0.0456	-0.0117	0.0451	-0.0117
1000	0.6229	-0.0067	0.0318	-0.0256	0.0318	-0.0256	0.0321	-0.0294	0.0319	-0.0286
	Standard Deviation and Skewness of the sampling distribution for $\rho = 0.8$									
70	0.1783	-0.0190	0.1116	0.0259	0.1287	0.0260	0.1139	0.0083	0.1258	0.0151
200	0.0741	0.0163	0.0537	0.0049	0.0751	0.0259	0.0537	0.0072	0.0718	0.0217
500	0.0316	-0.0289	0.0262	-0.0036	0.0473	0.0262	0.0261	-0.0042	0.0451	0.0165
1000	0.0159	0.0333	0.0142	-0.0192	0.0337	0.0024	0.0142	-0.0195	0.0320	0.0055
	Standard Deviation and Skewness of the sampling distribution for $\rho = 0.6$									
70	0.1031	0.0120	0.0891	0.0285	0.1382	0.0236	0.0889	0.0176	0.1267	0.0165
200	0.0390	-0.0307	0.0359	-0.0179	0.0796	-0.0006	0.0358	-0.0150	0.0721	-0.0068
500	0.0160	0.0284	0.0153	0.0628	0.0502	-0.0053	0.0153	0.0580	0.0451	0.0054
1000	0.0082	-0.0314	0.0080	-0.0351	0.0352	-0.0167	0.0080	-0.0410	0.0315	-0.0138

$\rho = 0.6$ the standard error of the Cochrane-Orcutt estimator is less than half of the standard error of the mis-specified first differenced OLS, and this is so even for moderately large sample sizes.[6]

4. It should be pointed out however that the one step feasible estimator in eq. (10) has comparable standard errors to the mis-specified first differenced OLS.

3.2.2 *Efficiency of various estimators when x_t is I(0)*

The following remarks can be made regarding efficiency of the five estimators.

1. OLS on levels $\hat{\beta}_{\text{OLS}}$ when $\rho = 1$ (y_t is I(1)) is not consistent, even when x_t is I(0), but heavily autocorrelated ($\theta = 0.8$). The standard deviation of its sampling distribution does not go to zero as the sample size increases. We observe this in the first column and first panel above. All the other four estimators are consistent.
2. The two feasible estimators that do not assume that ρ is known and estimate it from the data ($\hat{\beta}_{\text{CO}}$ and $\hat{\beta}_{\text{OS}}$) are roughly as efficient as the true generalised least squares estimators for moderately large sample sizes of 70 or more observations.
3. The first differenced OLS estimator is always mis-specified here, if we think of first differencing as something that we should apply to I(1) variables to achieve stationarity. Even if $\rho = 1$, still $\theta = 0.8$ so first differencing both the regressor and the regressand does not match the true data generating process. (It does coincide with the true GLS when $\rho = 1$.) We see from the table that the first differenced OLS is almost as efficient as the true GLS even for $\rho = 0.8$ and for $\rho = 0.6$.

[6]Every econometrician and practitioner would in principle agree that "always differencing the data" is a bad idea. However this is what we all do effectively nowadays – we pretest the null hypothesis of a unit root, the tests in this class are known to have notoriously low power, and then when we cannot reject the unit root null hypothesis , we difference the data. I believe this questionable practice is to be blamed on the exuberant fear of running a "spurious regression" and obtaining "results that do not make any sense."

Table IV: The parameter ρ in eq. (2) is set to 1, 0.8 and 0.6, and the parameter θ in eq. (3) is set to 0.8. To conserve space, results for selected *sample sizes* of $T = [70, 200, 500, 1000]$ observations are reported in the table.

