Revisiting Error Autocorrelation Correction: Common Factor Restrictions and Granger Non-Causality

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>

American Agricultural Economics Association Meetings, 2004
Key words: Linear Regression Model, error autocorrelation, common factor restrictions, Granger causality, Durbin- Watson test, OLS, GLS, temporal dependence, Dynamic Linear Regression Model.
JEL classification: C2, C4

Abstract

This paper demonstrates that linear regression models with an AR(1) error structure implicitly assume that y_t does not Granger cause any of the exogenous variables in $\mathbf{X}_t$. An indirect test of the common factor restrictions based on this Granger non-causality is proposed and shown to outperform existing tests.

1 Introduction

The **Linear Regression Model** (LRM) first used by pioneers like Moore and Schultz, has been the quintessential statistical model for econometric modeling since the early 20th century; see Morgan (1990). Yule (1921, 1926) scared econometricians away from the LRM for time series data by demonstrating that such regressions often lead to spurious results. The first attempt to address this problem 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 this testing problem in the case of the linear regression model:

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

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

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

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

$$H_0 : \rho = 0 \text{ vs. } H_1 : \rho \neq 0. \tag{3}$$

The traditional econometrics 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. Under this approach, error autocorrelation is viewed as a problem of efficiency for the Ordinary Least Squares (OLS) estimator $\widehat{\boldsymbol{\beta}}_{OLS} = (\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{y}$. It is argued that in the presence of error autocorrelation:

$$E(\mathbf{u}\mathbf{u}^\top \mid \mathbf{X}) = \boldsymbol{\Omega} \neq \sigma^2\mathbf{I}_T,$$

the OLS estimator maintains its unbiasedness and consistency, but it is no longer as efficient as the Generalized Least Squares (GLS) estimator $\widehat{\boldsymbol{\beta}}_{GLS} = (\mathbf{X}^\top\boldsymbol{\Omega}^{-1}\mathbf{X})^{-1}\mathbf{X}^\top\boldsymbol{\Omega}^{-1}\mathbf{y}$. The traditional way to deal with the inefficiency of OLS is to adopt the Autocorrelation-Corrected LRM (ACLRM):

$$y_t = \boldsymbol{\beta}^\top \mathbf{x}_t + \rho y_{t-1} - \rho\boldsymbol{\beta}^\top \mathbf{x}_{t-1} + \varepsilon_t,\ t \in \mathbb{T}. \tag{4}$$

That is, when the D-W test rejects H_0, the modeler adopts (4), estimated using GLS, as a way to 'correct' for serial correlation (see inter alia, Judge et al, 1985, Greene, 2000).

The practice of adopting the alternative model when the data reject the no-autocorrelation assumption is often inappropriate. The problem is that the presence of *residual autocorrelation* is interpreted as evidence *for* the ACLRM. This is an example of the classic **fallacy of rejection**: 'evidence *against* the null is interpreted as evidence *for* the alternative'. The ACLRM is presumed to be 'the' appropriate model, even though the *residual autocorrelation* could have arisen in numerous alternative ways, one of which is that the error follows an AR(1) process. It goes without saying that if the appropriate model is not the ACLRM, the OLS estimator is no longer unbiased or consistent, and the ACLRM simply constitutes another misspecified model; see Spanos (1986). Hence, adopting the ACLRM does not usually improve the reliability of inference.

Sargan (1964) was the first to view (4) as a restricted version of a more general statistical model:

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

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

$$H_0: \quad \boldsymbol{\beta}_1 - \boldsymbol{\beta}_0 \alpha_1 = \mathbf{0}. \tag{6}$$

He proposed a likelihood ratio test for these so-called *Common Factor (CF) restrictions*. This test is valid under the presumption that the DLRM is statistically adequate; its probabilistic assumptions are data-acceptable. Sargan's proposal was further elaborated upon by Hendry and Mizon (1978) and Sargan (1980) who suggested testing the CF restrictions before imposing them. Despite additional warnings concerning the unrealistic nature of the CF restrictions from Hoover (1988) and Mizon (1995), inter alia, the practice of autocorrelation correction without testing the CF restrictions 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 (see Anselin, 2001), and 'advances' in techniques for autocorrelation correcting systems of simultaneous equations and panel data models (see Greene, 2000).

In an attempt to demonstrate the restrictive nature of the ACLRM (4), Spanos (1988) investigated the probabilistic structure of the vector stochastic process $\{\mathbf{Z}_t,\ t \in \mathbb{T}\}$, $\mathbf{Z}_t := (y_t, \mathbf{x}_t^\top)^\top$, that would give rise to the CF restrictions (6). It was shown that the CF 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 Vector AutoRegressive (VAR) model for $\{\mathbf{Z}_t,\ t \in \mathbb{T}\}$:

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

$$\mathbf{A} = \rho \mathbf{I}_{k+1},\ \boldsymbol{\Omega} = (\boldsymbol{\Sigma}(0) - \rho^2 \boldsymbol{\Sigma}(0)).$$

That is, they imply y_t and $\mathbf{X}_t$ are mutually Granger non-causal and have "largely identical temporal structure". Mizon (1995), in a paper entitled "A Simple Message to Autocorrelation Correctors: Don't," elaborated on these sufficient conditions and recommended that the traditional way of 'correcting for serial correlation' is a bad idea. Unfortunately, that advice is largely ignored by the recent applied econometrics literature. As we will demonstrate, the consequences are very serious in terms of the reliability of inference based on such models.

In this paper, we elaborate on Spanos (1988) and Mizon (1995) by deriving necessary and sufficient conditions for the CF restrictions. Based on these conditions, we propose a new, easy-to-implement test of the common factor restrictions in the context of the VAR model. We then conduct Monte Carlo experiments to examine the relative performance of the LRM, the ACLRM, and the (unrestricted) DLRM when the common factor restrictions do and do not hold. We also examine the performance of various popular misspecification and common factor restriction tests under the different situations. Finally, we also briefly investigate the performance of the Heteroskedastic and Autocorrelation Consistent standard errors (HAC) proposed by Newey and West (1987) in dealing with the unreliability of inference issue.

2 Revisiting the Common Factor Restrictions

In this section, we try to elucidate the constrictive nature of the CF restrictions by examining the implicit restrictions imposed on the probabilistic structure of the observable vector stochastic process $\{\mathbf{Z}_t,\ t \in \mathbb{T}\}$, where $\mathbf{Z}_t := (y_t, \mathbf{x}_t^\top)^\top$. It is argued that by modeling the error term one implicitly imposes restrictions on the probabilistic structure of $\{\mathbf{Z}_t,\ t \in \mathbb{T}\}$, which are both unrealistic and unnecessary. The argument is based on deriving both *necessary* and *sufficient conditions* on the probabilistic structure of $\{\mathbf{Z}_t,\ t \in \mathbb{T}\}$ giving rise to the CF restrictions (6). We first establish that if the ACLRM holds then the implied parameterization $\boldsymbol{\phi}^*$ constitutes a restrictive form of $\boldsymbol{\phi}$ of the joint distribution $D(\mathbf{Z}_t, \mathbf{Z}_{t-1}; \boldsymbol{\phi})$[1]. We then show that if $\boldsymbol{\phi}^*$ is assumed, the reduction:

$$\begin{aligned} D(\mathbf{Z}_t, \mathbf{Z}_{t-1}; \boldsymbol{\phi}^*) &= D(\mathbf{Z}_t|\mathbf{Z}_{t-1}; \boldsymbol{\psi}_1^*) \cdot D(\mathbf{Z}_{t-1}; \boldsymbol{\psi}_2^*) = \\ &= D(y_t|\mathbf{X}_t, \mathbf{Z}_{t-1}; \boldsymbol{\theta}_1^*) \cdot D(\mathbf{X}_t|\mathbf{Z}_{t-1}; \boldsymbol{\theta}_2^*) \cdot D(\mathbf{Z}_{t-1}; \boldsymbol{\psi}_2^*), \end{aligned}$$

gives rise to the ACLRM based on $D(y_t|\mathbf{X}_t, \mathbf{Z}_{t-1}; \boldsymbol{\theta}_1^*)$, $\boldsymbol{\theta}_1^* := (\rho, \boldsymbol{\beta}, \sigma^2)$. It is then shown that the parameterization of the Vector Autoregressive (VAR) model ($\boldsymbol{\psi}_1^*$), specified in terms of $D(\mathbf{Z}_t|\mathbf{Z}_{t-1}; \boldsymbol{\psi}_1^*)$, is particularly useful in elucidating the nature of the CF restrictions.

The ACLRM is specified by:

$$y_t = \boldsymbol{\beta}^\top \mathbf{x}_t + u_t, \qquad u_t = \rho u_{t-1} + \varepsilon_t, \quad |\rho| < 1, \quad t \in \mathbb{T},$$

(8)

$$\begin{aligned} &[\text{a1}]\ E(\varepsilon_t) = 0,\ Var(\varepsilon_t) = \sigma^2,\ Cov(\varepsilon_t, \varepsilon_s) = 0, t \neq s, \\ &[\text{a2}]\ Cov(\mathbf{X}_t, u_s) = \mathbf{0},\ \forall\, (t, s) \in \mathbb{T}. \end{aligned}$$

Given the ACLRM (8) (including [a1]-[a2]), we can derive the implicit statistical parameterizations for the model parameters $(\boldsymbol{\beta}, \rho, \sigma^2)$ in terms of the primary parameters $\boldsymbol{\phi}$ of the joint distribution $D(\mathbf{Z}_t, \mathbf{Z}_{t-1}; \boldsymbol{\phi})$. In presenting the results we use the following notation for the variance-covariance of $(y_t, y_{t-1}, \mathbf{X}_t, \mathbf{X}_{t-1})$:

$$Cov(y_t, y_{t-1}, \mathbf{x}_t, \mathbf{x}_{t-1}) = \boldsymbol{\Sigma} = \begin{pmatrix} \sigma_{11}(0) & \sigma_{11}(1) & \boldsymbol{\sigma}_{21}^\top(0) & \boldsymbol{\sigma}_{21}^\top(1) \\ \sigma_{11}(1) & \sigma_{11}(0) & \boldsymbol{\sigma}_{21}^\top(1) & \boldsymbol{\sigma}_{21}^\top(0) \\ \boldsymbol{\sigma}_{21}(0) & \boldsymbol{\sigma}_{21}(1) & \boldsymbol{\Sigma}_{22}(0) & \boldsymbol{\Sigma}_{22}(1) \\ \boldsymbol{\sigma}_{21}(1) & \boldsymbol{\sigma}_{21}(0) & \boldsymbol{\Sigma}_{22}(1) & \boldsymbol{\Sigma}_{22}(0) \end{pmatrix} \tag{9}$$

We adopt the simplifying assumption that all random variables have mean zero, without any loss of generality.

Theorem 1. The mapping between the primary parameters $\boldsymbol{\Sigma}$ and the parameters $\boldsymbol{\theta}_1^* := (\rho, \boldsymbol{\beta}, \sigma^2)$ of the ACLRM (8) takes the form:

[1] $Var(u_t^2) = \sigma_{uu}(0) = \frac{\sigma^2}{1-\rho^2}$ **[4]** $Cov(\mathbf{x}_t, y_{t-1}) = \boldsymbol{\sigma}_{21}(1) = \boldsymbol{\Sigma}_{22}(1)\boldsymbol{\beta}$

[2] $Cov(u_t u_{t-1}) = \sigma_{uu}(1) = \frac{\rho\sigma^2}{1-\rho^2}$ **[5]** $Var(y_t) = \boldsymbol{\sigma}_{11}(0) = \boldsymbol{\beta}^\top \boldsymbol{\Sigma}_{22}(0)\boldsymbol{\beta} + \frac{\sigma^2}{1-\rho^2}$

[3] $Cov(\mathbf{x}_t, y_t) = \sigma_{21}(0) = \Sigma_{22}(0)\boldsymbol{\beta}$ **[6]** $Cov(y_t, y_{t-1}) = \boldsymbol{\sigma}_{11}(1) = \boldsymbol{\beta}^\top \boldsymbol{\Sigma}_{22}(1)\boldsymbol{\beta} + \frac{\rho\sigma^2}{1-\rho^2}$.

[1] The one lag restriction follows from the Markovness of the process.

See Appendix A for the derivations.

