The Linear Regression Model with Autocorrelated Errors: Just Say No to Error Autocorrelation

Anya McGuirk*
Department of Agricultural and Applied Economics,
Department of Statistics, Virginia Tech,
<mcguirk@vt.edu>

Aris Spanos
Department of Economics,
Virginia Tech,
<aris@vt.edu>

May 2002

Abstract

This paper focuses on the practice of serial correlation correcting of the Linear Regression Model (LRM) by modeling the error. Simple Monte Carlo experiments are used to demonstrate the following points regarding this practice. First, the common factor restrictions implicitly imposed on the temporal structure of y_t and $\mathbf{x}_t$ appear to be completely unreasonable for any real world application. Second, when one compares the Autocorrelation-Corrected LRM (ACLRM) model estimates with those from the (unrestricted) Dynamic Linear Regression Model (DLRM) encompassing the ACLRM, there is no significant gain in efficiency! Third, as expected, when the common factor restrictions do not hold the LRM model gives poor estimates of the true parameters and estimation of the ACLRM simply gives rise to different misleading results! On the other hand, estimates from the DLRM and the corresponding VAR model are very reliable. Fourth, the power of the usual Durbin Watson test (DW) of autocorrelation is much higher when the common factor restrictions do hold than when they do not. But, a more general test of autocorrelation is shown to perform almost as well as the DW when the common factor restrictions do hold and significantly better than the DW when the restrictions do not hold. Fifth, we demonstrate that the simple F-test suggested by Davidson and MacKinnon (1993) is quite powerful.

*Annual Meeting of the American Agricultural Economics Association, July 28-31, 2002.

1 Introduction

A key assumption underlying the Linear regression model (LRM) typically used in applied econometric studies is that of no autocorrelation. The traditional way of handling the LRM in cases where this assumption is false is to model the serial correlation explicitly using some error autocorrelation formulation, say an AR(1) process, and then use GLS to estimate the Autocorrelation-Corrected LRM (ACLRM). When one models the error term, restrictions are implicitly imposed on the structure of the observable random variables involved y_t and $\mathbf{x}_t$. The implicit (testable) restrictions implied in the AR(1) –the common factor restrictions–have been known for a long time (Sargan, 1964) yet they are still rarely tested in applied econometric studies. Further, despite the serious warnings from Henry and Mizon (1978), Sargan (1980), Spanos (1986, 1988), Hoover (1988), and Mizon (1995), inter alia, the practice of autocorrelation correcting is still common. In fact, its use may even be on the rise largely due to the increased use of spatial data which exhibit dependencies, and 'advances' in techniques for autocorrelation correcting systems of simultaneous equations, panel data models, etc.

The primary objective of this paper is make a strong case against the tradition of 'correcting' for serial correlation by modeling the error.

2 'Correcting' for serial correlation

The **Linear Regression Model** (LRM) has been the quintessential statistical model for econometric modeling since the early 20th century with pioneers like Moore and Schultz; see Morgan (1990). Yule (1921, 1926) scared econometricians away from regression by demonstrating that when using time series data regression often leads to spurious results. The first attempt to deal with the problem of regression with time series data was by Cochrane and Orcutt (1949) who proposed extending the LRM to include autocorrelated errors following a low order ARMA(p,q) formulation. They also demonstrated by simulation that the Von Neuman ratio test for autocorrelation was not very effective in detecting autocorrelated errors. Durbin and Watson (1950,1951) addressed the testing problem in the case where the linear regression model:

(1) $$y_t = \boldsymbol{\beta}^\top \mathbf{x}_t + u_t,\ t \in \mathbb{T},$$

where $\mathbf{x}_t$ is a $k \times 1$ vector, supplemented with an AR(1) error:

(2) $$u_t = \rho u_{t-1} + \varepsilon_t,\ |\rho| < 1,\ t \in \mathbb{T}.$$

Substituting (2) into (1) gives rise to:

(3) $$y_t = \boldsymbol{\beta}^\top \mathbf{x}_t + \rho y_{t-1} - \rho \boldsymbol{\beta}^\top \mathbf{x}_{t-1} + u_t,\ t \in \mathbb{T},$$

and the well known Durbin-Watson (D-W) test based on the hypotheses:

(4) $$H_0 : \rho = 0, \text{ vs. } H_0 : \rho \neq 0.$$

Since then, the traditional econometric literature has treated this extension of the LRM as providing a way to test for the presence of error autocorrelation in the data as well as a solution to the misspecification problem if one rejects H_0. That is, when the D-W test rejects H_0 the modeler adopts (3) as a way to 'correct' for serial correlation; the latter model is then estimated using Feasible Generalized Least Squares (GLS); see inter alia Green (2000).

Sargan (1964) was the first to view (3) as a restricted version of a more general model:

$$y_t = \alpha_1 y_{t-1} + \boldsymbol{\beta}_0^\top \mathbf{x}_t + \boldsymbol{\beta}_1^\top \mathbf{x}_{t-1} + u_t,\ t \in \mathbb{T}, \tag{5}$$

known as the **Dynamic Linear Regression Model** (DLRM), where the restrictions take the form:

$$H_0^{(cf)}:\ \ \boldsymbol{\beta}_1^\top - \alpha_1 \boldsymbol{\beta}_0^\top = \mathbf{0}. \tag{6}$$