T	se $\hat{\beta}_{\text{OLS}}$	skew $\hat{\beta}_{\text{OLS}}$	se $\tilde{\beta}_{\text{GLS}}$	skew $\tilde{\beta}_{\text{GLS}}$	se $\hat{\beta}_{\text{DOLS}}$	skew $\hat{\beta}_{\text{DOLS}}$	se $\hat{\beta}_{\text{CO}}$	skew $\hat{\beta}_{\text{CO}}$	se $\hat{\beta}_{\text{OS}}$	skew $\hat{\beta}_{\text{OS}}$
	Standard Deviation and Skewness of the sampling distribution for $\rho = 1$									
70	0.6852	0.0002	0.1152	-0.0120	0.1152	-0.0120	0.1297	-0.0686	0.1235	-0.0062
200	0.7137	0.0284	0.0679	-0.0193	0.0679	-0.0193	0.0686	-0.0205	0.0718	-0.0187
500	0.7322	-0.0071	0.0426	0.0076	0.0426	0.0076	0.0427	0.0072	0.0448	-0.0035
1000	0.7256	-0.0451	0.0301	0.0047	0.0301	0.0047	0.0301	0.0060	0.0317	0.0068
	Standard Deviation and Skewness of the sampling distribution for $\rho = 0.8$									
70	0.2492	-0.0066	0.1221	-0.0040	0.1246	-0.0045	0.1291	-0.0083	0.1246	-0.0028
200	0.1489	-0.0097	0.0710	0.0150	0.0729	0.0096	0.0720	0.0139	0.0715	0.0134
500	0.0954	0.0099	0.0447	-0.0261	0.0459	-0.0255	0.0449	-0.0280	0.0448	-0.0265
1000	0.0673	0.0001	0.0316	0.0182	0.0325	0.0085	0.0316	0.0161	0.0316	0.0163
	Standard Deviation and Skewness of the sampling distribution for $\rho = 0.6$									
70	0.1587	-0.0100	0.1185	-0.0168	0.1337	-0.0070	0.1229	-0.0176	0.1246	-0.0078
200	0.0918	-0.0092	0.0680	-0.0195	0.0780	0.0017	0.0689	-0.0230	0.0714	-0.0064
500	0.0568	-0.0257	0.0429	-0.0184	0.0495	-0.0059	0.0430	-0.0186	0.0451	-0.0143
1000	0.0400	0.0096	0.0300	0.0119	0.0350	0.0050	0.0301	0.0127	0.0317	0.0069

4. Discussion and conclusion

Let us contrast now the implications of the CRM above with the current econometric practice. The current state of affairs is that practitioner are taught to be very wary of "spurious regression" results arising from regressions involving nonstationary variables. Therefore practitioners pretest the null hypothesis of a unit root in the original series y and X and if they are not able to reject the unit root null hypothesis they routinely difference the data and carry on by regression analysis in differences. The results above stand in stark contrast, we see that what we should be wary of is heavy autocorrelation in the residual. For example, the second and third panels of Table I are regressions of I(0) variable on I(1) variable and the t-statistic of the OLS estimator is divergent, it rejects the true null more and more often as the sample size increases. The second and third panels of Table II are regressions of I(0) variable on I(0) variable and the t-statistic of the OLS estimator is still divergent.

The implied empirical strategy suggested by our analysis is the following (I disregard the unlikely case that the researcher knows for sure that the data has a unit root):

1. Estimate the structural equation in levels, i.e., apply OLS to eq. (1). Determine whether there is autocorrelation in the residuals by a Durbin-Watson, Durbin, Breusch-Godfrey or any other appropriate[7] test (Durbin and Watson, 1950; Durbin, 1970; Wooldridge, 1991). Alternatively, just eyeball a plot of the autocorrelation function of the residual with confidence bands – the heavy and persistent autocorrelation arising in "spurious regression" cases is very obvious and hard to miss. (See Appendix A for a demonstration that one simply cannot miss the point that there is autocorrelation in the error process, not only in the $\rho = 1$ "spurious regression" case, but also when ρ is as low as 0.6.)
2. If autocorrelation in the OLS residual is detected, proceed directly with a method geared toward rendering the estimating equation complying with the CRM assumption, e.g., feasible GLS or eq. (10). In effect let the data determine the value of ρ that is needed.

[7] In the context of "spurious regression" the Durbin-Watson test is appropriate because strict exogeneity holds.

Alternatively, if one likes pretesting, one can reverse the current practice

1. Pretest the null hypothesis of stationarity in the original data series in levels, say with KPSS test (Kwiatkowski, Phillips, Schmidt and Shin, 1992).
2. If the test does *not* reject the null hypothesis of stationarity proceed with transforming the data in CRM form, or simply use HAC covariance estimate.
3. Only if one can reject the null hypothesis of stationarity, one should proceed with first differencing the data and estimating the structural equation in first differences.