Using [3]-[6] the variance-covariance matrix of $(y_t, y_{t-1}, \mathbf{X}_t, \mathbf{X}_{t-1})$ implied by the ACLRM is:

$$(10)\qquad \boldsymbol{\Sigma}^* = \begin{pmatrix} \boldsymbol{\beta}^\top \boldsymbol{\Sigma}_{22}(0)\boldsymbol{\beta} + \frac{\sigma^2}{1-\rho^2} & \boldsymbol{\beta}^\top \boldsymbol{\Sigma}_{22}(1)\boldsymbol{\beta} + \frac{\rho\sigma^2}{1-\rho^2} & \boldsymbol{\beta}^\top \Sigma_{22}(0) & \boldsymbol{\beta}^\top \Sigma_{22}(1) \\ \boldsymbol{\beta}^\top \boldsymbol{\Sigma}_{22}(1)\boldsymbol{\beta} + \frac{\rho\sigma^2}{1-\rho^2} & \boldsymbol{\beta}^\top \boldsymbol{\Sigma}_{22}(0)\boldsymbol{\beta} + \frac{\sigma^2}{1-\rho^2} & \boldsymbol{\beta}^\top \Sigma_{22}(1) & \boldsymbol{\beta}^\top \Sigma_{22}(0) \\ \Sigma_{22}(0)\boldsymbol{\beta} & \Sigma_{22}(1)^\top \boldsymbol{\beta} & \boldsymbol{\Sigma}_{22}(0) & \boldsymbol{\Sigma}_{22}(1) \\ \Sigma_{22}(1)^\top \boldsymbol{\beta} & \Sigma_{22}(0)\boldsymbol{\beta} & \boldsymbol{\Sigma}_{22}(1) & \boldsymbol{\Sigma}_{22}(0) \end{pmatrix}$$

Note that the sufficient conditions in Spanos (1988), $\boldsymbol{\Sigma}(1) = \rho\boldsymbol{\Sigma}(0)$, follow from [3]-[6] above:

$$(11)\qquad \begin{array}{rl} \boldsymbol{\beta} = \Sigma_{22}(0)^{-1}\sigma_{21}(0) & = \Sigma_{22}(1)^{-1}\sigma_{21}(1). \\ \sigma^2 = \left[\sigma_{11}(0) - \boldsymbol{\beta}^\top \boldsymbol{\Sigma}_{22}(0)\boldsymbol{\beta}\right](1-\rho^2) & = \left(\frac{1}{\rho}\right)\left[\sigma_{11}(1) - \boldsymbol{\beta}^\top \boldsymbol{\Sigma}_{22}(1)\boldsymbol{\beta}\right](1-\rho^2), \end{array}$$

giving rise to:

$$(12)\qquad \sigma_{11}(1) - \rho\sigma_{11}(0) = \boldsymbol{\beta}^\top \left[\boldsymbol{\Sigma}_{22}(1) - \rho\boldsymbol{\Sigma}_{22}(0)\right]\boldsymbol{\beta}.$$

Hence, $\sigma_{11}(1) = \rho\sigma_{11}(0)$ iff $\boldsymbol{\Sigma}_{22}(1) = \rho\boldsymbol{\Sigma}_{22}(0)$.

In order to derive **necessary** and **sufficient conditions** for the common factor restrictions we make use of $\boldsymbol{\Sigma}^*$ in (10) and the following lemma from Spanos and McGuirk (2002).

Lemma 1. Consider the Linear Regression Model:

$$y_t = \gamma_0 + \boldsymbol{\gamma}_1^\top \mathbf{x}_t + u_t,\ u_t \backsim \mathsf{IID}(0, \sigma_u^2),\ t \in \mathbb{T},$$

where $\mathbf{x}_t$ is a $k \times 1$ vector. Under the assumptions $E(u_t|\mathbf{X}_t) = 0$ and $E(u_t^2|\mathbf{X}_t) = \sigma_u^2$, the model parameters $(\gamma_0, \boldsymbol{\gamma}_1, \sigma_u^2)$ are related to the primary parameters of the stochastic process $\{\mathbf{Z}_t, t \in \mathbb{T}\}, \mathbf{Z}_t \equiv (y_t, \mathbf{X}_t^\top)^\top$:

$$E(\mathbf{Z}_t) := \mu = \begin{pmatrix} \mu_y \\ \boldsymbol{\mu}_x \end{pmatrix},\ Cov(\mathbf{Z}_t) := \boldsymbol{\Sigma} = \begin{pmatrix} \sigma_{yy} & \boldsymbol{\sigma}_{xy}^\top \\ \boldsymbol{\sigma}_{xy} & \boldsymbol{\Sigma}_{xx} \end{pmatrix},$$

assuming $\boldsymbol{\Sigma} > \mathbf{0}$, via:

$$(13)\qquad \gamma_0 = \mu_y - \gamma_1^\top \boldsymbol{\mu}_x,\quad \boldsymbol{\gamma}_1 = \boldsymbol{\Sigma}_{xx}^{-1}\boldsymbol{\sigma}_{xy},\quad \sigma_u^2 = \sigma_{yy} - \boldsymbol{\sigma}_{xy}^\top \boldsymbol{\Sigma}_{xx}^{-1}\boldsymbol{\sigma}_{xy}.$$

The relationship between the model parameters $\boldsymbol{\theta}_1 := (\alpha_1, \boldsymbol{\beta}_0, \boldsymbol{\beta}_1, \sigma_\varepsilon^2)$ of the (unrestricted) DLRM model (5):

$$y_t = \alpha_1 y_{t-1} + \boldsymbol{\beta}_0^\top \mathbf{x}_t + \boldsymbol{\beta}_1^\top \mathbf{x}_{t-1} + \varepsilon_t,\ \varepsilon_t \backsim \mathsf{IID}(0, \boldsymbol{\sigma}_\varepsilon^2),\ t \in \mathbb{T},$$

and the primary parameters of the joint distribution (9) was derived in Spanos (1986). Next we derive this relationship for the DLRM under the CF restrictions.

Theorem 2. The statistical parameterization of the DLRM model parameters $(\alpha_1, \boldsymbol{\beta}_0, \boldsymbol{\beta}_1, \sigma_\varepsilon^2)$ implied by the restricted $\boldsymbol{\Sigma}^*$ in (10) is:
(14)

$$\begin{pmatrix} \alpha_1 \\ \boldsymbol{\beta}_0 \\ \boldsymbol{\beta}_1 \end{pmatrix} = \begin{pmatrix} \boldsymbol{\beta}^\top \boldsymbol{\Sigma}_{22}(0)\boldsymbol{\beta} + \frac{\sigma^2}{1-\rho^2} & \boldsymbol{\beta}^\top \Sigma_{22}(1) & \boldsymbol{\beta}^\top \boldsymbol{\Sigma}_{22}(0) \\ \boldsymbol{\Sigma}_{22}(1)^\top \boldsymbol{\beta} & \boldsymbol{\Sigma}_{22}(0) & \boldsymbol{\Sigma}_{22}(1) \\ \boldsymbol{\Sigma}_{22}(0)\boldsymbol{\beta} & \boldsymbol{\Sigma}_{22}(1) & \boldsymbol{\Sigma}_{22}(0) \end{pmatrix}^{-1} \begin{pmatrix} \boldsymbol{\beta}^\top \boldsymbol{\Sigma}_{22}(1)\boldsymbol{\beta} + \frac{\rho\sigma^2}{1-\rho^2} \\ \boldsymbol{\Sigma}_{22}(0)\boldsymbol{\beta} \\ \boldsymbol{\Sigma}_{22}(1)\boldsymbol{\beta} \end{pmatrix}$$

$$= \begin{pmatrix} \rho \\ \boldsymbol{\Delta}(\boldsymbol{\Sigma}_{22}(0) - \boldsymbol{\Sigma}_{22}(1)\boldsymbol{\Sigma}_{22}(0)^{-1}\boldsymbol{\Sigma}_{22}(1))\boldsymbol{\beta} \\ -\boldsymbol{\rho}\boldsymbol{\beta} + (\boldsymbol{\Delta}\boldsymbol{\Sigma}_{22}(1) - \boldsymbol{\Sigma}_{22}(0)^{-1}\boldsymbol{\Sigma}_{22}(1)\boldsymbol{\Delta}\boldsymbol{\Sigma}_{22}(0))\boldsymbol{\beta} \end{pmatrix} = \begin{pmatrix} \rho \\ \boldsymbol{\beta} \\ -\rho\boldsymbol{\beta} \end{pmatrix}$$

$$\boldsymbol{\sigma}_\varepsilon^2 = \boldsymbol{\beta}^\top \boldsymbol{\Sigma}_{22}(0)\boldsymbol{\beta} + \frac{\sigma^2}{1-\rho^2} - \begin{pmatrix} \boldsymbol{\beta}^\top \boldsymbol{\Sigma}_{22}(1)\boldsymbol{\beta} + \frac{\rho\sigma^2}{1-\rho^2} & \boldsymbol{\beta}^\top \boldsymbol{\Sigma}_{22}(0) & \boldsymbol{\beta}^\top \boldsymbol{\Sigma}_{22}(1) \end{pmatrix} \begin{pmatrix} \alpha_1 \\ \boldsymbol{\beta}_0 \\ \boldsymbol{\beta}_1 \end{pmatrix} = \sigma^2.$$

giving rise to the ACLRM, i.e. a DLRM where the common factor (CF) restrictions (6) hold. See Appendix A for the details.

Taken together, Theorems 1 and 2 indicate that (10) constitutes a set of necessary and sufficient conditions for the common factor restrictions to hold. However, the restrictive nature of these CF conditions is not completely evident from (10). To shed light on the restrictiveness of these conditions in terms of the probabilistic structure of the vector process $\{\mathbf{Z}_t,\ t \in \mathbb{T}\}$, we use $\boldsymbol{\Sigma}^*$ in (10) and Lemma 1 to derive the Vector Autoregressive (VAR) model based on $D(\mathbf{Z}_t|\mathbf{Z}_{t-1}; \boldsymbol{\psi}_1^*)$.

Theorem 3. The implicit statistical parameterization of the VAR model, based on $D(\mathbf{Z}_t|\mathbf{Z}_{t-1}; \boldsymbol{\psi}_1^*)$:

$$\mathbf{Z}_t = \mathbf{A}^\top \mathbf{Z}_{t-1} + \mathbf{E}_t, \quad \mathbf{E}_t \backsim \mathsf{IID}(0, \boldsymbol{\Omega}),\ t \in \mathbb{T}. \tag{15}$$

implied by the restricted variance-covariance matrix $\boldsymbol{\Sigma}^*$ in (10), takes the form:

$$\mathbf{A}^\top = \begin{pmatrix} \rho & (\mathbf{D} - \boldsymbol{\rho}\mathbf{I}_k)\boldsymbol{\beta} \\ \mathbf{0} & \mathbf{D} \end{pmatrix}, \quad \boldsymbol{\Omega} = \begin{pmatrix} \sigma^2 + \boldsymbol{\beta}^\top \boldsymbol{\Lambda}\boldsymbol{\beta} & \boldsymbol{\beta}^\top \boldsymbol{\Lambda} \\ \boldsymbol{\Lambda}\boldsymbol{\beta} & \boldsymbol{\Lambda} \end{pmatrix}$$

where:

$$\mathbf{D} = \Sigma_{22}(0)^{-1}\Sigma_{22}(1) \quad \boldsymbol{\Lambda} = \boldsymbol{\Sigma}_{22}(0) - \Sigma_{22}(1)\Sigma_{22}(0)^{-1}\Sigma_{22}(1)$$

and $\mathbf{I}_k$ is a $k \times k$ identity matrix. See Appendix A for the proof.

Theorem 3 sheds ample light on just how unappetizing the necessary and sufficient conditions for an ACLRM are in terms of the implied restrictions on the vector stochastic process $\{\mathbf{Z}_t,\ t \in \mathbb{T}\}$:

(a) y_t does **not Granger cause** any of the regressors in $\mathbf{X}_t$, and

(b) $Cov(\mathbf{X}_t, y_t|\mathbf{Z}_{t-1}) = Cov(\mathbf{X}_t|\mathbf{X}_{t-1})Cov(\mathbf{X}_t)^{-1}Cov(\mathbf{X}_t, y_t) = \boldsymbol{\Lambda}\boldsymbol{\beta}$.

In addition, these results suggest that one can test the appropriateness of the CF restrictions (6) by testing the implied Granger non-causality in the context of the unrestricted VAR (15). The performance of this new test is considered in the Monte Carlo experiments presented below.

3 Monte Carlo Simulations

In an attempt to illustrate the restrictive nature of 'correcting for serial correlation' by modeling the error, even in the unlikely event that the Common Factor (CF) restrictions do hold, we consider a number of Monte Carlo experiments. All experimental results reported are based on 10,000 replications of sample sizes $T = 25$, $T = 50$ and $T = 100$. The experiments and results below use the following notation for the model parameters of the DLRM:

$$y_t = \alpha_0 + \beta_1 x_{1t} + \beta_2 x_{2t} + \alpha_1 y_{t-1} + \beta_3 x_{1t-1} + \beta_4 x_{2t-1} + \varepsilon_t,\ t \in \mathbb{T}. \tag{16}$$

When the CF restrictions hold this model is rewritten as:

$$y_t = \alpha_0^* + \beta_1 x_{1t} + \beta_2 x_{2t} + u_t; \quad u_t = \rho u_{t-1} + \varepsilon_t,\ t \in \mathbb{T}, \tag{17}$$

where $\rho = \alpha_1$, and $\alpha_0^* = \frac{\alpha_0}{1-\rho}$.