Sargan proposed a likelihood ratio test for testing these, so-called *common factor restrictions,* before imposing them. His proposal was further elaborated upon by Hendry and Mizon (1978) and Sargan (1980). In an attempt to show the restrictive nature of (6), Spanos (1988) investigated the probabilistic structure of the vector stochastic process $\{\mathbf{Z}_t, t \in \mathbb{T}\}$, where $\mathbf{Z}_t := (y_t, \mathbf{x}_t^\top)^\top$ that would give rise to such restrictions. It was shown that the common factor restrictions arise naturally when $\{\mathbf{Z}_t, t \in \mathbb{T}\}$ is a Normal, Markov and Stationary process:

$$\begin{pmatrix} \mathbf{Z}_t \\ \mathbf{Z}_{t-1} \end{pmatrix} \backsim \mathsf{N}\left(\begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \boldsymbol{\Sigma}(0) & \boldsymbol{\Sigma}(1)^\top \\ \boldsymbol{\Sigma}(1) & \boldsymbol{\Sigma}(0) \end{pmatrix} \right),\ t \in \mathbb{T},$$

with a temporal covariance structure of the form:

$$\boldsymbol{\Sigma}(1) = \rho \boldsymbol{\Sigma}(0). \tag{7}$$

The sufficient conditions in (7) are 'highly unrealistic' because, as shown in Spanos (1988), they give rise to a very restrictive VAR(1) model for the $\{\mathbf{Z}_t, t \in \mathbb{T}\}$ process of the form:

$$\mathbf{Z}_t = \mathbf{A}^\top \mathbf{Z}_{t-1} + \mathbf{E}_t,\ \mathbf{E}_t \backsim \mathsf{N}(\mathbf{0}, \boldsymbol{\Omega}),\ t \in \mathbb{T},$$

$$\mathbf{A} = \begin{pmatrix} \rho & 0 \\ \mathbf{0} & \rho \mathbf{I}_k \end{pmatrix}, \quad \boldsymbol{\Omega} = (1-\rho^2)\boldsymbol{\Sigma}(0).$$

That is, y_t and $\mathbf{x}_t$ are mutually Granger non-causal and have identical AR(1) representations! Mizon (1995) elaborated on the sufficient condition and recommended that the traditional way of 'correcting for serial correlation' is a bad idea; his paper is entitled "A simple message to autocorrelation correctors: Don't". Unfortunately, that advice seems to continue to be ignored by the recent traditional literature with dire consequences for the reliability of inference based on such models.

The primary objective of this paper is to elaborate on Spanos (1988) by showing that the temporal structure (7) is not just sufficient for the common factors restrictions, but also necessary and proceed to illustrate the unreliability of inference when this is ignored.

3 A specification/respecification perspective

One possible reason why the empirical econometric practice continues to ignore the warnings concerning the inappropriateness of the 'serial correlation correction' is that this problem raises several methodological issues that have not been addressed adequately in the literature. For instance, 'why is it problematic to adopt the alternative hypothesis in the case of a D-W test?' It is generally accepted that there is no problem when one adopts the alternative in the case of a t-test for the hypothesis:

$$H_0: \beta_1 = 0,\ \text{vs.}\ H_0: \beta_1 \neq 0. \tag{8}$$

The purpose of this section is to address briefly these methodological issues in the context of the Probabilistic Reduction framework; see Spanos (1986,1995).

Despite the apparent similarity between the D-W test and a significance t-test in terms of the hypotheses being tested, the fact is that they are very different in nature. As argued in Spanos (1998,1999), the D-W test is a misspecification test, but the later is a proper Neyman-Pearson test. The crucial difference between them is that the D-W is probing beyond the boundaries of the original model, the LRM, but the t-test is probing within those boundaries. In a Neyman-Pearson test there are only two types of errors (reject the null when true and accept the null when false) because one assumes that the prespecified statistical model (LRM) contains the true model, and thus rejection of the null leaves only one other choice, the alternative, because between them they span the original model. In the case of a misspecification test one is probing beyond the boundaries of the prespecified model by extending it in specific directions; the D-W test extends the LRM by attaching an AR(1) error. A rejection of the null in a misspecification test does not entitle the modeler to infer that the extended model is true; only that the original model is misspecified! In order to infer the validity of the extended model one needs to test its own assumptions. In the case of the D-W test, if the null is rejected one can only infer that the LRM is misspecified in so far as the no autocorrelation assumption is rejected by the data and thus the data exhibit some kind of dependence. However, the type of dependence present in the data can only be established by thorough misspecification testing of alternative statistical models which allow for such dependence. The alternative model involved in a D-W test (5) is only of one of an infinite number of such models one can contemplate and so is (5); the only advantage of the latter is that it nests the former and thus if (5) is misspecified so is (3). Hence, in terms of respecifying the LRM to allow for temporal dependence the DLR (5) is considerably more realistic than (3) because it allows for a much less restrictive form of temporal dependence than (3) does.

4 Revisiting the Common Factor Restrictions

Consider the Linear Regression Model (LRM):

$$y_t = \boldsymbol{\beta}^\top \mathbf{x}_t + u_t,$$

with an AR(m) error process:

$$(9) \qquad u_t = \rho_1 u_{t-1} + \rho_2 u_{t-2} + ... + \rho_m u_{t-m} + \varepsilon_t, \quad \varepsilon_t \backsim \mathsf{NIID}(0, \sigma^2), \; t \in \mathbb{T},$$

the most typical formulation being an AR(1). Substituting the AR(m) model of the error into the main model yields:

$$(10) \qquad y_t = \boldsymbol{\beta}^\top \mathbf{x}_t + \sum_{i=1}^{m} \rho_i \left[y_{t-i} - \boldsymbol{\beta}^\top \mathbf{x}_{t-i} \right] + \varepsilon_t, \; t \in \mathbb{T}.$$

Thus, the AR(m) error formulation is simply a restricted version of the Dynamic Linear Regression Model (DLRM):

$$(11) \qquad y_t = \boldsymbol{\beta}_0^\top \mathbf{x}_t + \sum_{i=1}^{m} \left[\alpha_i y_{t-i} + \boldsymbol{\beta}_i^\top \mathbf{x}_{t-i} \right] + v_t, \; v_t \backsim \mathsf{NIID}(0, \sigma^2), \; t \in \mathbb{T}.$$

The specific common factor restrictions for (11) viewed in the context of (10) are:

$$\boldsymbol{\beta}_0 \alpha_i = -\boldsymbol{\beta}_i, i = 1, ..., m.$$

To understand the implications of these common factor restrictions in terms of the parameters of the joint distribution underlying the regression model (11), we make use of Theorem 1 in Spanos and McGuirk (2002).