As a matter of language, the simulation results here suggest that we should dispense altogether with the "spurious regression" terminology as it is not useful as all. For one thing, when one has substantial autocorrelation in the error process, but *no* "spurious regression" (e.g., first columns in Table I and Table II, $\rho = 0.6$ or $\rho = 0.8$) the OLS on levels delivers t-test with actual rejection rate of the true null hypothesis from 20% to 50% when the stated nominal level is 5%, which is unacceptable, and the problem gets worse as the sample size increases, i.e., the usual OLS t-statistic is divergent). For another, even when one *has* the prototypical Granger and Newbold "spurious regression" case ($\rho = 1$ in eq. (2) and $\theta = 1$ in eq. (3)), but takes care of the easy to detect autocorrelation in the residual through the Cochrane-Orcutt feasible GLS or any other technique that takes care of the serial correlation like eq. (10), it turns out that these two techniques deliver correctly sized tests, are unbiased, are consistent, and are roughly as efficient as the OLS on first differenced data (which in this case is the exact GLS best unbiased estimator).

Appendix A: Can one miss the autocorrelation in the error term?

The following Table A.1 and Table A.2 show the means of three statistics that are used to detect autocorrelation in the residuals, and the fraction of time that each of them *failed* to detect autocorrelation when it was in fact present. For "spurious regression" even for sample sizes as small as 30 observations, all three methods fail to detect it less than 5% of the time (not reported in the table below). If the sample size is 70 or larger, all three methods virtually *never* fail to reject the null hypothesis of no autocorrelation. The situation is similar when autocorrelation is somewhat lower 0.8 or 0.6 – autocorrelation at these level is impossible to miss.

Table A.1:
The parameter
ρ in eq. (2) is set to 1, 0.8 and 0.6, and the parameter θ in eq. (3) is set to 1. To conserve space, results for selected *sample sizes* of $T = [70, 200, 500, 1000]$ observations are reported in the table.

T	$\widehat{\mathrm{DW}}$	$[\widehat{\mathrm{DW}}]$	$\tilde{\rho}$	$[\lvert t_{\tilde{\rho}}\rvert < 1.96]$	$\hat{\rho}$	$[\lvert t_{\hat{\rho}}\rvert < 1.96]$
	Means of the statistics for $\rho = 1$					
70	0.2442	0.0000	0.8755	0.0000	0.9007	0.0000
200	0.0891	0.0000	0.9551	0.0000	0.9642	0.0000
500	0.0360	0.0000	0.9819	0.0000	0.9856	0.0000
1000	0.0181	0.0000	0.9909	0.0000	0.9928	0.0000
	Means of the statistics for $\rho = 0.8$					
70	0.5545	0.0000	0.7182	0.0000	0.7248	0.0001
200	0.4519	0.0000	0.7728	0.0000	0.7738	0.0000
500	0.4204	0.0000	0.7894	0.0000	0.7895	0.0000
1000	0.4099	0.0000	0.7948	0.0000	0.7949	0.0000
	Means of the statistics for $\rho = 0.6$					
70	0.9193	0.0030	0.5334	0.0059	0.5362	0.0058
200	0.8402	0.0000	0.5777	0.0000	0.5781	0.0000
500	0.8163	0.0000	0.5911	0.0000	0.5911	0.0000
1000	0.8080	0.0000	0.5956	0.0000	0.5956	0.0000

In the Table A.1 and Table A.2, $\widehat{\text{DW}}$ is the Durbin-Watson statistic, $\tilde{\rho}$ is the estimated ρ from regressing the OLS residual on the lagged OLS residual, and $\hat{\rho}$ is the directly estimated ρ in eq. (10). $[\widehat{\text{DW}}]$ is an indicator taking the value of 1 if the Durbin-Watson statistic exceeds 1.5 *and* the OLS t-statistic on the slope parameter rejects the Ho: $\beta_1 = 0$ (rough rule of thumb that approximates the case that in "spurious regression" we would think that the two variables are related and in the same time we would fail to detect the severe positive autocorrelation in the residual). $[|t_{\tilde{\rho}}| < 1.96]$ is an indicator taking the value of 1 if we regress OLS residuals on the lagged OLS residuals and fail to reject the null hypothesis that the autoregressive parameter is different from 0. $[|t_{\hat{\rho}}| < 1.96]$ is an indicator that is 1 if we fail to reject the null hypothesis that $\rho = 0$ in eq. (10).

Table A.2:
The parameter
ρ in eq. (2) is set to 1, 0.8 and 0.6, and the parameter θ in eq. (3) is set to 0.8. To conserve space, results for selected *sample sizes* of $T = [70, 200, 500, 1000]$ observations are reported in the table.