3.1 Experiment 1 - Unrestricted DLRM vs. ACLRM ($\rho = .588$)

In **Experiment 1** we generate data with two very similar implied Dynamic Linear regression Models (**1A-1B**). They differ by the model parameters, (β_3, β_4), the coefficients of (x_{1t-1}, x_{2t-1}) and, therefore, also by the intercept α_0. All other model parameters are the same.[2]

Experiment 1A - Primary parameters underlying DLRM

$$\begin{array}{cccc}
E(Y_t) = 2, & Var(Y_t) = 1.115, & Cov(Y_t, X_{1t}) = -.269, & Cov(Y_t, X_{2t}) = 0.5, \\
E(X_{1t}) = 1, & Var(X_{1t}) = 1, & Cov(X_{1t}, X_{1t-1}) = 0.6, & Cov(Y_t, Y_{t-1}) = 0.446, \\
E(X_{2t}) = .5, & Var(X_{2t}) = 1, & Cov(X_{2t}, X_{2t-1}) = 0.54, & Cov(Y_t, X_{1t-1}) = -.678, \\
 & Cov(X_{1t}, X_{2t}) = -.4, & Cov(X_{1t}, X_{2t-1}) = -.32, & Cov(Y_t, X_{2t-1}) = .42,
\end{array}$$

giving rise to an (unrestricted) DLRM:

$$\begin{gathered} y_t = 0.946 + 0.749 x_{1t} + 0.215 x_{2t} + 0.589 y_{t-1} - 0.936 x_{1t-1} - 0.089 x_{2t-1} + \varepsilon_t, \\ \sigma_\varepsilon^2 = 0.349, \quad \Re^2 = 0.687. \end{gathered} \tag{18}$$

The implied VAR parameters are:

(19)

$$\mathbf{a}_0 = \begin{pmatrix} 2.102 \\ 1.519 \\ 0.081 \end{pmatrix}, \ \mathbf{A}^\top = \begin{pmatrix} 0.236 & -0.596 & 0.045 \\ -0.535 & 0.498 & 0.105 \\ 0.222 & -0.156 & 0.262 \end{pmatrix}, \ \boldsymbol{\Omega} = \begin{pmatrix} 0.586 & 0.263 & 0.186 \\ 0.263 & 0.372 & -0.073 \\ 0.186 & -.073 & 1.116 \end{pmatrix}.$$ [3]

Consider the ACLRM experimental set-up:

Experiment 1B - Primary parameters underlying ACLRM

$$\begin{array}{cccc}
E(Y_t) = 2, & Var(Y_t) = 1.032, & Cov(Y_t, X_{1t}) = .663, & Cov(Y_t, X_{2t}) = 0.001, \\
E(X_{1t}) = 1, & Var(X_{1t}) = 1, & Cov(X_{1t}, X_{1t-1}) = 0.6, & Cov(Y_t, Y_{t-1}) = 0.574, \\
E(X_{2t}) = .5, & Var(X_{2t}) = 1.4, & Cov(X_{2t}, X_{2t-1}) = 0.54, & Cov(Y_t, X_{1t-1}) = .381, \\
 & Cov(X_{1t}, X_{2t}) = -.4, & Cov(X_{1t}, X_{2t-1}) = -.32, & Cov(Y_t, X_{2t-1}) = -.124,
\end{array}$$

[2] Experiments 1A and 1B not only have several model parameters in common, but they are also similar in the sense that in both cases, $\det(\boldsymbol{\Sigma}) = 0.163$.

[3] The data in these experiments are all generated using the implied VAR. The initial values (y_0, x_{10}, x_{20}) for each run are drawn randomly from the joint distribution of (y_0, x_{10}, x_{20}) implied by the experimental setup.

giving rise to the ACLRM:

$$
(20)\qquad \begin{aligned} y_t &= 1.143 + 0.749x_{1t} + 0.251x_{2t} + u, \qquad\qquad u_t = 0.589u_{t-1} + \varepsilon_t, \\ &= 0.470 + 0.749x_{1t} + 0.215x_{2t} + 0.589y_{t-1} - 0.441x_{1t-1} - 0.127x_{2t-1} + \varepsilon_t, \\ &\qquad\qquad \sigma_\varepsilon^2 = 0.349;\ \Re^2 = 0.662. \end{aligned}
$$

That is, for experiment 1B, the *common factor restrictions* implicitly imposed by an error AR(1) process hold. The implied VAR model parameters for this case are:

(21)

$$
\mathbf{a}_0 = \begin{pmatrix} 0.925 \\ 0.458 \\ 0.521 \end{pmatrix},\ \mathbf{A}^\top = \begin{pmatrix} 0.589 & -0.051 & -0.104 \\ 0 & 0.574 & -0.0645 \\ 0 & -0.187 & 0.332 \end{pmatrix},\ \mathbf{\Omega} = \begin{pmatrix} 0.701 & 0.437 & 0.114 \\ 0.437 & 0.635 & -0.181 \\ 0.114 & -0.181 & 1.161 \end{pmatrix}.
$$

We begin our simulation, as a typical modeler would and estimate the LRM:

$$
y_t = \alpha_0 + \beta_1 x_{1t} + \beta_2 x_{2t} + u_t,\ t \in \mathbb{T},
$$

We test for first-order autocorrelation using the Durbin-Watson (D-W) and related tests based on the following auxiliary regressions:

$$
\begin{array}{lll} \text{(i) } \hat{u}_t = \rho_1 \hat{u}_{t-1} + v_{1t}, & \text{(ii) } \hat{u}_t = \boldsymbol{\gamma}_0^\top \mathbf{x}_t + \gamma_1 \hat{u}_{t-1} + v_{2t}, & \text{(iii) } \hat{u}_t = \boldsymbol{\delta}_0^\top \mathbf{x}_t + \delta_1 y_{t-1} + \boldsymbol{\delta}_2^\top \mathbf{x}_{t-1} + v_{3t}, \\ \quad H_0: \rho_1 = 0, & \quad H_0: \gamma_1 = 0, & \quad H_0: \delta_1 = 0,\ \boldsymbol{\delta}_2 = 0. \end{array}
$$

The test based on (i) corresponds to the D-W test, the test based on (ii) was proposed by Breusch (1978) and Godfrey (1978), and the test based on (iii) by Spanos (1986). The last auxiliary regression arises when comparing the original regression with one based on replacing the Independence assumption with Markov dependence leading to:

$$
\begin{aligned} y_t &= \gamma_0 + \boldsymbol{\gamma}^\top \mathbf{x}_t + u_t, \\ y_t &= \alpha_0 + \boldsymbol{\beta}^\top \mathbf{x}_t + \delta_1 y_{t-1} + \boldsymbol{\delta}_2^\top \mathbf{x}_{t-1} + v_{3t}, \\ \hline u_t &= (\alpha_0 - \gamma_0) + (\boldsymbol{\beta}^\top - \boldsymbol{\gamma}^\top)\mathbf{x}_t + \delta_1 y_{t-1} + \boldsymbol{\delta}_2^\top \mathbf{x}_{t-1} + v_{3t}, \end{aligned}
$$

where the last auxiliary regression is equivalent to:

$$
(22)\qquad \widehat{u}_t = (\alpha_0 - \hat{\gamma}_0) + (\boldsymbol{\beta}^\top - \hat{\boldsymbol{\gamma}}^\top)\mathbf{x}_t + \delta_1 y_{t-1} + \boldsymbol{\delta}_2^\top \mathbf{x}_{t-1} + v_{3t}.
$$

These four autocorrelation misspecification tests cover the range of autocorrelation tests usually implemented in applied work. The LRM simulation results for Experiments 1A (DLRM) and 1B (ACLRM) are reported in Tables 1A-1B.

The results in table 1A suggest most clearly that the OLS estimator is biased and inconsistent when the CF restrictions do not hold; the true model is not the ACLRM, but a DLRM. In addition, any form of inference based on the estimated LRM is likely to be unreliable; both the inconsistency and the unreliability are accentuated as $T \to \infty$. The t-test significance results reported are indicative of the extent of the unreliability; the actual type I error rate is as large as 0.988 instead of the nominal 0.05. It is also interesting to note that the use of Heteroskedasticity and Autocorrelation Consistent (HAC) standard errors with $m = 2$ (see Newey and West, 1987), does not improve the reliability of inference – see the results in square brackets. Looking at the misspecification tests for temporal dependence in the DLRM case, we observe that the test based on the most general auxiliary regression (22) detects temporal dependence problems a large percentage of the time, even for small T, and it substantially out-performs all the other tests.

These DLRM results are in marked contrast to those of table 1B, where the CF restrictions do hold. When the CF restrictions hold, the OLS estimators of the regression coefficients seem both unbiased and consistent, even though the estimates of σ^2 and $\Re^2$ are still problematic. Notice also that the difference between the actual type I errors and the nominal significance level is not as large as it was in table 1A, though there is still some indication of unreliable inferences. Also the use of the HAC standard errors does not significantly improve reliability. As expected, the D-W test and auxiliary regression-based tests utilizing $\hat{u}_{t-1}$, (i)-(ii), do better than the test based on y_{t-1} and $\mathbf{X}_{t-1}$, (iii), but the difference in effectiveness disappears with increasing sample size. Given the unrealistic nature of the CF restrictions and the dominance of (iii) in the DLRM case, sole reliance on the D-W test is not recommended.

Suppose that on the basis of the misspecification tests in 1A and 1B, a modeler decides to 'correct' the apparent autocorrelation problem by adopting an ACLRM model. Tables 2A-2B present simulation results for this ACLRM estimated using an iterative Cochrane-Orcutt correction (GLS) estimation procedure; the results based on other iterative procedures are very similar. In table 2A the true model is a DLRM and, as we can see, the ACLRM estimators are both biased and inconsistent, giving rise to unreliable inferences; the unreliability is as bad as estimating the LRM. Not surprisingly, when the true model is the ACLRM, the ACLRM estimation results (table 2B) are quite reliable, particularly for larger T.

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

		T=25		T=50		T=100	
	True	Mean	Std	Mean	Std	Mean	Std
$\hat{\alpha}_0$	0.946	1.515	0.834	1.693	0.580	1.829	0.389
$\hat{\beta}_1$	0.749	0.346	0.358	0.157	0.289	0.020	0.219
$\hat{\beta}_2$	0.215	0.264	0.179	0.286	0.129	0.300	0.093
$\hat{\sigma}^2$	0.349	0.631	0.257	0.742	0.215	0.824	0.166
$\Re^2$	0.687	0.271	0.165	0.201	0.115	0.176	0.089
t-statistics		Mean	% R(.05)	Mean	% R(.05)	Mean	% R(.05)
$\tau_{\alpha_0} = \frac{\hat{\alpha}_0 - \alpha_0}{\hat{\sigma}_{\alpha_0}}$		2.460 [2.749]	.577 [.609]	3.874 [4.003]	.731 [730]	5.996 [5.938]	0.907 [.898]
$\tau_{\beta_1} = \frac{\hat{\beta}_1 - \beta_1}{\hat{\sigma}_{\beta_1}}$		-1.857 [-2.197]	0.446 [.495]	-3.962 [-4.172]	0.823 [.815]	-7.015 [-6.895]	0.988 [.984]
$\tau_{\beta_2} = \frac{\hat{\beta}_2 - \beta_2}{\hat{\sigma}_{\beta_2}}$		0.318 [.378]	0.102 [.164]	0.625 [.677]	0.122 [.159]	1.025 [1.060]	0.198 [.214]
M-S (F-test)		Statistic	% R(.05)	Statistic	% R(.05)	Statistic	% R(.05)
D-W		1.518	0.272	1.590	0.398	1.685	0.413
(i) $\hat{u}_{t-1}$		1.025	0.157	1.346	0.282	1.509	0.341
(ii) $\hat{u}_{t-1}, \mathbf{x}_t$		2.718	0.209	4.225	0.371	5.290	0.441
(iii) $y_{t-1}, \mathbf{x}_{t-1}, \mathbf{x}_t$		8.257	0.829	19.922	0.998	46.255	1.00

*'% R(.05)' denotes the percentage of actual rejections at .05 significance level and numbers in square brackets refer to the t-test using the Newey West HAC standard errors with m=2.

Table 1B- True: ACLRM // Estimated: LRM (OLS)*

	True	T=25 Mean	T=25 Std	T=50 Mean	T=50 Std	T=100 Mean	T=100 Std
$\hat{\alpha}_0^*$	1.143	1.138	0.383	1.137	0.269	1.142	0.190
$\hat{\beta}_1$	0.749	0.752	0.220	0.751	0.154	0.750	0.109
$\hat{\beta}_2$	0.215	0.214	0.159	0.216	0.111	0.216	0.079
$\hat{\sigma}^2$	0.349	0.451	0.184	0.488	0.142	0.512	0.105
$\Re^2$	0.662	0.527	0.173	0.506	0.131	0.493	0.097
t-statistics		Mean	% R(.05)	Mean	% R(.05)	Mean	% R(.05)
$\tau_{\alpha_0^*} = \frac{\hat{\alpha}_0^* - \alpha_0^*}{\hat{\sigma}_{\alpha_0^*}}$		0.018 [.015]	0.226 [.260]	-.036 [-.027]	0.220 [.207]	-.012 [-.008]	0.217 [.178]
$\tau_{\beta_1} = \frac{\hat{\beta}_1 - \beta_1}{\hat{\sigma}_{\beta_1}}$		0.011 [.011]	0.151 [.205]	0.019 [.017]	0.152 [.159]	0.005 [.002]	0.151 [.131]
$\tau_{\beta_2} = \frac{\hat{\beta}_2 - \beta_2}{\hat{\sigma}_{\beta_2}}$		-0.010 [-.016]	0.106 [.167]	0.006 [.004]	0.101 [.122]	0.014 [.014]	0.101 [.100]
M-S Tests		Statistic	% R(.05)	Statistic	% R(.05)	Statistic	% R(.05)
D-W		1.174	0.652	1.004	0.967	0.912	1.0
(i) $\hat{u}_{t-1}$		2.130	0.535	3.991	0.938	6.447	0.999
(ii) $\hat{u}_{t-1}, \mathbf{x}_t$		6.318	0.504	17.997	0.935	43.86	1.0
(iii) $y_{t-1}, \mathbf{x}_{t-1}, \mathbf{x}_t$		3.637	0.451	7.548	0.885	16.275	0.998

*'% R(.05)' denotes the percentage of actual rejections at .05 significance level and numbers in square brackets refer to the t-test using the Newey West HAC standard errors with m=2.