Theorem 1. Consider the Linear Regression Model:

$$y_t = \gamma_0 + \gamma_1' \mathbf{x}_t + u_t,\ u_t \backsim \mathsf{NIID}(0, \sigma^2),\ t \in \mathbb{T},$$

where $\mathbf{x}_t$ is a $kx1$ vector. Under the assumptions $E(u_t|\mathbf{X}_t) = 0$ and $E(u_t^2|\mathbf{X}_t) = \sigma^2$ the model parameters $(\gamma_0, \gamma_1, \sigma^2)$ are related to the primary parameters of the stochastic process $\{\mathbf{Z}_t, t \in \mathbb{T}\}$:

$$E(\mathbf{Z}_t) := \mu = \begin{pmatrix} \mu_1 \\ \boldsymbol{\mu}_2 \end{pmatrix},\ Cov(\mathbf{Z}_t) := \Sigma = \begin{pmatrix} \sigma_{11} & \boldsymbol{\sigma}_{21}^\top \\ \boldsymbol{\sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix},$$

where $\mathbf{Z}_t \equiv (y_t, \mathbf{X}_t^\top)$ via:

$$\gamma_0 = \mu_1 - \gamma_1' \boldsymbol{\mu}_2 \text{ and } \gamma_1 = \Sigma_{22}^{-1} \sigma_{21}.$$

In the case of the unrestricted dynamic linear regression model (11) with $m = 1$, assuming for simplicity that all variables are in mean deviation form, $\mathbf{Z}_t := (y_t, \mathbf{X}_t^\top)$, $\mathbf{x}_t^\top := (y_{t-1}, x_t, x_{t-1})$, x_t is a scalar. $\{\mathbf{Z}_t, t \in \mathbb{T}\}$ is a second order stationary process, the primary parameters defining the process are:

$$(12)\quad E(\mathbf{Z}_t) := \boldsymbol{\mu} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix};\ Cov(\mathbf{Z}_t) := \boldsymbol{\Sigma} = \begin{pmatrix} \sigma_{11}(0) & \sigma_{11}(1) & \sigma_{12}(0) & \sigma_{12}(1) \\ \sigma_{11}(1) & \sigma_{11}(0) & \sigma_{12}(1) & \sigma_{12}(0) \\ \sigma_{12}(0) & \sigma_{12}(1) & \sigma_{22}(0) & \sigma_{22}(1) \\ \sigma_{12}(1) & \sigma_{12}(0) & \sigma_{22}(1) & \sigma_{22}(0) \end{pmatrix}$$

where $\sigma_{11}(i) = Cov(y_t, y_{t-i})$, $\sigma_{12}(i) = Cov(y_t, x_{t-i}) = Cov(y_{t-i}, x_t)$, and $\sigma_{22}(i) = Cov(x_t, x_{t-i})$, $i = 0, 1$. Further the model parameters of the unrestricted dynamic linear regression model are $(\alpha_1, \beta_0, \beta_1, \sigma^2)$ and these parameters are related to the primary parameters via the following system of equations:

$$(13)\qquad \boldsymbol{\beta} = \begin{pmatrix} \alpha_1 \\ \beta_0 \\ \beta 1 \end{pmatrix} = \begin{pmatrix} \sigma_{11}(0) & \sigma_{12}(1) & \sigma_{12}(0) \\ \sigma_{12}(1) & \sigma_{22}(0) & \sigma_{22}(1) \\ \sigma_{12}(0) & \sigma_{22}(1) & \sigma_{22}(0) \end{pmatrix}^{-1} \begin{pmatrix} \sigma_{11}(1) \\ \sigma_{12}(0) \\ \sigma_{12}(1) \end{pmatrix}$$

$$\sigma^2 = \sigma_{11}(0) - \begin{pmatrix} \sigma_{11}(1) & \sigma_{12}(0) & \sigma_{12}(1) \end{pmatrix} \begin{pmatrix} \alpha_1 \\ \beta_0 \\ \beta 1 \end{pmatrix}.$$

When the common factor restriction holds, $\alpha_1 = \rho$ and $\boldsymbol{\beta} = \begin{pmatrix} \rho & \beta_0 & -\rho\beta_0 \end{pmatrix}^\top$. It is instructive to solve this (restricted) non-linear system of 4 equations for $\sigma_{11}(0)$, $\sigma_{11}(1)$, $\sigma_{12}(0)$, $\sigma_{22}(1)$.[1] The solution yields:

$$(14)\qquad \sigma_{11}(1) = \sigma_{12}(1)\beta_0 + \tfrac{\rho\sigma^2}{1-\rho^2},$$

$$(15)\qquad \sigma_{11}(0) = \sigma_{12}(0)\beta_0 + \tfrac{\sigma^2}{1-\rho^2},$$

[1] This system of equations was solved using a combination of Matlab (symbolics toolkit) and Mathematica.

$$\sigma_{22}(1) = \sigma_{12}(1)/\beta_0, \tag{16}$$

$$\sigma_{22}(0) = \sigma_{12}(0)/\beta_0. \tag{17}$$

Note that (16)-(17) taken together indicate:

$$\frac{\sigma_{22}(1)}{\sigma_{22}(0)} = \frac{\sigma_{12}(1)}{\sigma_{12}(0)}$$

and thus, the common factor restrictions imply $Cov(x_t, x_{t-1}) = \lambda Var(x_t)$ and $Cov(y_t, x_{t-1}) = Cov(y_{t-1}, x_t) = \lambda Cov(y_t, x_t)$, where λ is simply an unknown constant of proportionality.