T	$\widehat{\text{DW}}$	$[\widehat{\text{DW}}]$	$\tilde{\rho}$	$[\lvert t_{\tilde{\rho}}\rvert < 1.96]$	$\hat{\rho}$	$[\lvert t_{\hat{\rho}}\rvert < 1.96]$
	Means of the statistics for $\rho = 1$					
70	0.2084	0.0000	0.8939	0.0000	0.9194	0.0000
200	0.0728	0.0000	0.9634	0.0000	0.9726	0.0000
500	0.0288	0.0000	0.9856	0.0000	0.9892	0.0000
1000	0.0145	0.0000	0.9928	0.0000	0.9946	0.0000
	Means of the statistics for $\rho = 0.8$					
70	0.5454	0.0000	0.7231	0.0000	0.7357	0.0000
200	0.4493	0.0000	0.7742	0.0000	0.7785	0.0000
500	0.4193	0.0000	0.7899	0.0000	0.7915	0.0000
1000	0.4095	0.0000	0.7950	0.0000	0.7958	0.0000
	Means of the statistics for $\rho = 0.6$					
70	0.9098	0.0026	0.5382	0.0063	0.5449	0.0058
200	0.8384	0.0000	0.5787	0.0000	0.5808	0.0000
500	0.8159	0.0000	0.5912	0.0000	0.5921	0.0000
1000	0.8075	0.0000	0.5958	0.0000	0.5962	0.0000

Appendix B: Simulation with an I(2) regressor

In the main body of the paper I studied spurious regression where the regressor is an I(1) variable. The order of integration of the dependent variable is determined by the value of ρ. When $\rho = 1$, the dependent variable is I(1). When $\rho < 1$ the dependent variable is I(0). Here I study the situation where the regressor is I(2) variable.

I generate 30,000 replications from the model

$$y_t = \beta_0 + \beta_1 x_t + u_t, \qquad [\beta_0\, \beta_1] = [0\,0]$$

$$u_t = \rho u_{t-1} + e_t, \quad u_0 = e_0, \qquad \rho = [1, 0.8, 0.6]$$

$$x_t = 2x_{t-1} - x_{t-2} + w_t \Leftrightarrow (x_t - x_{t-1}) - (x_{t-1} - x_{t-2}) = w_t, \quad x_0 = w_0, x_1 = w_1$$

and each $w_t \sim \text{iid}\,\text{Normal}(0,1)$ is independent of each $e_t \sim \text{iid}\,\text{Normal}(0,1)$

for sample sizes of $T = [15, 30, 70, 100, 200, 500, 1000, 9000]$.

For each replication round I generate $1000 + T$ observations and discard the first 1000, using only the last T observations for estimation.

Means for the estimators and t-statistic rejection rates for the true null hypothesis

For a description of the estimators in Table B see Section 3.1. Each entry in the Table B that follows contains the mean of a quantity computed over the 30,000 replication rounds. The mean of the estimate is followed by the mean of an indicator variable taking the value of 1 if the t-statistic associated with the given estimate and with the (true) null hypothesis Ho: $\beta_1 = 0$ exceeds in absolute value the appropriate critical value for Student's t random variable. In other words we have a column of the means of the estimates followed by the rejection rate under the null hypothesis. The test of the null hypothesis is always carried out at the 5% significance level. To conserve space, results for selected *sample sizes* of $T = [70, 200, 500, 1000]$ observations are reported in the table.

Table B

T	$\hat{\beta}_{OLS}$	$[\lvert t_{\hat{\beta}_{OLS}}\rvert > t_c]$	$\tilde{\beta}_{GLS}$	$[\lvert t_{\tilde{\beta}_{GLS}}\rvert > t_c]$	$\hat{\beta}_{DOLS}$	$[\lvert t_{\hat{\beta}_{DOLS}}\rvert > t_c]$	$\hat{\beta}_{CO}$	$[\lvert t_{\hat{\beta}_{CO}}\rvert > t_c]$	$\hat{\beta}_{OS}$	$[\lvert t_{\hat{\beta}_{OS}}\rvert > t_c]$
	Means of the statistics for $\rho = 1$									
70	0.0000	0.8435	-0.0004	0.0501	-0.0004	0.0501	-0.0002	0.4928	-0.0012	0.1545
200	0.0000	0.9081	-0.0000	0.0515	-0.0000	0.0515	0.0000	0.4795	0.0002	0.1635
500	-0.0000	0.9403	0.0000	0.0498	0.0000	0.0498	0.0000	0.4616	0.0000	0.1686
1000	-0.0000	0.9563	-0.0000	0.0498	-0.0000	0.0498	-0.0000	0.4450	-0.0000	0.1736
	Means of the statistics for $\rho = 0.8$									
70	-0.0000	0.5126	-0.0000	0.0458	-0.0002	0.0032	-0.0000	0.1480	-0.0005	0.1031
200	-0.0000	0.5203	-0.0000	0.0508	-0.0000	0.0002	-0.0000	0.0890	-0.0003	0.0732
500	-0.0000	0.5126	0.0000	0.0488	0.0000	0.0000	0.0000	0.0675	0.0001	0.0602
1000	0.0000	0.5153	-0.0000	0.0524	0.0000	0.0000	0.0000	0.0595	0.0000	0.0541
	Means of the statistics for $\rho = 0.6$									
70	0.0000	0.3246	0.0000	0.0512	-0.0003	0.0005	-0.0000	0.0973	-0.0001	0.0768
200	0.0000	0.3253	0.0000	0.0501	0.0001	0.0000	0.0000	0.0677	0.0002	0.0603
500	0.0000	0.3252	0.0000	0.0494	-0.0000	0.0000	0.0000	0.0574	-0.0001	0.0559
1000	0.0000	0.3268	0.0000	0.0501	-0.0000	0.0000	0.0000	0.0532	-0.0000	0.0539