Table 2A- True: DLRM // Estimated: ACLRM*

	True	T=25 Mean	T=25 Std	T=50 Mean	T=50 Std	T=100 Mean	T=100 Std
$\hat{\alpha}_0$	0.946	1.045	1.352	1.025	1.300	1.017	1.236
$\hat{\beta}_1$	0.749	0.837	0.458	0.864	0.451	0.873	0.454
$\hat{\beta}_2$	0.215	0.215	0.143	0.213	0.097	0.212	0.074
$\hat{\rho}$		0.586	0.421	0.705	0.365	0.755	0.345
$\hat{\sigma}^2$	0.349	0.414	0.167	0.452	0.175	0.473	0.182
t-statistics		Mean	% R(.05)	Mean	% R(.05)	Mean	% R(.05)
$\tau_{\alpha_0} = \frac{\hat{\alpha}_0 - \alpha_0}{\hat{\sigma}_{\alpha_0}}$		0.826	0.366	0.767	0.318	0.9603	0.270
$\tau_{\beta_1} = \frac{\hat{\beta}_1 - \beta_1}{\hat{\sigma}_{\beta_1}}$		0.622	0.453	1.143	0.682	1.785	0.917
$\tau_{\beta_2} = \frac{\hat{\beta}_2 - \beta_2}{\hat{\sigma}_{\beta_2}}$		-0.034	0.098	-0.091	0.082	-0.148	0.095
$\tau_{\rho} = \frac{\hat{\rho}}{\hat{\sigma}_{\rho}}$		6.103	0.839	11.774	0.926	19.218	0.934
CF Test		Statist.	% R(.05)	Statistic	% R(.05)	Statistic	% R(.05)
D-M F		10.289	0.764 (0.381)	19.204	0.945 (0.753)	39.11	0.997 (0.980)
LR		3.719	0.169 (0.381)	6.592	0.389 (0.753)	12.853	0.812 (0.980)
G-C VAR		15.184	0.787 (0.595)	30.982	0.987 (0.974)	66.36	1.0 (1.0)

*'% R(.05)' denotes the percentage of actual rejections at .05 significance level and other numbers in parentheses refer to size-corrected power.

Tables 2A-2B also report the results a modeler would likely obtain if the common factor restrictions (CF), implicitly imposed by the ACLRM, are tested. The first test (D-M F) is the approximate F-test recommended in Davidson and MacKinnon (1993) and the second is the traditional Likelihood ratio (LR) test (see Spanos, 1986). The last test (G-C VAR) is an F-type test of the assumption that y_t does not Granger Cause $\mathbf{X}_t$ in the context of (15). In conducting this test, the VAR is estimated using Iterative Seemingly Unrelated Regression (SUR).

The test results in table 2B (where the CF restrictions hold) indicate that the actual sizes of the D-M F and the LR tests differ from the nominal systematically; the former overestimates and the latter underestimates the nominal size. Although the actual size of the G-C VAR F-type test is high for small T, it approaches the nominal size as T increases.

Given these size problems, table 2A reports both the actual and size corrected percent rejected for the various CF tests. The results indicate that while all three of the CF tests have reasonably good size adjusted power, the proposed Granger non-causality test is the most probative for all sample sizes, including small T.

Table 2B- True: ACLRM // Estimated: ACLRM*

		T=25		T=50		T=100	
	True	Mean	Std	Mean	Std	Mean	Std
$\hat{\alpha}_0^*$	1.143	1.138	0.372	1.137	0.244	1.141	0.169
$\hat{\beta}_1$	0.749	0.751	0.178	0.752	0.114	0.750	0.078
$\hat{\beta}_2$	0.215	0.211	0.126	0.216	0.081	0.216	0.055
$\hat{\rho}$	0.589	0.456	0.224	0.528	0.133	0.561	0.088
$\hat{\sigma}^2$	0.349	0.331	0.105	0.341	0.073	0.346	0.050
t-statistics		Mean	% R(.05)	Mean	% R(.05)	Mean	% R(.05)
$\tau_{\alpha_0} = \frac{\hat{\alpha}_0 - \alpha_0^*}{\hat{\sigma}_{\alpha_0}}$		-0.015	0.116	-0.027	0.083	-0.012	0.064
$\tau_{\beta_1} = \frac{\hat{\beta}_1 - \beta_1}{\hat{\sigma}_{\beta_1}}$		0.017	0.093	0.025	0.061	0.009	0.054
$\tau_{\beta_2} = \frac{\hat{\beta}_2 - \beta_2}{\hat{\sigma}_{\beta_2}}$		-0.001	0.081	0.006	0.059	0.016	0.052
$\tau_{\rho} = \frac{\hat{\rho} - \rho}{\hat{\sigma}_{\rho}}$		-0.567	0.132	-0.398	0.078	-0.264	0.063
CF Test		Statistic	% R(.05)	Statistic	% R(.05)	Statistic	% R(.05)
D-M F		3.111	0.320	2.454	0.265	2.252	0.253
LR		1.360	0.008	1.080	0.003	0.986	0.002
G-C VAR		3.222	0.151	2.523	0.096	2.219	0.071

*'% R(.05)' denotes the percentage of actual rejections at .05 significance level.

Tables 3A-3B summarize the implications of estimating the DLRM with and without the Common Factor restrictions. These tables indicate that estimation is generally accurate and the usual t-tests reliable (particularly for large T), whether or not the CF restrictions hold. That is, even in the ACLRM case, estimating the DLRM would give rise to very reliable inferences. In fact, a comparison of the GLS results with these unrestricted DLRM results for the ACLRM case suggests no advantage to using GLS even when the restrictions hold, except maybe for very small T. Given the unrealistic nature of the common factor restrictions, and the potential for biased and inconsistent estimators when the restrictions do not hold, the adoption of the ACLRM is not recommended.

Table 3A- True: DLRM // Estimated: DLRM (OLS)							
		T=25		T=50		T=100	
	True	Mean	Std	Mean	Std	Mean	Std
$\hat{\alpha}_0$	0.946	1.026	0.772	0.987	0.361	0.970	0.233
$\hat{\beta}_1$	0.749	0.772	0.228	0.758	0.149	0.752	0.102
$\hat{\beta}_2$	0.215	0.212	0.134	0.215	0.086	0.216	0.058
$\hat{\alpha}_1$	0.589	0.458	0.206	0.522	0.131	0.556	0.087
$\hat{\beta}_3$	-0.936	-0.788	0.229	-0.861	0.137	-0.900	0.089
$\hat{\beta}_4$	-0.09	-0.068	0.140	-0.078	0.091	-0.084	0.062
$\hat{\sigma}^2$	0.349	0.339	0.111	0.345	0.074	0.348	0.051
$\Re^2$	0.687	0.642	0.120	0.637	0.095	0.653	0.076
t-statistics		Mean	% R(.05)	Mean	% R(.05)	Mean	% R(.05)
$\tau_{\alpha_0} = \frac{\hat{\alpha}_0 - \alpha_0}{\hat{\sigma}_{\alpha_0}}$		0.137	0.087	0.109	0.067	0.096	0.056
$\tau_{\beta_1} = \frac{\hat{\beta}_1 - \beta_1}{\hat{\sigma}_{\beta_1}}$		0.104	0.063	0.062	0.053	0.023	0.054
$\tau_{\beta_2} = \frac{\hat{\beta}_2 - \beta_2}{\hat{\sigma}_{\beta_2}}$		-0.023	0.066	0.004	0.057	0.015	0.048
$\tau_{\alpha_1} = \frac{\hat{\alpha}_1 - \alpha_1}{\hat{\sigma}_{\alpha_1}}$		-0.633	0.09	-0.498	0.079	-0.369	0.065
$\tau_{\beta_3} = \frac{\hat{\beta}_3 - \beta_3}{\hat{\sigma}_{\beta_3}}$		0.664	0.111	0.541	0.089	0.395	0.069
$\tau_{\beta_4} = \frac{\hat{\beta}_4 - \beta_4}{\hat{\sigma}_{\beta_4}}$		0.175	0.064	0.134	0.055	0.096	0.054

Table 3B- True: ACLRM // Estimated: DLRM (OLS)							
		T=25		T=50		T=100	
	True	Mean	Std	Mean	Std	Mean	Std
$\hat{\alpha}_0$	0.470	0.648	0.399	0.557	0.243	0.512	0.158
$\hat{\beta}_1$	0.749	0.752	0.181	0.751	0.116	0.750	0.080
$\hat{\beta}_2$	0.215	0.215	0.135	0.216	0.087	0.216	0.058
$\hat{\alpha}_1$	0.589	0.432	0.210	0.511	0.131	0.551	0.088
$\hat{\beta}_3$	-0.441	-0.326	0.239	-0.384	0.152	-0.413	0.103
$\hat{\beta}_4$	-0.127	-0.095	0.140	-0.110	0.090	-0.119	0.061
$\hat{\sigma}^2$	0.349	0.343	0.113	0.347	0.075	0.349	0.051
$\Re^2$	0.662	0.681	0.122	0.665	0.091	0.662	0.066
t-statistics		Mean	% R(.05)	Mean	% R(.05)	Mean	% R(.05)
$\tau_{\alpha_0} = \frac{\hat{\alpha}_0 - \alpha_0}{\hat{\sigma}_{\alpha_0}}$		0.427	0.091	0.315	0.072	0.222	0.059
$\tau_{\beta_1} = \frac{\hat{\beta}_1 - \beta_1}{\hat{\sigma}_{\beta_1}}$		0.013	0.068	0.016	0.054	0.006	0.053
$\tau_{\beta_2} = \frac{\hat{\beta}_2 - \beta_2}{\hat{\sigma}_{\beta_2}}$		-0.003	0.066	0.012	0.060	0.019	0.050
$\tau_{\alpha_1} = \frac{\hat{\alpha}_1 - \alpha_1}{\hat{\sigma}_{\alpha_1}}$		-0.698	0.107	-0.532	0.075	-0.377	0.064
$\tau_{\beta_3} = \frac{\hat{\beta}_3 - \beta_3}{\hat{\sigma}_{\beta_3}}$		0.491	0.090	0.367	0.072	0.266	0.063
$\tau_{\beta_4} = \frac{\hat{\beta}_4 - \beta_4}{\hat{\sigma}_{\beta_4}}$		0.245	0.071	0.186	0.055	0.126	0.050

3.2 Experiment 2 - Unrestricted DLRM vs. ACLRM ($\rho = .2$)

The main purpose of **Experiment 2** is to examine the extent to which the previous experimental results change when the parameter on y_{t-1} (the degree of autocorrelation in the ACLRM) is reduced. To reduce this model parameter, we made only two changes to the moments of the joint distribution (primary parameters) of Experiment 1A. First, the $Cov(y_t, y_{t-1})$ is reduced from 0.446 to 0.185. This decreased the coefficient of y_{t-1} (to .20 from .60), but also reduced the overall fit of the DLRM. In order to maintain a model fit close to that in Experiment 1A, we also reduced $Var(y_t)$ from 1.115 to 0.85.