It turns out that the common factor restrictions also imply $Cov(y_t, y_{t-1}) = \lambda Var(y_t)$. To see this note that (14)-(15), in conjunction with (16)-(17) imply:

$$\rho = \frac{\sigma_{11}(1) - \sigma_{12}(1)\beta_0}{\sigma_{11}(0) - \sigma_{12}(0)\beta_0} = \frac{\sigma_{11}(1) - \lambda\sigma_{12}(0)\beta_0}{\sigma_{11}(0) - \sigma_{12}(0)\beta_0}. \tag{18}$$

Further, the (messy) equation for ρ found by expanding (13) simplifies nicely when one substitutes in $\lambda\sigma_{12}(0)$ for $\sigma_{12}(1)$ and $\lambda\sigma_{22}(0)$ for $\sigma_{22}(1)$. These simplifications yield:

$$\rho = \frac{\lambda\sigma_{12}(0)^2 - \sigma_{11}(1)\sigma_{22}(0)}{\sigma_{12}(0)^2 - \sigma_{11}(1)\sigma_{22}(0)} \tag{19}$$

Setting the expressions for ρ in (18) and (19) equal to each other, and solving for $\sigma_{11}(1)$ yields:

$$\sigma_{11}(1) = \lambda\sigma_{11}(0) \tag{20}$$

or $Cov(y_t, y_{t-1}) = \lambda Var(y_t)$. Further, substituting (20) into (19) and simplifying yields:

$$\lambda = \rho.$$

Thus, the common factor restrictions imply identical temporal structure between the observable random variables in the sense that:

$$\begin{aligned} Cov(x_t, x_{t-1}) &= \rho\, Cov(x_t, x_t);\ Cov(y_t, y_{t-1}) = \rho\, Cov(y_t, y_t); \\ Cov(y_t, x_{t-1}) &= Cov(y_{t-1}, x_t) = \rho\, Cov(y_t, x_t). \end{aligned} \tag{21}$$

The conditions (21) are identical to those given in Spanos (1988) as sufficient for the common factor restrictions to hold. The derivations above, however, show that they are also necessary.

More insight can be gained into the restrictiveness of (21), by deriving the vector autoregressive model (VAR) model based on $D(\mathbf{Z}_t|\mathbf{Z}_{t-1};\boldsymbol{\psi})$ which is derivable directly from $D(\mathbf{Z}_1, \mathbf{Z}_2, ..., \mathbf{Z}_T; \boldsymbol{\phi})$. Taking into account the restrictions in (21), $\boldsymbol{\Sigma}$ of (12) can be simplified as follows:

$$Cov(\mathbf{Z}_t) := \boldsymbol{\Sigma} = \begin{pmatrix} \sigma_{11}(0) & \rho\sigma_{11}(0) & \sigma_{12}(0) & \rho\sigma_{12}(0) \\ \rho\sigma_{11}(0) & \sigma_{11}(0) & \rho\sigma_{12}(0) & \sigma_{12}(0) \\ \sigma_{12}(0) & \rho\sigma_{12}(0) & \sigma_{22}(0) & \rho\sigma_{22}(0) \\ \rho\sigma_{12}(0) & \sigma_{12}(0) & \rho\sigma_{22}(0) & \sigma_{22}(0) \end{pmatrix} \tag{22}$$

From this variance-covariance matrix we can use the theorem above to show that the model parameters underlying the VAR(1) model:

$$\mathbf{Z}_t = \mathbf{A}^\top \mathbf{Z}_{t-1} + \mathbf{E}_t, \mathbf{E}_t \backsim \mathsf{NIID}(0, \boldsymbol{\Omega}) \tag{23}$$

are related to the primary parameters in (22) via:

$$\mathbf{A}^{\top} = \begin{pmatrix} \sigma_{11}(0) & \sigma_{12}(0) \\ \sigma_{12}(0) & \sigma_{22}(0) \end{pmatrix}^{-1} \begin{pmatrix} \rho\sigma_{11}(0) & \rho\sigma_{12}(0) \\ \rho\sigma_{12}(0) & \rho\sigma_{22}(0) \end{pmatrix}$$

$$\mathbf{\Omega} = \begin{pmatrix} \sigma_{11}(0) & \sigma_{12}(0) \\ \sigma_{12}(0) & \sigma_{22}(0) \end{pmatrix} - \begin{pmatrix} \rho\sigma_{11}(0) & \rho\sigma_{12}(0) \\ \rho\sigma_{12}(0) & \rho\sigma_{22}(0) \end{pmatrix} \mathbf{A}^{\top}$$

which simplify to:

$$\mathbf{A} = \begin{pmatrix} \rho & 0 \\ 0 & \rho \end{pmatrix}, \quad \mathbf{\Omega} = (1-\rho^2) \begin{pmatrix} \sigma_{11}(0) & \sigma_{12}(0) \\ \sigma_{12}(0) & \sigma_{22}(0) \end{pmatrix}.$$

These derivations confirm that the sufficient conditions $\mathbf{\Sigma}(1) = \rho\mathbf{\Sigma}(0)$,are also necessary in the case of $k = 1$.

5 Monte Carlo Simulations

In an attempt to illustrate the restrictive nature of 'correcting for serial correlation' by modeling the error, we consider a number of Monte Carlo experiments. These experiments relate to the linear regression model and departures from the no autocorrelation assumption. All experimental results reported are based on 10,000 replications of sample sizes $T = 25$ and $T = 50$.

5.1 Experiment 1 - Linear Regression with temporal dependence

In **Experiment 1** we generate data with two very similar implied Dynamic Linear regression Models (**1A-1B**). They differ by only one model parameter, β_1,the parameter on x_{t-1}.