Looking through the entries of the Table B, we observe the following.

1. All five estimators, including the OLS estimator are *unbiased*.
2. The t-statistic associated with the OLS estimator is divergent for the regression of I(1) on I(2) variable (the first panel). It is severely distorted, but not divergent for regressions of I(0) on I(2) variable (second and third panel).
3. The t-statistic associated with the true GLS is precisely sized for any sample size, as predicted by theory.
4. The t-statistic associated with the mis-specified first differenced OLS is correctly sized for regression of I(1) on I(2) variable (first panel). However for regressions of I(0) on I(2) variable the rejection rate does not approach the nominally stated level of the test of 5%, but instead approaches 0.
5. The t-statistics associated with the Cochrane-Orcutt feasible GLS and the one step estimator in eq. (10) exhibit notable size distortions for regressions of I(0) on I(2) variable (second and third panel), and as the sample size increases they approach the nominally stated size of the test of 5% (however they do so rather slowly). They exhibit severe size distortions for regressions of I(1) on I(2) (first panel). Note however that the t-statistics associated with the feasible versions do *not* diverge.

These results are interesting, however I leave more careful investigation of what is going on here for future research. As a practical matter, it should be noted that it is hard to miss the point that an I(2) variable is non-stationary, as just looking at plots of I(2) variables clearly reveal that they are trending. Seeing plots of I(2) variables against time, one might still wonder whether what one sees is a deterministic time trend, or stochastic trend.

References

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

Durbin, J. (1970). Testing for serial correlation in least squares regressions when some of the regressors are lagged dependent variables. *Econometrica*, 38, pp. 410–421.

Durbin, J., and Watson, G. (1950). Testing for serial correlation in least squares regressions I. *Biometrika*, 37, pp. 409–428.

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

Hayashi, F. (2000). *Econometrics*. Princeton University Press.

Kwiatkowski, D., Phillips, P., Schmidt, P. and Shin, Y. (1992). Testing the null hypothesis of stationarity against the alternative of a unit root. How sure are we that economic time series have a unit root? *Journal of Econometrics*, 54(1–3), pp. 159–178.

McCallum, B. (2010). Is the Spurious Regression Problem Spurious? *NBER Working Paper* No. 15690, pp. 1–8.

Mendenhall, W. and Sinich, T. (2003). *A second course in statistics: regression analysis*. Prentice Hall.

Pearson, K. (1897). On a form of spurious correlation which may arise when indices are used in the measurement of organs. *Proceedings of the Royal Society of London*, 60, pp. 489–498.

Phillips, P. (1986). Understanding spurious regressions in econometrics. *Journal of Econometrics,* 33(3), pp. 311–340.

Sun, Y. (2004). A Convergent t-statistic in spurious regressions. *Econometric Theory,* 20(5), pp. 943–962.

Wooldridge, J. (1991). On the Application of robust, regression-based diagnostics to models of conditional means and conditional variances. *Journal of Econometrics*, 47, pp. 5–46.

Ventosa-Santaularia, D. (2009). Spurious regression. *Journal of Probability and Statistics*, pp. 1–27.

Yule, G. (1897). On the theory of correlation. *Journal of the Royal Statistical Society,* 60(4), pp. 812–854.

Yule, G. (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), pp. 1–63.