Experiment 2B differs from Experiment 2A, in the same way as Experiment 1B differs from 1A. That is, the two DLRMs differ by the model parameters, (β_3, β_4), the coefficients of (x_{1t-1}, x_{2t-1}), and, therefore, also by the intercept α_0. All other model parameters remain the same.[4]

Experiment 2A - Primary parameters underlying DLRM

$$
\begin{array}{cccc}
E(Y_t) = 2, & Var(Y_t) = .85, & Cov(Y_t, X_{1t}) = -.269, & Cov(Y_t, X_{2t}) = 0.5, \\
E(X_{1t}) = 1, & Var(X_{1t}) = 1, & Cov(X_{1t}, X_{1t-1}) = 0.6, & Cov(Y_t, Y_{t-1}) = 0.185, \\
E(X_{2t}) = .5, & Var(X_{2t}) = 1, & Cov(X_{2t}, X_{2t-1}) = 0.54, & Cov(Y_t, X_{1t-1}) = -.678, \\
 & Cov(X_{1t}, X_{2t}) = -.4, & Cov(X_{1t}, X_{2t-1}) = -.32, & Cov(Y_t, X_{2t-1}) = .42,
\end{array}
$$

giving rise to the DLRM:

$$
\begin{gathered}
y_t = 1.842 + 0.455x_{1t} + 0.237x_{2t} + 0.200y_{t-1} - 0.818x_{1t-1} + 0.0076x_{2t-1} + \varepsilon_t, \\
\sigma_\varepsilon^2 = 0.259, \ \Re^2 = 0.695
\end{gathered} \tag{23}
$$

The implied VAR parameters are:

(24)

$$
\mathbf{a}_0 = \begin{pmatrix} 2.705 \\ 1.950 \\ -0.097 \end{pmatrix}, \ \mathbf{A}^\top = \begin{pmatrix} -0.068 & -0.640 & 0.142 \\ -0.752 & 0.46712 & 0.173 \\ 0.311 & -0.143 & 0.234 \end{pmatrix}, \boldsymbol{\Omega} = \begin{pmatrix} 0.369 & 0.114 & 0.247 \\ 0.114 & 0.265 & -0.028 \\ 0.247 & -0.028 & 1.097 \end{pmatrix}.
$$

For the ACLRM, the experimental set-up is as follows:

Experiment 2B - Primary parameters underlying ACLRM

$$
\begin{array}{cccc}
E(Y_t) = 2, & Var(Y_t) = .469, & Cov(Y_t, X_{1t}) = .360, & Cov(Y_t, X_{2t}) = 0.150, \\
E(X_{1t}) = 1, & Var(X_{1t}) = 1, & Cov(X_{1t}, X_{1t-1}) = 0.6, & Cov(Y_t, Y_{t-1}) = 0.139, \\
E(X_{2t}) = .5, & Var(X_{2t}) = 1.4, & Cov(X_{2t}, X_{2t-1}) = 0.54, & Cov(Y_t, X_{1t-1}) = .197, \\
 & Cov(X_{1t}, X_{2t}) = -.4, & Cov(X_{1t}, X_{2t-1}) = -.32, & Cov(Y_t, X_{2t-1}) = -.017,
\end{array}
$$

giving rise to the ACLRM:

$$
\begin{aligned}
y_t &= 1.427 + 0.455x_{1t} + 0.237x_{2t} + u_t, \qquad u_t = 0.200u_{t-1} + \varepsilon_t \\
&= 1.142 + 0.455x_{1t} + 0.237x_{2t} + 0.200y_{t-1} - 0.091x_{1t-1} + 0.047x_{2t-1} + \varepsilon_t, \\
&\qquad \sigma_\varepsilon^2 = 0.259; \ \Re^2 = 0.448,
\end{aligned} \tag{25}
$$

The implied VAR parameters for this case are:

(26)

$$
\mathbf{a}_0 = \begin{pmatrix} 1.473 \\ 0.458 \\ 0.521 \end{pmatrix}, \ \mathbf{A}^\top = \begin{pmatrix} 0.200 & 0.126 & 0.002 \\ 0 & 0.574 & -0.064 \\ 0 & -0.187 & 0.332 \end{pmatrix}, \ \boldsymbol{\Omega} = \begin{pmatrix} 0.416 & 0.246 & 0.193 \\ 0.246 & 0.635 & -0.181 \\ 0.193 & -0.181 & 1.161 \end{pmatrix}.
$$

[4] Experiments 2A and 2B are also similar in the sense that in both cases, $Det(\Sigma) = 0.061$.

The main conclusions from tables 4A and 4B are basically confirmed in tables 1A and 1B. Not surprisingly, the smaller autocorrelation coefficient reduces (slightly) both the bias and the inconsistency in the LRM estimators. However, inference based on the t-tests remains unreliable and the use of HAC standard errors does not seem to improve the reliability. The most significant difference in these new results is the reduced power of the misspecification tests. When the common factor restrictions hold, the power of all four misspecification tests is drastically reduced. However, in the case of the DLRM, there is only a significant drop in the power of the $\hat{u}_{t-1}$ based tests, including the D-W. Interestingly the auxiliary regression test based on y_{t-1} and $\mathbf{x}_{t-1}$ drastically outperforms the other tests for all T and its ability to pick up departures from independence is even higher than in Experiment 1A (table 1A). Another surprising observation regarding the misspecification test results in table 4A concerns the changes in power as T increases. While the power of all the $\hat{u}_{t-1}$based tests is very low for $T = 25$, we only see an increase in power with T for the auxiliary regression based tests; the power of the D-W test actually decreases with T. Taken altogether, this new evidence confirms the earlier observation that sole reliance on the usual D-W test is not recommended.

Tables 5A-5B report the simulation results obtained when an ACLRM is estimated under the two different scenarios: no common factors (DLRM) and common factors (ACLRM). The main findings of experiment 1 (tables 2A-2B) are simply corroborated in experiment 2. What is surprising is the increased effectiveness of all the CF tests in the DLRM case (experiment 2A). As in experiment 1, however, the G-C VAR-based F test clearly outperforms the other two tests in terms of both size and power.

Table 4A - True: DLRM // Estimated: LRM (OLS)*

		T=25		T=50		T=100	
	True	Mean	Std	Mean	Std	Mean	Std
$\hat{\alpha}_0$	1.842	1.601	0.682	1.755	0.457	1.864	0.298
$\hat{\beta}_1$	0.455	0.256	0.331	0.094	0.253	-0.017	0.182
$\hat{\beta}_2$	0.237	0.273	0.149	0.292	0.104	0.304	0.073
$\hat{\sigma}^2$	0.259	0.491	0.194	0.558	0.158	0.605	0.120
$\Re^2$	0.695	0.282	0.157	0.233	0.117	0.219	0.096
t-statistics		Mean	% R(.05)	Mean	% R(.05)	Mean	% R(.05)
$\tau_{\alpha_0} = \frac{\hat{\alpha}_0 - \alpha_0}{\hat{\sigma}_{\alpha_0}}$		-0.433 [-.466]	0.393 [.456]	-0.086 [.004]	0.364 [.399]	0.436 [.571]	0.364 [.384]
$\tau_{\beta_1} = \frac{\hat{\beta}_1 - \beta_1}{\hat{\sigma}_{\beta_1}}$		-1.095 [-1.346]	0.342 [.397]	-2.856 [-3.115]	0.652 [.644]	-5.356 [-5.446]	0.920 [.903]
$\tau_{\beta_2} = \frac{\hat{\beta}_2 - \beta_2}{\hat{\sigma}_{\beta_2}}$		0.265 [.330]	0.082 [.151]	0.558 [.629]	0.095 [.141]	0.935 1.011]	0.155 [.193]
M-S (F-test)		Statistic	% R(.05)	Statistic	% R(.05)	Statistic	% R(.05)
D-W		2.011	0.025	2.195	0.011	2.351	0.002
$\hat{u}_{t-1}$		-0.276	0.045	-0.884	0.171	-1.932	0.485
$\hat{u}_{t-1}, \mathbf{x}_t$		1.56	0.089	3.426	0.291	8.869	0.638
$y_{t-1}, \mathbf{x}_{t-1}, \mathbf{x}_t$		8.952	0.865	20.352	0.999	45.477	1.0

*'% R(.05)' denotes the percentage of actual rejections at .05 significance level and numbers in square brackets refer to the t-test using the Newey West HAC standard errors with m=2.

Table 4B- True: ACLRM // Estimated: LRM (OLS)*	True	T=25 Mean	T=25 Std	T=50 Mean	T=50 Std	T=100 Mean	T=100 Std
$\hat{\alpha}_0^*$	1.142	1.424	0.214	1.423	0.145	1.426	0.100
$\hat{\beta}_1$	0.455	0.457	0.140	0.457	0.093	0.455	0.064
$\hat{\beta}_2$	0.237	0.237	0.110	0.238	0.074	0.238	0.051
$\hat{\sigma}^2$	0.259	0.260	0.081	0.265	0.057	0.267	0.040
$\Re^2$	0.448	0.449	0.156	0.437	0.114	0.430	0.083
t-statistics		Mean	% R(.05)	Mean	% R(.05)	Mean	% R(.05)
$\tau_{\alpha_0^*}=\frac{\hat{\alpha}_0^*-\alpha_0^*}{\hat{\sigma}_{\alpha_0^*}}$		0.014 [-.074]	0.091 [.146]	-.030 [-.032]	0.091 [.111]	-.014 [-.012]	0.085 [.85]
$\tau_{\beta_1}=\frac{\hat{\beta}_1-\beta_1}{\hat{\sigma}_{\beta_1}}$		0.014 [.015]	0.084 [.044]	0.024 [.024]	0.081 [.107]	0.009 [.008]	0.079 [.085]
$\tau_{\beta_2}=\frac{\hat{\beta}_2-\beta_2}{\hat{\sigma}_{\beta_2}}$		-0.005 [.009]	0.076 [.140]	0.014 [.014]	0.070 [.100]	0.018 [.017]	0.065 [.075]
M-S Tests		Statistic	% R(.05)	Statistic	% R(.05)	Statistic	% R(.05)
D-W		1.759	0.110	1.686	0.259	1.644	0.483
$\hat{u}_{t-1}$		0.418	0.069	0.997	0.163	1.723	0.403
$\hat{u}_{t-1},\mathbf{x}_t$		1.190	0.056	2.020	0.153	4.033	0.402
$y_{t-1},\mathbf{x}_{t-1},\mathbf{x}_t$		1.304	0.077	1.460	0.121	2.107	0.278

*'% R(.05)' denotes the percentage of actual rejections at .05 significance level

Table 5A- True: DLRM // Estimated: ACLRM*	True	T=25 Mean	T=25 Std	T=50 Mean	T=50 Std	T=100 Mean	T=100 Std
$\hat{\alpha}_0$	1.842	1.607	0.975	1.877	0.721	2.124	0.399
$\hat{\beta}_1$	0.455	0.261	0.649	-0.017	0.583	-0.272	0.361
$\hat{\beta}_2$	0.237	0.258	0.136	0.280	0.093	0.298	0.061
$\hat{\rho}$		-0.051	0.611	-0.245	0.542	-0.455	0.335
$\hat{\sigma}^2$	0.259	0.366	0.130	0.428	0.108	0.473	0.080
t-statistics		Mean	% R(.05)	Mean	% R(.05)	Mean	% R(.05)
$\tau_{\alpha_0}=\frac{\hat{\alpha}_0-\alpha_0}{\hat{\sigma}_{\alpha_0}}$		0.243	0.502	1.593	0.648	3.921	0.863
$\tau_{\beta_1}=\frac{\hat{\beta}_1-\beta_1}{\hat{\sigma}_{\beta_1}}$		-1.616	0.871	-5.531	0.987	-12.233	1.0
$\tau_{\beta_2}=\frac{\hat{\beta}_2-\beta_2}{\hat{\sigma}_{\beta_2}}$		0.204	0.130	0.599	0.161	1.205	0.265
$\tau_{\rho}=\frac{\hat{\rho}}{\hat{\sigma}_{\rho}}$		0.272	0.817	-1.587	0.962	-5.257	0.998
CF Test		Statist.	% R(.05)	Statistic	% R(.05)	Statistic	% R(.05)
D-M: F		16.669	0.939 (0.722)	38.111	0.999 (0.993)	88.007	1.0 (1.0)
LR		5.646	0.401 (0.722)	12.367	0.920 (0.993)	27.516	1.0 (1.0)
G-C - VAR		29.452	0.952(0.896)	62.257	0.999 (0.999)	133.05	1.0 (1.0)

*'% R(.05)' denotes the percentage of actual rejections at .05 significance level and numbers in parentheses refer to size-corrected power.

Table 5B- True: ACLRM // Estimated: ACLRM*							
		T=25		T=50		T=100	
	True	Mean	Std	Mean	Std	Mean	Std
$\hat{\alpha}_0^*$	1.427	1.424	0.222	1.423	0.145	1.425	0.099
$\hat{\beta}_1$	0.455	0.457	0.147	0.457	0.093	0.455	0.063
$\hat{\beta}_2$	0.237	0.237	0.114	0.238	0.073	0.238	0.050
$\hat{\rho}$	0.200	0.099	0.229	0.151	0.151	0.176	0.103
$\hat{\sigma}^2$	0.259	0.245	0.078	0.253	0.054	0.256	0.038
t-statistics		Mean	% R(.05)	Mean	% R(.05)	Mean	% R(.05)
$\tau_{\alpha_0} = \frac{\hat{\alpha}_0 - \alpha_0^*}{\hat{\sigma}_{\alpha_0}}$		-0.0122	0.091	-0.029	0.071	-0.015	0.056
$\tau_{\beta_1} = \frac{\hat{\beta}_1 - \beta_1}{\hat{\sigma}_{\beta_1}}$		0.013	0.096	0.0234	0.071	0.011	0.060
$\tau_{\beta_2} = \frac{\hat{\beta}_2 - \beta_2}{\hat{\sigma}_{\beta_2}}$		-0.008	0.094	0.0115	0.069	0.017	0.056
$\tau_{\rho} = \frac{\hat{\rho} - \rho}{\hat{\sigma}_{\rho}}$		-0.454	0.108	-0.318	0.077	-0.219	0.063
CF Test		Statistic	% R(.05)	Statistic	% R(.05)	Statistic	% R(.05)
D-M: F-test		2.813	0.274	2.320	0.242	2.181	0.239
LR test		1.241	0.004	1.024	0.002	0.956	0.002
G-C - VAR		2.886	0.127	2.394	0.085	2.155	0.060

*'% R(.05)' denotes the percentage of actual rejections at .05 significance level.