Experiment 1A - Primary parameters

$$E(Y_t) = 2, \quad Var(Y_t) = 1.115, \quad Cov(Y_t, Y_{t-1}) = 0.446, \quad Cov(Y_t, X_{t-1}) = -.678,$$
$$E(X_t) = 1, \quad Var(X_t) = 1, \quad Cov(X_t, X_{t-1}) = 0.6, \quad Cov(Y_t, X_t) = -.269,$$

giving rise to an Unrestricted Dynamic Linear regression Model (UDLRM) model:

$$y_t = 1.037 + 0.6y_{t-1} + 0.7x_t - 0.9369x_{t-1}, \quad \sigma^2 = 0.4; \Re^2 = 0.641. \tag{24}$$

Experiment 1B - Primary parameters[2]

$$E(Y_t) = 2, \quad Var(Y_t) = 1.115, \quad Cov(Y_t, Y_{t-1}) = 0.669, \quad Cov(Y_t, X_{t-1}) = .42,$$
$$E(X_t) = 1, \quad Var(X_t) = 1, \quad Cov(X_t, X_{t-1}) = 0.6, \quad Cov(Y_t, X_t) = .70,$$

giving rise to the Restricted Dynamic Linear regression Model (RDLRM) model:

$$y_t = 0.52 + 0.6y_{t-1} + 0.7x_t - 0.42x_{t-1}, \quad \sigma^2 = 0.4; \Re^2 = 0.641. \tag{25}$$

Note that the true regression model in 1B is considered 'restricted' because it can also be written as:

$$y_t = 1.3 + 0.7x_t + u_t, \quad u_t = 0.6u_{t-1} + v_t, \ t \in \mathbb{T}. \tag{26}$$

[2] Experiments 1A and 1B are also similar in the sense that $Det(\Sigma) = 0.16$.

That is, for experiment 1B, the *common factor restrictions* implicitly imposed by an error AR(1) process hold.

We begin our simulation, as a typical modeler would and estimate the Normal, linear regression (LRM):

$$y_t = \alpha_0 + \beta_0 x_t + u_t,\ t \in \mathbb{T},$$

We test for first order autocorrelation using the typical Durbin-Watson (D-W) test, and by running auxiliary regressions of $\hat{u}_t$ on (i) $\hat{u}_{t-1}$ and x_t, (ii) y_{t-1}, x_{t-1}, and x_t, and (iii) only $\hat{u}_{t-1}$. These three autocorrelation misspecification tests cover the gambit of tests usually implemented in applied work. The LRM simulation results for Experiments 1A (UDLRM) and 1B (RDLRM) are reported in Tables 1A-1B.

The results in tables 1A-B suggest that the temporal dependence would be detected at a reasonably high percentage in both cases (UDLRM and RDLRM). Interestingly, despite the similar fit of the two models (based on R^2), the dependence is detected more often in the UDLRM case. Not surprisingly, given our discussions above, the D-W test is most likely to detect temporal dependence departures in the RDLRM, while the most general auxiliary regression test ($\hat{u}_t$ on y_{t-1}, x_{t-1}, and x_t) performs best in the UDLRM.

Table 1A - True: UDLRM // Estimated: LRM (OLS)

		T=25		T=50	
	True	Mean	Std	Mean	Std
$\hat{\alpha}_0$	1.037	1.6109	0.3611	1.6978	0.2456
$\hat{\beta}_0$	0.70	0.3887	0.2939	0.3023	0.2229
$\hat{\sigma}^2$	0.40	1.0754	0.4174	1.1766	0.3105
R^2	0.641	0.1556	0.1426	0.0975	0.0940
t-statistics		Mean	% reject (.05)	Mean	% reject (.05)
$\tau_{\alpha_0} = \frac{\hat{\alpha}_0 - \alpha_0}{\hat{\sigma}_{\alpha_0}}$		1.9335	0.459	3.0109	0.816
$\tau_{\beta_0} = \frac{\hat{\beta}_0 - \beta_0}{\hat{\sigma}_{\beta_0}}$		-1.4208	0.332	-2.5235	0.634
Misspecification Test		Statistic	% reject (.05)	Statistic	% reject (.05)
Durbin-Watson		0.9315	0.877	0.8559	0.997
$\hat{u}_{t-1}$ (with x_t)		10.8535	0.768	25.057	0.991
y_{t-1}, x_{t-1}(with x_t)		21.948	0.984	0.488	1.00
$\hat{u}_{t-1}$		9.00	0.794	23.271	0.992

Tables 1A-1B also illustrate that in the case of the UDLRM the estimators of α_0, β_0, and σ^2 are badly biased and any inference associated with these parameters will be invalid. Further, these problems are accentuated as $T \to \infty$. In contrast, yet consistent with theory, $\hat{\beta}_0$ in the RDLRM appears unbiased, while $\hat{\sigma}^2$ is biased. Interestingly, $\hat{\alpha}_0$ is a reasonably good estimator of the intercept in the AR(1) model (26), where the common factor restrictions are imposed, and a biased estimator of the intercept in the general model formulation (25) where $\alpha_0 = 0.52$, even though the errors have not been modeled as an AR(1) process.

Table 1B- True: RDLRM // Estimated: LRM (OLS)					
		T=25		T=50	
	True	Mean	Std	Mean	Std
$\hat{\alpha}_0$	0.52	1.320	0.2382	1.3104	0.1671
$\hat{\beta}_0$	0.70	0.6810	0.2216	0.6896	0.1610
$\hat{\sigma}^2$	0.40	0.5512	0.1696	0.5855	0.1255
R^2	0.641	0.4237	0.1717	0.4307	0.1297
t-statistics		Mean	% reject (.05)	Mean	% reject (.05)
$\tau_{\alpha_0} = \frac{\hat{\alpha}_0 - \alpha_0}{\hat{\sigma}_{\alpha_0}}$		3.733	0.919	5.1127	0.997
$\tau_{\beta_0} = \frac{\hat{\beta}_0 - \beta_0}{\hat{\sigma}_{\beta_0}}$		-0.1266	0.163	-0.0950	0.168
Misspecification Test		Statistic	% reject (.05)	Statistic	% reject (.05)
Durbin-Watson		1.0699	0.743	0.9370	0.980
$\hat{u}_{t-1}$ (with x_t)		8.1093	0.607	21.175	0.958
y_{t-1}, x_{t-1}(with x_t)		5.491	0.576	12.112	0.941
$\hat{u}_{t-1}$		6.195	0.633	19.011	0.961

Suppose that, on the basis of the D-W tests in 1A and 1B, the modeler decided to 'correct' the apparent autocorrelation problem by adopting the alternative of an LRM with an AR(1) error model (see (3)). Tables 2A-2B give simulation results for the models re-estimated using an iterative Cochrane-Orcutt correction (EGLS). As expected, for the RDLRM case, where the common factor restrictions hold, inferences regarding β_0 and α_0^*, the intercept in (26), are reliable. In contrast in the UDLRM, though the EGLS estimators seems to do better than the OLS estimators of the LRM, both are still biased and the extent of the bias increases with T.

Tables 2A-2B also report the results a modeler would likely obtain if the common factor restrictions imposed by implementing EGLS are tested. The first test is the approximate F-test recommended in Davidson and MacKinnon (1993) and the second is the more theoretically appealing Likelihood ratio (LR) test (see Spanos, 1986). The last test is a Wald test of the restrictions implicitly imposed on the VAR(1) model by the AR(1) error formulation (see above). Specifically, we test the restriction that $\mathbf{A} = \begin{pmatrix} \rho & 0 \\ 0 & \rho \end{pmatrix}$, in (23) estimated using Iterative Seemingly Unrelated Regression.

When the common factor restrictions do not hold, all three of these tests have particularly good power. Somewhat surprisingly the more appealing LR test is the least powerful of all. When the common factors do hold (RDLRM), the actual size of the approximate F-test is close to the nominal size and the actual size of the Wald test is high yet it does approach the nominal as T increases. For the LR, the actual size is low and does not improve with T. To see the extent to which the size distortions affect the results, we also report the empirically adjusted percentage of rejections for both the Wald and LR tests. As can be seen, the LR test, still underperforms relative to the other 2 tests. What is encouraging from a practical point of view, is the ability of the easy to implement approximate F-test to detect when the common factor restrictions do not hold.

Also, of interest, in terms of the claims of this paper is the estimator of ρ in the VAR(1). Specifically, it is on interest to see how this estimator of ρ differs from that obtained using iterative Cochrane-Orcutt. The last row in Tables 2A and 2B, report the mean and standard deviations of $\hat{\rho}$. As illustrated, in the RDLRM, where the common factor restrictions hold, the (fully restricted) VAR(1) estimator of $\hat{\rho}$ easily outperforms the EGLS estimator. To illustrate more clearly the differences in these two estimators, figure 1, presents the two associated smoothed histogram of $\hat{\rho}$.

Table 2A- True: UDLRM // Estimated: LRM (EGLS)					
		T=25		T=50	
	True	Mean	Std	Mean	Std
$\hat{\alpha}_0$	1.037	1.2035	0.8025	1.1962	0.6130
$\hat{\beta}_0$	0.70	0.7897	0.1983	0.8044	0.1321
$\hat{\rho}$		0.6830	0.1856	0.7442	0.1056
$\hat{\sigma}^2$	0.40	0.6101	0.1384	0.6160	0.0903
t-statistics		Mean	% reject (.05)	Mean	% reject (.05)
$\tau_{\alpha_0} = \frac{\hat{\alpha}_0 - \alpha_0}{\hat{\sigma}_{\alpha_0}}$		0.3665	0.025	0.3666	0.004
$\tau_{\beta_0} = \frac{\hat{\beta}_0 - \beta_0}{\hat{\sigma}_{\beta_0}}$		0.4619	0.086	0.7709	0.1138
$\tau_{\rho} = \frac{\hat{\rho}}{\hat{\sigma}_{\rho}}$		5.42	0.940	8.54	0.999
AR(1) Test		Statistic	% reject (.05)	Statistic	% reject (.05)
Com. Factor F-test		14.497	0.931	27.992	0.999
Com. Fact.–LR test		5.387	0.705 (0.787)*	10.073	0.985 (0.998)*
Com. Fact.–VAR		20.40	0.918 (0.884)*	34.464	0.999 (0.997)*
Rest. VAR(1)		Mean	Std.	Mean	Std.
$\hat{\rho}$		0.5743	0.1415	0.6358	0.932

Table 2B- True: RDLRM // Estimated: LRM (EGLS)					
		T=25		T=50	
	True	Mean	Std	Mean	Std
$\hat{\alpha}_0^* = \hat{\alpha}_0(1-\hat{\rho})$	1.30	1.3082	0.3095	1.3039	0.1489
$\hat{\beta}_0$	0.70	0.6914	0.1698	0.6963	0.1148
$\hat{\rho}$	0.60	0.4748	0.2080	0.5395	0.1289
$\hat{\sigma}^2$	0.40	0.3814	0.0418	0.3912	0.0201
t-statistics		Mean	% reject (.05)	Mean	% reject (.05)
$\tau_{\alpha_0} = \frac{\hat{\alpha}_0 - \alpha_0^*}{\hat{\sigma}_{\alpha_0}}$		0.0338	0.017	0.0171	0.959
$\tau_{\beta_0} = \frac{\hat{\beta}_0 - \beta_0}{\hat{\sigma}_{\beta_0}}$		-0.0573	0.085	-0.0344	0.06
$\tau_{\rho} = \frac{\hat{\rho} - \rho}{\hat{\sigma}_{\rho}}$		-0.529	0.128	-0.3996	0.076
AR(1) Test		Statistic	% reject (.05)	Statistic	% reject (.05)
Common Factor F-test		1.481	0.076	1.2271	0.065
Com. Fact.–LR test		0.7041	0.012	0.560	0.005
Com. Fact.–VAR		4.316	0.141	3.65	0.098
$\hat{\rho}$ (VAR)		0.5028	0.1298	0.5537	0.0866