3.3 Summary of Monte Carlo results

The above Monte Carlo results demonstrate most clearly that when the common factor (CF) restrictions do not hold:

(a) the Linear Regression Model (LRM) OLS estimators are both biased and inconsistent and inference based on them highly unreliable,

(b) the 'autocorrelation correction' GLS estimators, based on the Autocorrelation Corrected LRM (ACLRM), do not improve the situation; they gives rise to different but equally misleading inferences, and

(c) utilizing heteroskedastic and autocorrelation consistent standard errors does not ameliorate the reliability of inference.

When the common factor (CF) restrictions do hold, our Monte Carlo results demonstrate that:

(d) the Dynamic Linear Regression Model (DLRM) OLS estimators are equally reliable as the ACLRM GLS estimators with hardly any loss of efficiency, particularly for $T \geqq 50$; the same is true for the VAR estimators.

Taking (a)-(d) together, the main conclusion is that, in view of the misleading results from the ACLRM when the CF restrictions do not hold, the adoption of the ACLRM when residual autocorrelation is detected in practice is not a good idea.

The Monte Carlo experiments in this paper also call into question the usual practice of relying solely on the Durbin-Watson test (D-W) to assess the independence assumption. Not surprisingly, we find that the power of the D-W is much higher when the common factor restrictions do hold than when they do not. However, a more general test of autocorrelation is shown to perform almost as well as the D-W when the common factor restrictions do hold and *significantly* better than the D-W when the restrictions do not hold. Given that the CF restrictions may be unlikely for most data applications anyway, sole reliance on D-W test seems problematic. Our Monte Carlo results recommend the use of the test based on the auxiliary regression:

$$\widehat{u}_t = (\alpha_0 - \hat{\gamma}_0) + (\boldsymbol{\beta}^\top - \hat{\boldsymbol{\gamma}}^\top)\mathbf{x}_t + \delta_1 y_{t-1} + \boldsymbol{\delta}_2^\top \mathbf{x}_{t-1} + v_{3t}. \tag{27}$$

The most important advise for practitioners is that they should test the common factor restrictions before imposing them, irrespective of the test of autocorrelation used to detect residual autocorrelation. The Monte Carlo results suggest that the simple (approximate) F-test suggested by Davidson and MacKinnon (1993) as well as a Likelihood Ratio test suffer from size problems. A more probative test for the CF restrictions, based on the Granger non-causality of y_t on $\mathbf{X}_t$, is proposed.

As mentioned in the introduction, however, these CF restriction tests are based on the presumption that the DLRM is statistically adequate for data $(\mathbf{y}, \mathbf{X})$. This is the issue we address next.

4 'Autocorrelation Correction' as a modeling strategy

The question of whether or not autocorrelation correcting is appropriate 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 hypotheses:

$$H_0 : \beta_1 = 0 \text{ vs. } H_1 : \beta_1 \neq 0. \tag{28}$$

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 a t-test (28) and the D-W test based on:

$$H_0 : \rho = 0 \text{ vs. } H_1 : \rho \neq 0, \tag{29}$$

in terms of the hypotheses being tested, they are very different in nature. As argued in Spanos (1999), the D-W test is a misspecification test, but the t-test 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 (see table A):

Table A - The Linear Regression Model (LRM)

$$\boxed{y_t = \beta_0 + \boldsymbol{\beta}_1^\top \mathbf{x}_t + u_t,\ t \in \mathbb{T},}$$

[1]	**Normality:**	$D(y_t \mid \mathbf{x}_t; \boldsymbol{\theta})$, is Normal
[2]	**Linearity:**	$E(y_t \mid \mathbf{x}_t) = \beta_0 + \boldsymbol{\beta}_1^\top \mathbf{x}_t$,
[3]	**Homoskedasticity:**	$Var(y_t \mid \mathbf{x}_t) = \sigma^2$, free of $\mathbf{x}_t$,
[4]	**t-homogeneity:**	$\boldsymbol{\theta} :=(\beta_0, \boldsymbol{\beta}_1, \sigma^2)$ are t-invariant $\forall t \in \mathbb{T}$,
[5]	**Independence:**	$\{(y_t \mid \mathbf{x}_{t-1}),\ t \in \mathbb{T}\}$ - independent process.

The t-test, on the other hand, 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) is valid. This ensures that the estimated model contains the 'true' model. Hence, rejection of the null leaves only one choice–the alternative–as the union of the null and the alternative 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 alternative model one needs to test its own assumptions. More formally, the alternative hypothesis in (29) *has not passed a severe test*; see Mayo (1996), Mayo and Spanos (2004).

In the case of the D-W test, if the null is rejected one can only infer that the LRM is misspecified since the data exhibit some kind of dependence over the dimension of the index $t = 1, 2, ..., T$. 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.[5] The alternative model involved in a D-W test (4) is only one of an infinite number of potential models that could have given rise to data $(\mathbf{y}, \mathbf{X})$; the DLRM (5) is another such model. The advantage of the latter is that it nests the former and thus, if (5) is misspecified, so is (4). In terms of respecifying the LRM to allow for dependence, the DLRM (5) is considerably more general than (4) because it allows for a much less restrictive form of dependence. This suggests that if one wanted to consider the ACLRM as a respecification of the original LRM, one has to establish two things. To begin with, one should estimate the DLRM (5) and ensure its statistical adequacy by testing and not rejecting the assumptions [1]-[5] in table B. That will provide the framework to ensure the statistical validity of the CF restrictions test itself. If these restrictions are tested and not rejected, the ACLRM model will provide a reliable framework for any inference concerning the model parameters $\boldsymbol{\theta}$. If either of these conditions is not met, the inference is likely to be unreliable.

Table B - The Dynamic Linear Regression (DLR) Model

$$\boxed{y_t = \boldsymbol{\beta}_0 + \boldsymbol{\beta}_1^\top \mathbf{x}_t + \gamma_1 y_{t-1} + \boldsymbol{\beta}_2^\top \mathbf{x}_{t-1} + u_t,\ t \in \mathbb{T},}$$

[1]	**Normality:**	$D(y_t \mid \mathbf{x}_t, \mathbf{Z}_{t-1}; \boldsymbol{\theta})$, $\mathbf{Z}_t = (y_t, \mathbf{x}_t)$, is Normal
[2]	**Linearity:**	$E(y_t \mid \mathbf{x}_t, \mathbf{Z}_{t-1}) = \boldsymbol{\beta}_0 + \boldsymbol{\beta}_1^\top \mathbf{x}_t + \gamma_1 y_{t-1} + \boldsymbol{\beta}_2^\top \mathbf{x}_{t-1}$,
[3]	**Homoskedasticity:**	$Var(y_t \mid \mathbf{x}_t, \mathbf{Z}_{t-1}) = \sigma_0^2$, free of $(\mathbf{x}_t, \mathbf{Z}_{t-1})$,
[4]	**t-homogeneity:**	$\boldsymbol{\theta} := (\boldsymbol{\beta}_0, \boldsymbol{\beta}_1, \gamma_1, \boldsymbol{\beta}_2, \sigma_0^2)$ are t-invariant $\forall t \in \mathbb{T}$,
[5]	**Markovness:**	$\{(\mathbf{Z}_t \mid \mathbf{Z}_{t-1}^0),\ t \in \mathbb{T}\}$ - Markov process, where $\mathbf{Z}_{t-1}^0 := (\mathbf{Z}_{t-1}, \mathbf{Z}_{t-2}, ..., \mathbf{Z}_0)$.

5 Conclusion

The primary aim of this paper is to demonstrate that the restrictions implicitly imposed on the dependence structure of y_t and $\mathbf{X}_t$ when an AR(1) error formulation is adopted seem unreasonably restrictive for most real applications. We derive necessary and sufficient conditions on the probabilistic structure of the vector stochastic process $\{\mathbf{Z}_t,\ t \in \mathbb{T}\}$, $\mathbf{Z}_t := (y_t, \mathbf{x}_t^\top)^\top$ for the common factor restrictions (CF) to hold. These restrictions, in the context of the VAR model, amount to:

(a) y_t does **not Granger cause** any of the regressors in $\mathbf{X}_t$, and

(b) $Cov(\mathbf{X}_t, y_t|\mathbf{Z}_{t-1}) = Cov(\mathbf{X}_t|\mathbf{X}_{t-1})Cov(\mathbf{X}_t)^{-1}Cov(\mathbf{X}_t, y_t) = \boldsymbol{\Lambda\beta}$.

[5] By thorough misspecification testing, we mean that *all* the testable assumptions underlying the DLRM are tested (see Spanos, 1986 or 1999).

Condition (a) seems most unrealistic, but it is a testable hypothesis. A new test of the CF restrictions is proposed based on Granger non-causality, and the results of our Monte Carlo experiments suggest that the proposed test outperforms, in terms of correct size and high power, the approximate F-test suggested by Davidson and MacKinnon (1993) as well as a Likelihood Ratio test first proposed by Sargan (1964).

The Monte Carlo evidence also suggests that testing for dependence using only the D-W test is not an effective strategy because its power is considerably reduced when the CF restrictions do not hold. In contrast, the misspecification test based on the DLRM (see (27)) is most probative in general.

The main conclusion of this paper, based on both the theoretical as well as extensive Monte Carlo results, is that adopting the ACLRM as an alternative to the LRM when residual autocorrelation is detected is a bad modeling strategy; echoing the main message in Hendry and Mizon (1978), Hoover (1988), Spanos (1988) and Mizon (1995). In general, we do not recommend the modeling of the error term when misspecifications are detected. This is because the modeling of the error imposes implicit assumptions on the probabilistic structure of the observable vector processes, which are often both unrealistic and unnecessary.

Although we have only considered the implications of error 'correcting' under the very simple AR(1) scenario, typically implemented using time series data, similar unrealistic restrictions are implicitly imposed when researchers use spatial data and model errors as a function of 'near-by' errors. Similarly, it is also clear, that more complicated error structures (i.e. higher order AR or ARMA models) will impose more complicated and unrealistic restrictions. In general, before any kind of 'correcting' for misspecifications by modeling the error takes place, it seems prudent to make explicit, and then test, the restrictions being implicitly imposed on the probabilistic structure of the data.

References

[1] Anselin, L. (2001), "Spatial Econometrics," ch. 14 in *A Companion to Theoretical Econometrics*, edited by B. H. Baltagi, Blackwell Publishers, Oxford.

[2] Breusch, T. S. (1978), "Testing for autocorrelation in dynamic linear models," *Australian Economic Papers*, **17**, 334-355.

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

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

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

[6] Godfrey, L. G. (1978), "Testing against general autoregressive and moving average error models when the regressors include lagged endogenous variables," *Econometrica*, **46**, 1293-1302.

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

[8] 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.

[9] 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.

[10] Mayo, D. G. (1996), *Error and Growth of Scientific Knowledge*, Chicago University Press, Chicago.

[11] Mayo, D. G. and A. Spanos (2004), "Methodology in Practice: Statistical Misspecification Testing," forthcoming, *Philosophy of Science*.

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

[13] Newy, W. K. and K. West (1987), "A Simple, Positive Semi-Definite, Heteroskedastic and Autocorrelation Consistent Covariance Matrix," *Econometrica*, **55**, 703-708.

[14] 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.

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

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

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

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

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

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

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

[22] 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.