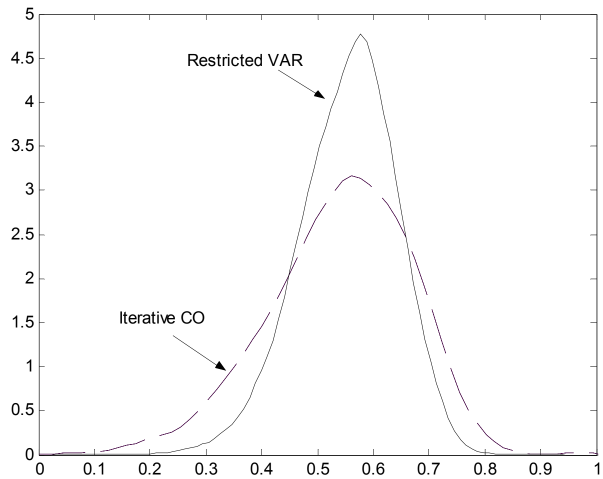

Fig. 1: Histograms of estimators of ρ

Given the superiority of the fully restricted VAR(1) estimator of $\hat{\rho}$, we run additional simulations to explore the properties of this estimator as fewer and fewer of the restrictions of the VAR(1) were imposed. The results from these simulations are reported in Tables 3A-3B. As, one would expect, given the analytical results above, the estimator of $\hat{\rho}$ becomes less efficient (and more biased) as the number of restrictions imposed are reduced. Interestingly, the estimator of $\hat{\rho}$ from each VAR scenario analyzed seems to outperform the EGLS estimator.

Table 4A - Experiment 1, T=25

	True: RDLRM				**True: UDLRM**	
	Summary Statistics for $\hat{\rho}$ ($\rho = 0.6$)				Test of Restrictions	
Model-Restrictions // Test*	Mean	Std	Min	Max	Mean	% reject (.05)
EGLS-CF // F-test	0.4748	0.2080	-0.4434	1.0845	14.497	0.931
VAR-r1 // Wald $\chi^2(1)$	0.4834	0.1809	-0.3552	0.9249	11.559	0.747
VAR-r2 // Wald test $\chi^2(2)$	0.4887	0.1564	-0.1700	0.9285	12.585	0.802
VAR-r3 // Wald test $\chi^2(3)$	0.5028	0.1298	-0.0689	0.8901	20.40	0.918

*VAR Restrictions: r1: $x_{t-1} = 0$ in y_t;r2: r1 & $y_{t-1} = 0$ in x_t;r3:r2 & x_{t-1} in x_t and y_{t-1} in y_t same parm. GLS-Common Factor restrictions (CF)

Table 4B- Experiment 1, T=50

	True: RDLRM				**True: UDLRM**	
	Summary Statistics for $\hat{\rho}$ ($\rho = 0.6$)				Test of Restrictions	
Model-Restrictions // Test*	Mean	Std	Min	Max	Mean	% reject (.05)
EGLS-CF // F-test	0.5395	0.1289	-0.0564	1.0164	27.992	0.999
VAR-r1 // Wald $\chi^2(1)$	0.5414	0.1228	-0.0767	0.8632	20.318	0.975
VAR-r2 // Wald test $\chi^2(2)$	0.5454	0.1033	-0.0134	0.8572	21.291	0.984
VAR-r3 // Wald test $\chi^2(3)$	0.5537	0.0866	0.1331	0.8328	34.464	0.999

*VAR Restrictions: r1: $x_{t-1} = 0$ in y_t;r2: r1 & $y_{t-1} = 0$ in x_t;r3:r2 & x_{t-1} in x_t and y_{t-1} in y_t same parm. GLS-Common Factor restrictions (CF)

The last two tables of simulation results summarize the implications of estimating an unrestricted DLRM (tables 5A-5B). These tables indicate that estimation results are very accurate and the usual t-tests are reliable, whether or not the common factor restrictions hold. That is, even in the RDLRM case, estimating the DLRM would yield very reliable inferences. In fact, a comparison of the EGLS results with these unrestricted DLRM results, for the RDLRM case, suggests no advantage to using EGLS even when the restrictions hold! Given the unrealistic nature of the common factor restrictions, and the potential for unreliable inferences when the restrictions do not hold, estimation of error AR(1) type models is not recommended!

Table 3A- True: UDLRM // Estimated: UDLRM (OLS)

		T=25		T=50	
	True	Mean	Std	Mean	Std
$\hat{\alpha}_0$	1.037	1.1643	0.2785	1.0998	0.1833
$\hat{\beta}_0$	0.70	0.6997	0.1669	0.6995	0.1138
$\hat{\alpha}_1 = \hat{\rho}$	0.60	0.5217	0.1447	0.5635	0.0901
$\hat{\beta}_1$	-0.937	-0.9069	0.1854	-0.9265	0.1211
$\hat{\sigma}^2$	0.40	0.3982	0.0432	0.3995	0.0203
R^2	0.641	0.6858	0.0913	0.6910	0.0654
t-statistics		Mean	% reject (.05)	Mean	% reject (.05)
$\tau_{\alpha_0} = \frac{\hat{\alpha}_0 - \alpha_0}{\hat{\sigma}_{\alpha_0}}$		0.3995	0.052	0.2885	0.037
$\tau_{\beta_0} = \frac{\hat{\beta}_0 - \beta_0}{\hat{\sigma}_{\beta_0}}$		-0.0039	0.057	-0.0049	0.049
$\tau_{\alpha_1} = \frac{\hat{\alpha}_1 - \alpha_1}{\hat{\sigma}_{\alpha_1}}$		-0.4898	0.082	-0.3521	0.064
$\tau_{\beta_1} = \frac{\hat{\beta}_1 - \beta_1}{\hat{\sigma}_{\beta_1}}$		0.1531	0.069	0.0814	0.054

Table 3B- True: RDLRM // Estimated: UDLRM (OLS)

		T=25		T=50	
	True	Mean	Std	Mean	Std
$\hat{\alpha}_0$	0.52	0.7034	0.3270	0.6116	0.2048
$\hat{\beta}_0$	0.70	0.6958	0.1683	0.6985	0.1143
$\hat{\alpha}_1 = \hat{\rho}$	0.60	0.4650	0.1909	0.5316	0.1248
$\hat{\beta}_1$	-0.42	-0.3290	0.2119	-0.3734	0.1440
$\hat{\sigma}^2$	0.40	0.3963	0.0439	0.3989	0.0207
R^2	0.641	0.6053	0.1270	0.6196	0.0920
t-statistics		Mean	% reject (.05)	Mean	% reject (.05)
$\tau_{\alpha_0} = \frac{\hat{\alpha}_0 - \alpha_0}{\hat{\sigma}_{\alpha_0}}$		0.4998	0.070	0.3706	0.0498
$\tau_{\beta_0} = \frac{\hat{\beta}_0 - \beta_0}{\hat{\sigma}_{\beta_0}}$		-0.0276	0.058	-0.0129	0.050
$\tau_{\alpha_1} = \frac{\hat{\alpha}_1 - \alpha_1}{\hat{\sigma}_{\alpha_1}}$		-0.6420	0.096	-0.4800	0.072
$\tau_{\beta_1} = \frac{\hat{\beta}_1 - \beta_1}{\hat{\sigma}_{\beta_1}}$		0.4220	0.082	0.3130	0.068

6 Conclusion

This main results of this paper can be summarized quite simply. First, the restrictions implicitly imposed on the temporal dependence structure of y_t and x_t when an AR(1) error formulation is adopted are completely unreasonable for any real world application. We show that the conditions:

$$\begin{aligned} Cov(x_t, x_{t-1}) &= \rho\, Cov(x_t, x_t);\ Cov(y_t, y_{t-1}) = \rho\, Cov(y_t, y_t); \\ Cov(y_t, x_{t-1}) &= Cov(y_{t-1}, x_t) = \rho\, Cov(y_t, x_t) \end{aligned}$$

are both necessary and sufficient in the one regressor case, and we show that the implied VAR(1) model is absurdly restrictive. Second, when one compares the Autocorrelation-Corrected LRM (ACLRM) model estimates with those from the (unrestricted) Dynamic Linear Regression Model (DLRM) encompassing the ACLRM, there is no significant gain in efficiency! Third, as expected, when the common factor restrictions do not hold the LRM model gives poor estimates of the true parameters and estimation of the ACLRM simply gives rise to different misleading results! On the other hand, estimates from the DLRM and the corresponding VAR model are very reliable. Fourth, the power of the usual Durbin Watson test (DW) of autocorrelation is much higher when the common factor restrictions do hold than when they do not. But, a more general test of autocorrelation is shown to perform almost as well as the DW when the common factor restrictions do hold and significantly better than the DW when the restrictions do not hold. Fifth, we demonstrate that the simple F-test suggested by Davidson and MacKinnon (1993) is quite powerful.

References

[1] Davidson, R. and J. G. MacKinnon (1993), *Estimation and Inference in Econometrics*, Oxford University Press, Oxford.

[2] Durbin, J. and G. S. Watson (1950) "Testing for serial correlation in least squares regression I", *Biometrika*, **37**, 409-428.

[3] Durbin, J. and G. S. Watson (1951) "Testing for serial correlation in least squares regression II", *Biometrika*, **38**, 159-178.

[4] Greene, W. H. (2000), *Econometric Analysis*, 4rd ed., Prentice-Hall, New Jersey.

[5] Hendry, D. F. and G. E. Mizon (1978) "Serial correlation as a convenient simplification, not a nuisance: A comment on a study of the demand for money by the Bank of England," *Economic Journal*, **88**, 549-563.

[6] Hoover, K. D. (1988) "On the pitfalls of untested common-factor restrictions: The case of the inverted Fisher hypothesis", *Oxford Bulletin of Economics and Statistics*, **50**, 125-138.

[7] Morgan, M. S. (1990) *The history of econometric ideas*, Cambridge University Press, Cambridge.

[8] Sargan, J. D. (1964) "Wages and Prices in the United Kingdon: A study in econometric methodology," in P.E. Hart, G. Mills and J.K. Whitaker, eds., *Econometric Analysis for Economic Planning*, Butterworths, London.

[9] Sargan, J. D. (1980) "Some tests of dynamic specification for a single equation," *Econometrica*, **48**, 879-897.

[10] Spanos, A. (1986), *Statistical Foundations of Econometric Modelling*, (1986), Cambridge University Press, Cambridge.

[11] Spanos, A. (1995) "On theory testing in Econometrics: modeling with nonexperimental data", *Journal of Econometrics,* **67**, 189-226.

[12] Spanos, A. (1988) "Error-autocorrelation revisited: the AR(1) case", *Econometric Reviews*, 1988, **7**, pp. 285-94.

[13] Spanos, A. (1998) "Econometric Testing", pp. 116-130, in *The Handbook of Economic Methodology*, edited by J. B. Davis, D. W. Hands and U. Maki, 1998, Edward Elgar.

[14] Spanos, A. (1999) *Probability Theory and Statistical Inference: Econometric modeling with observational data*, Cambridge University Press, Cambridge.

[15] Spanos, A. and A. McGuirk (2002) "The problem of near-multicollinearity revisited: erratic vs systematic volatility", (with Anya McGuirk), *The Journal of Econometrics*, **108**, 365-393.

[16] Yule, G.U. (1921) "On the time-correlation problem", *Journal of the Royal Statistical Society,* **84**, 497-526.

[17] 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**, 2-41