6 Appendix A

The ACLRM is specified as follows:

$$y_t = \boldsymbol{\beta}^\top \mathbf{x}_t + u_t, \qquad u_t = \rho u_{t-1} + \varepsilon_t, \quad |\rho| < 1, \quad t \in \mathbb{T},$$

(30)

$$\begin{aligned}&[\text{a1}]\ E(\varepsilon_t) = 0,\ Var(\varepsilon_t) = \sigma^2,\ Cov(\varepsilon_t, \varepsilon_s) = 0, t \neq s,\\ &[\text{a2}]\ Cov(\mathbf{X}_t, u_s) = \mathbf{0},\ \forall\,(t,s) \in \mathbb{T}.\end{aligned}$$

Theorem 1. The mapping between the primary parameters $\boldsymbol{\Sigma}$ and the parameters $\boldsymbol{\theta}_1^* := (\alpha_1, \boldsymbol{\beta}_0, \sigma^2)$ of the ACLRM (8) takes the form [1]-[6].

$$\boxed{[\mathbf{1}] \quad Var(u_t^2) = \sigma_{uu}(0) = \tfrac{\sigma^2}{1-\rho^2}.}$$

$Var(u_t^2) = E(u_t^2) = E\left[(\rho u_{t-1} + \varepsilon_t)(\rho u_{t-1} + \varepsilon_t)\right] = \rho^2 E(u_{t-1}^2) + E(\varepsilon_t^2) = \rho^2 E(u_t^2) + E(\varepsilon_t^2)$, from [a1]. Given $|\rho| < 1$, $E(u_t^2) = \sigma_{uu}(0) = \frac{\sigma^2}{1-\rho^2}$.

$$\boxed{[\mathbf{2}] \quad Cov(u_t u_{t-1}) = \sigma_{uu}(1) = \tfrac{\rho\sigma^2}{1-\rho^2}.}$$

$Cov(u_t u_{t-1}) = E(u_t u_{t-1}) = E\left[(\rho u_{t-1} + \varepsilon_t)u_{t-1}\right] = \rho E(u_{t-1}^2) = \rho\sigma_{uu}(0) = \frac{\rho\sigma^2}{1-\rho^2}$ from [a1] and [**1**].

$$\boxed{[\mathbf{3}] \quad \sigma_{21}(0) = \Sigma_{22}(0)\boldsymbol{\beta}.}$$

$\boldsymbol{\sigma}_{21}(0) = E(\mathbf{x}_t y_t) = E\left[\mathbf{x}_t\left(\boldsymbol{\beta}^\top \mathbf{x}_t + u_t\right)'\right] = E\left[\mathbf{x}_t \mathbf{x}_t'\right]\boldsymbol{\beta} = \Sigma_{22}(0)\boldsymbol{\beta}$ from [a2]. Thus, $\boldsymbol{\sigma}_{21}(0) = \boldsymbol{\Sigma}_{22}(0)\boldsymbol{\beta}$ or $\boldsymbol{\beta} = \boldsymbol{\Sigma}_{22}^{-1}(0)\boldsymbol{\sigma}_{21}(0)$.

$$\boxed{[\mathbf{4}] \quad \boldsymbol{\sigma}_{21}(1) = \boldsymbol{\Sigma}_{22}(1)\boldsymbol{\beta}.}$$

$\boldsymbol{\sigma}_{21}(1) = E(\mathbf{x}_t y_{t-1}) = E\left[\mathbf{x}_t\left(\boldsymbol{\beta}^\top \mathbf{x}_{t-1} + u_{t-1}\right)'\right] = E\left[\mathbf{x}_t \mathbf{x}_{t-1}'\right]\boldsymbol{\beta} = \Sigma_{22}(1)\boldsymbol{\beta}$ from [a2]. Thus, $\boldsymbol{\sigma}_{21}(1) = \boldsymbol{\Sigma}_{22}(1)\boldsymbol{\beta}$ or $\boldsymbol{\beta} = \boldsymbol{\Sigma}_{22}^{-1}(1)\boldsymbol{\sigma}_{21}(1)$.

$$\boxed{[\mathbf{5}] \quad \boldsymbol{\sigma}_{11}(0) = \boldsymbol{\beta}^\top \boldsymbol{\Sigma}_{22}(0)\boldsymbol{\beta} + \tfrac{\sigma^2}{1-\rho^2}.}$$

$\boldsymbol{\sigma}_{11}(0) = E(y_t y_t) = E\left[\left(\boldsymbol{\beta}^\top \mathbf{x}_t + u_t\right)\left(\boldsymbol{\beta}^\top \mathbf{x}_t + u_t\right)^\top\right] = E\left[\left(\boldsymbol{\beta}^\top \mathbf{x}_t \mathbf{x}_t^\top \boldsymbol{\beta} + u_t^2\right)\right]$ using [a2]. Using $E(u_t^2) = \frac{\sigma^2}{1-\rho^2}$ from [**1**], $\boldsymbol{\sigma}_{11}(0) = \boldsymbol{\beta}^\top \boldsymbol{\Sigma}_{22}(0)\boldsymbol{\beta} + \frac{\sigma^2}{1-\rho^2}$.

$$\boxed{[\mathbf{6}] \quad \boldsymbol{\sigma}_{11}(1) = \boldsymbol{\beta}^\top \boldsymbol{\Sigma}_{22}(1)\boldsymbol{\beta} + \tfrac{\rho\sigma^2}{1-\rho^2}.}$$

$\boldsymbol{\sigma}_{11}(1) = E(y_t y_{t-1}) = E\left[\left(\boldsymbol{\beta}^\top \mathbf{x}_t + u_t\right)\left(\boldsymbol{\beta}^\top \mathbf{x}_{t-1} + u_{t-1}\right)^\top\right] = E\left[\left(\boldsymbol{\beta}^\top \mathbf{x}_t \mathbf{x}_{t-1}^\top \boldsymbol{\beta} + u_t u_{t-1}\right)\right]$ using [a2]. From [**2**], $\boldsymbol{\sigma}_{11}(1) = \boldsymbol{\beta}^\top \boldsymbol{\Sigma}_{22}(1)\boldsymbol{\beta} + \frac{\rho\sigma^2}{1-\rho^2}$.

Theorem 2. The implicit statistical parameterization for $(\alpha_1, \beta_0, \beta_1, \boldsymbol{\sigma}_\varepsilon^2)$ implied by the variance-covariance matrix $\mathbf{\Sigma}^*$ in (10) is:
(31)

$$\begin{pmatrix} \alpha_1 \\ \boldsymbol{\beta}_0 \\ \boldsymbol{\beta}_1 \end{pmatrix} = \begin{pmatrix} \boldsymbol{\beta}^\top \mathbf{\Sigma}_{22}(0)\boldsymbol{\beta} + \frac{\sigma^2}{1-\rho^2} & \boldsymbol{\beta}^\top \Sigma_{22}(1) & \boldsymbol{\beta}^\top \mathbf{\Sigma}_{22}(0) \\ \mathbf{\Sigma}_{22}(1)^\top \boldsymbol{\beta} & \mathbf{\Sigma}_{22}(0) & \mathbf{\Sigma}_{22}(1) \\ \mathbf{\Sigma}_{22}(0)\boldsymbol{\beta} & \mathbf{\Sigma}_{22}(1) & \mathbf{\Sigma}_{22}(0) \end{pmatrix}^{-1} \begin{pmatrix} \boldsymbol{\beta}^\top \mathbf{\Sigma}_{22}(1)\boldsymbol{\beta} + \frac{\rho\sigma^2}{1-\rho^2} \\ \mathbf{\Sigma}_{22}(0)\boldsymbol{\beta} \\ \mathbf{\Sigma}_{22}(1)\boldsymbol{\beta} \end{pmatrix}$$

$$\boldsymbol{\sigma}_\varepsilon^2 = \boldsymbol{\beta}^\top \mathbf{\Sigma}_{22}(0)\boldsymbol{\beta} + \frac{\sigma^2}{1-\rho^2} - \begin{pmatrix} \boldsymbol{\beta}^\top \mathbf{\Sigma}_{22}(1)\boldsymbol{\beta} + \frac{\rho\sigma^2}{1-\rho^2} & \boldsymbol{\beta}^\top \mathbf{\Sigma}_{22}(0) & \boldsymbol{\beta}^\top \mathbf{\Sigma}_{22}(1) \end{pmatrix} \begin{pmatrix} \alpha_1 \\ \boldsymbol{\beta}_0 \\ \boldsymbol{\beta}_1 \end{pmatrix}.$$

Using the partitioned matrix inverse result in Searle (1982):
(32)

$$\begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{B}^\top & \mathbf{C} \end{bmatrix}^{-1} = \begin{bmatrix} \left(\mathbf{A} - \mathbf{B}\mathbf{C}^{-1}\mathbf{B}^\top\right)^{-1} & -\left(\mathbf{A} - \mathbf{B}\mathbf{C}^{-\mathbf{1}}\mathbf{B}^\top\right)^{-\mathbf{1}} \mathbf{B}\mathbf{C}^{-1} \\ -\mathbf{C}^{-1}\mathbf{B}^\top \left(\mathbf{A} - \mathbf{B}\mathbf{C}^{-1}\mathbf{B}^\top\right)^{-1} & \mathbf{C}^{-1} + \mathbf{C}^{-1}\mathbf{B}^\top \left(\mathbf{A} - \mathbf{B}\mathbf{C}^{-1}\mathbf{B}^\top\right)^{-1} \mathbf{B}\mathbf{C}^{-1} \end{bmatrix}$$

we proceed to derive the inverse of the matrix:

$$\begin{pmatrix} \boldsymbol{\beta}^\top \mathbf{\Sigma}_{22}(0)\boldsymbol{\beta} + \frac{\sigma^2}{1-\rho^2} & \beta^\top \Sigma_{22}(1) & \beta^\top \Sigma_{22}(0) \\ \Sigma_{22}(1)\beta & \mathbf{\Sigma}_{22}(0) & \mathbf{\Sigma}_{22}(1) \\ \Sigma_{22}(0)\beta & \mathbf{\Sigma}_{22}(1) & \mathbf{\Sigma}_{22}(0) \end{pmatrix}$$

where we define $\mathbf{A} = \begin{bmatrix} \boldsymbol{\beta}^\top \mathbf{\Sigma}_{22}(0)\boldsymbol{\beta} + \frac{\sigma^2}{1-\rho^2} & \beta^\top \Sigma_{22}(1) \\ \Sigma_{22}(1)\beta & \mathbf{\Sigma}_{22}(0) \end{bmatrix}$, $\mathbf{B} = \begin{bmatrix} \beta^\top \Sigma_{22}(0) \\ \mathbf{\Sigma}_{22}(1) \end{bmatrix}$, and $\mathbf{C} = \mathbf{\Sigma}_{22}(0)$, to find:

$$\begin{pmatrix} \boldsymbol{\beta}^\top \mathbf{\Sigma}_{22}(0)\boldsymbol{\beta} + \frac{\sigma^2}{1-\rho^2} & \beta^\top \Sigma_{22}(1) & \beta^\top \Sigma_{22}(0) \\ \Sigma_{22}(1)\beta & \mathbf{\Sigma}_{22}(0) & \mathbf{\Sigma}_{22}(1) \\ \Sigma_{22}(0)\beta & \mathbf{\Sigma}_{22}(1) & \mathbf{\Sigma}_{22}(0) \end{pmatrix}^{-1} =$$

$$= \begin{pmatrix} \frac{(1-\rho^2)}{\sigma^2} & 0_{(1xk)} & -\frac{(1-\rho^2)}{\sigma^2}\beta^\top \\ 0_{(kx1)} & \Delta & -\Delta\Sigma_{22}(1)\Sigma_{22}(0)^{-1} \\ \left(-\frac{(1-\rho^2)}{\sigma^2}\beta^\top\right) & \left(-\Sigma_{22}(0)^{-1}\Sigma_{22}(1)\Delta\right) & \left(\Psi + \beta\beta^\top \frac{(1-\rho^2)}{\sigma^2}\right) \end{pmatrix}$$

where:

$$\Psi = \Sigma_{22}(0)^{-1} + \Sigma_{22}(0)^{-1}\Sigma_{22}(1)\Delta\Sigma_{22}(1)\Sigma_{22}(0)^{-1},$$

$$\mathbf{\Delta} = \left[\mathbf{\Sigma}_{22}(0) - \mathbf{\Sigma}_{22}(1)\mathbf{\Sigma}_{22}(0)^{-1}\mathbf{\Sigma}_{22}(1)\right]^{-1}.$$

Noting that:

$$\mathbf{\Delta} = \Psi$$

(see Henderson and Searle, 1981, p. 53), the inverse simplifies to:

$$\begin{pmatrix} \frac{(1-\rho^2)}{\sigma^2} & \mathbf{0}_{(1xk)} & -\frac{(1-\rho^2)}{\sigma^2}\boldsymbol{\beta}^\top \\ \mathbf{0}_{(kx1)} & \mathbf{\Delta} & -\Delta\mathbf{\Sigma}_{22}(1)\mathbf{\Sigma}_{22}(0)^{-1} \\ -\frac{(1-\rho^2)}{\sigma^2}\boldsymbol{\beta}^\top & -\mathbf{\Sigma}_{22}(0)^{-1}\mathbf{\Sigma}_{22}(1)\Delta & \mathbf{\Delta} + \boldsymbol{\beta}\boldsymbol{\beta}^\top \frac{(1-\rho^2)}{\sigma^2} \end{pmatrix}.$$

Thus,

(33)

$$\begin{pmatrix} \alpha_1 \\ \boldsymbol{\beta}_0 \\ \boldsymbol{\beta}_1 \end{pmatrix} = \begin{pmatrix} \frac{(1-\rho^2)}{\sigma^2} & \mathbf{0}_{(1xk)} & -\frac{(1-\rho^2)}{\sigma^2}\boldsymbol{\beta}^\top \\ \mathbf{0}_{(kx1)} & \boldsymbol{\Delta} & -\boldsymbol{\Delta}\boldsymbol{\Sigma}_{22}(1)\boldsymbol{\Sigma}_{22}(0)^{-1} \\ -\frac{(1-\rho^2)}{\sigma^2}\boldsymbol{\beta}^\top & -\boldsymbol{\Sigma}_{22}(0)^{-1}\boldsymbol{\Sigma}_{22}(1)\boldsymbol{\Delta} & \boldsymbol{\Delta}+\boldsymbol{\beta}\boldsymbol{\beta}^\top\frac{(1-\rho^2)}{\sigma^2} \end{pmatrix} \begin{pmatrix} \boldsymbol{\beta}^\top\boldsymbol{\Sigma}_{22}(1)\boldsymbol{\beta}+\frac{\rho\sigma^2}{1-\rho^2} \\ \Sigma_{22}(0)\boldsymbol{\beta} \\ \Sigma_{22}(1)\boldsymbol{\beta} \end{pmatrix}$$

Multiplying and simplifying the right-hand-side yields:

$$\begin{pmatrix} \frac{(1-\rho^2)}{\sigma^2}\boldsymbol{\beta}^\top\boldsymbol{\Sigma}_{22}(1)\boldsymbol{\beta}+\rho-\frac{(1-\rho^2)}{\sigma^2}\boldsymbol{\beta}^\top\boldsymbol{\Sigma}_{22}(1)\boldsymbol{\beta} \\ \boldsymbol{\Delta}\Sigma_{22}(0)\boldsymbol{\beta}-\Delta\boldsymbol{\Sigma}_{22}(1)\boldsymbol{\Sigma}_{22}(0)^{-1}\Sigma_{22}(1)\boldsymbol{\beta} \\ -\frac{(1-\rho^2)}{\sigma^2}\boldsymbol{\beta}\boldsymbol{\beta}^\top\boldsymbol{\Sigma}_{22}(1)\boldsymbol{\beta}-\boldsymbol{\rho}\boldsymbol{\beta}-\boldsymbol{\Sigma}_{22}(0)^{-1}\boldsymbol{\Sigma}_{22}(1)\Delta\Sigma_{22}(0)\boldsymbol{\beta}+\Delta\Sigma_{22}(1)\boldsymbol{\beta}+\frac{(1-\rho^2)}{\sigma^2}\boldsymbol{\beta}^\top\boldsymbol{\beta}^\top\boldsymbol{\Sigma}_{22}(1)\boldsymbol{\beta} \end{pmatrix},$$

thus,

$$\begin{pmatrix} \alpha_1 \\ \boldsymbol{\beta}_0 \\ \boldsymbol{\beta}_1 \end{pmatrix} = \begin{pmatrix} \rho \\ \boldsymbol{\Delta}(\boldsymbol{\Sigma}_{22}(0)-\boldsymbol{\Sigma}_{22}(1)\boldsymbol{\Sigma}_{22}(0)^{-1}\boldsymbol{\Sigma}_{22}(1))\boldsymbol{\beta} \\ -\boldsymbol{\rho}\boldsymbol{\beta}+(\boldsymbol{\Delta}\boldsymbol{\Sigma}_{22}(1)-\boldsymbol{\Sigma}_{22}(0)^{-1}\boldsymbol{\Sigma}_{22}(1)\boldsymbol{\Delta}\boldsymbol{\Sigma}_{22}(0))\boldsymbol{\beta} \end{pmatrix} = \begin{pmatrix} \rho \\ \boldsymbol{\beta} \\ -\rho\boldsymbol{\beta} \end{pmatrix}$$

since $\boldsymbol{\Sigma}_{22}(0)-\boldsymbol{\Sigma}_{22}(1)\boldsymbol{\Sigma}_{22}(0)^{-1}\boldsymbol{\Sigma}_{22}(1)=\boldsymbol{\Delta}^{-1}$ and $\boldsymbol{\Delta}\boldsymbol{\Sigma}_{22}(1)=\boldsymbol{\Sigma}_{22}(0)^{-1}\boldsymbol{\Sigma}_{22}(1)\boldsymbol{\Delta}\boldsymbol{\Sigma}_{22}(0)$. This follows from:

$$(D-VA^{-1}U)^{-1}VA^{-1}=D^{-1}V(A-UD^{-1}V)^{-1},$$

(see Henderson and Searle, 1981, p.56) by letting $A=D=\Sigma_{22}(0)$ and $U=V=\Sigma_{22}(1)$.

Substituting this result into the expression for $\boldsymbol{\sigma}_\varepsilon^2$:

$$\boldsymbol{\sigma}_\varepsilon^2=\boldsymbol{\beta}^\top\boldsymbol{\Sigma}_{22}(0)\boldsymbol{\beta}+\frac{\sigma^2}{1-\rho^2}-\begin{pmatrix} \boldsymbol{\beta}^\top\boldsymbol{\Sigma}_{22}(1)\boldsymbol{\beta}+\frac{\rho\sigma^2}{1-\rho^2} & \boldsymbol{\beta}^\top\boldsymbol{\Sigma}_{22}(0) & \boldsymbol{\beta}^\top\boldsymbol{\Sigma}_{22}(1) \end{pmatrix}\begin{pmatrix} \rho \\ \beta \\ -\rho\boldsymbol{\beta} \end{pmatrix}=\sigma^2,$$

confirming that the variance-covariance matrix (10) leads to a dynamic linear regression model where the common factor restrictions hold.

Theorem 3. The implicit statistical parameterization of the VAR model, based on $D(\mathbf{Z}_t|\mathbf{Z}_{t-1};\boldsymbol{\psi})$:

$$\mathbf{Z}_t=\mathbf{A}^\top\mathbf{Z}_{t-1}+\mathbf{E}_t,\quad \mathbf{E}_t\backsim \mathsf{IID}(0,\boldsymbol{\Omega}),\ t\in\mathbb{T}. \tag{34}$$

implied by the restricted variance-covariance matrix $\boldsymbol{\Sigma}^*$ in (10), takes the form:

$$\mathbf{A}^\top=\begin{pmatrix} \rho & (\mathbf{D}-\boldsymbol{\rho}\mathbf{I}_k)\boldsymbol{\beta} \\ \mathbf{0} & \mathbf{D} \end{pmatrix},\quad \boldsymbol{\Omega}=\begin{pmatrix} \sigma^2+\boldsymbol{\beta}^\top\boldsymbol{\Lambda}\boldsymbol{\beta} & \boldsymbol{\beta}^\top\boldsymbol{\Lambda} \\ \boldsymbol{\Lambda}\boldsymbol{\beta} & \boldsymbol{\Lambda} \end{pmatrix}$$

where:

$$\mathbf{D}=\Sigma_{22}(0)^{-1}\Sigma_{22}(1)\quad \boldsymbol{\Lambda}=\boldsymbol{\Sigma}_{22}(0)-\Sigma_{22}(1)\Sigma_{22}(0)^{-1}\Sigma_{22}(1)$$

and $\mathbf{I}_k$ is a $k\times k$ identity matrix. From Lemma 1:

$$\mathbf{A}=\begin{pmatrix} \boldsymbol{\beta}^\top\boldsymbol{\Sigma}_{22}(0)\boldsymbol{\beta}+\frac{\sigma^2}{1-\rho^2} & \boldsymbol{\beta}^\top\boldsymbol{\Sigma}_{22}(0) \\ \boldsymbol{\Sigma}_{22}(0)\boldsymbol{\beta} & \boldsymbol{\Sigma}_{22}(0) \end{pmatrix}^{-1}\begin{pmatrix} \boldsymbol{\beta}^\top\boldsymbol{\Sigma}_{22}(1)\boldsymbol{\beta}+\frac{\rho\sigma^2}{1-\rho^2} & \boldsymbol{\beta}^\top\boldsymbol{\Sigma}_{22}(1) \\ \boldsymbol{\Sigma}_{22}(1)\boldsymbol{\beta} & \boldsymbol{\Sigma}_{22}(1) \end{pmatrix}$$

and

$$\boldsymbol{\Omega} \;=\left(\begin{array}{cc} \boldsymbol{\beta}^{\top}\boldsymbol{\Sigma}_{22}(0)\boldsymbol{\beta}+\frac{\sigma^2}{1-\rho^2} & \boldsymbol{\beta}^{\top}\boldsymbol{\Sigma}_{22}(0) \\ \boldsymbol{\Sigma}_{22}(0)\boldsymbol{\beta} & \boldsymbol{\Sigma}_{22}(0) \end{array}\right)-\left(\begin{array}{cc} \boldsymbol{\beta}^{\top}\boldsymbol{\Sigma}_{22}(1)\boldsymbol{\beta}+\frac{\rho\sigma^2}{1-\rho^2} & \boldsymbol{\Sigma}_{22}(1)\boldsymbol{\beta} \\ \boldsymbol{\Sigma}_{22}(1)\boldsymbol{\beta} & \boldsymbol{\Sigma}_{22}(1) \end{array}\right)\mathbf{A}.$$

Using (32) to find the inverse of:

$$\left(\begin{array}{cc} \boldsymbol{\beta}^{\top}\boldsymbol{\Sigma}_{22}(0)\boldsymbol{\beta}+\frac{\sigma^2}{1-\rho^2} & \boldsymbol{\beta}^{\top}\boldsymbol{\Sigma}_{22}(0) \\ \boldsymbol{\Sigma}_{22}(0)\boldsymbol{\beta} & \boldsymbol{\Sigma}_{22}(0) \end{array}\right)$$

yields:

$$\left(\begin{array}{cc} \left(\frac{1-\rho^2}{\sigma^2}\right) & -\left(\frac{1-\rho^2}{\sigma^2}\right)\boldsymbol{\beta}^{\top} \\ -\left(\frac{1-\rho^2}{\sigma^2}\right)\boldsymbol{\beta} & \left(\boldsymbol{\Sigma}_{22}(0)^{-1}+\left(\frac{1-\rho^2}{\sigma^2}\right)\boldsymbol{\beta}\boldsymbol{\beta}^{\top}\right) \end{array}\right).$$

Thus,

$$\mathbf{A} \;=\left(\begin{array}{cc} \left(\frac{1-\rho^2}{\sigma^2}\right) & -\left(\frac{1-\rho^2}{\sigma^2}\right)\boldsymbol{\beta}^{\top} \\ -\left(\frac{1-\rho^2}{\sigma^2}\right)\boldsymbol{\beta} & \left(\boldsymbol{\Sigma}_{22}(0)^{-1}+\left(\frac{1-\rho^2}{\sigma^2}\right)\boldsymbol{\beta}\boldsymbol{\beta}^{\top}\right) \end{array}\right)\left(\begin{array}{cc} \left(\boldsymbol{\beta}^{\top}\boldsymbol{\Sigma}_{22}(1)\boldsymbol{\beta}+\frac{\rho\sigma^2}{1-\rho^2}\right) & \boldsymbol{\beta}^{\top}\boldsymbol{\Sigma}_{22}(1) \\ \boldsymbol{\Sigma}_{22}(1)\boldsymbol{\beta} & \boldsymbol{\Sigma}_{22}(1) \end{array}\right)$$

$$\mathbf{A} \;=\left(\begin{array}{cc} \rho & \mathbf{0} \\ \boldsymbol{\Sigma}_{22}(0)^{-1}\boldsymbol{\Sigma}_{22}(1)\boldsymbol{\beta}-\boldsymbol{\rho}\boldsymbol{\beta} & \boldsymbol{\Sigma}_{22}(0)^{-1}\boldsymbol{\Sigma}_{22}(1) \end{array}\right).$$

Substituting this result for $\mathbf{A}$ into the expression for $\boldsymbol{\Omega}$ and multiplying yields:

$$\left(\begin{array}{cc} \boldsymbol{\beta}^{\top}\boldsymbol{\Sigma}_{22}(0)\boldsymbol{\beta}+\frac{\sigma^2}{1-\rho^2} & \boldsymbol{\beta}^{\top}\boldsymbol{\Sigma}_{22}(0) \\ \boldsymbol{\Sigma}_{22}(0)\boldsymbol{\beta} & \boldsymbol{\Sigma}_{22}(0) \end{array}\right)-\left(\begin{array}{cc} \boldsymbol{\beta}^{\top}\boldsymbol{\Sigma}_{22}(1)\boldsymbol{\Sigma}_{22}(0)^{-1}\boldsymbol{\Sigma}_{22}(1)\boldsymbol{\beta}+\frac{\rho^2\sigma^2}{1-\rho^2} & \boldsymbol{\beta}^{\top}\boldsymbol{\Sigma}_{22}(1)\boldsymbol{\Sigma}_{22}(0)^{-1}\boldsymbol{\Sigma}_{22}(1) \\ \boldsymbol{\Sigma}_{22}(1)\boldsymbol{\Sigma}_{22}(0)^{-1}\boldsymbol{\Sigma}_{22}(1)\boldsymbol{\beta} & \boldsymbol{\Sigma}_{22}(1)\boldsymbol{\Sigma}_{22}(0)^{-1}\boldsymbol{\Sigma}_{22}(1) \end{array}\right).$$

Simplifying further we obtain:

$$\boldsymbol{\Omega}=\left(\begin{array}{cc} \sigma^2+\boldsymbol{\beta}^{\top}(\boldsymbol{\Sigma}_{22}(0)-\Sigma_{22}(1)\Sigma_{22}(0)^{-1}\Sigma_{22}(1))\boldsymbol{\beta} & \boldsymbol{\beta}^{\top}(\boldsymbol{\Sigma}_{22}(0)-\Sigma_{22}(1)\Sigma_{22}(0)^{-1}\Sigma_{22}(1)) \\ (\boldsymbol{\Sigma}_{22}(0)-\Sigma_{22}(1)\Sigma_{22}(0)^{-1}\Sigma_{22}(1))\boldsymbol{\beta} & \boldsymbol{\Sigma}_{22}(0)-\Sigma_{22}(1)\Sigma_{22}(0)^{-1}\Sigma_{22}(1) \end{array}\right).$$

Using Lemma 1, to obtain the parameterization of the VAR model for $\mathbf{X}_t$ based on $D(\mathbf{X}_t \mid \mathbf{X}_{t-1};\boldsymbol{\theta})$:

$$\mathbf{X}_t=\mathbf{D}^{\top}\mathbf{X}_{t-1}+\mathbf{V}_t, \mathbf{V}_t \backsim \mathsf{IID}(0,\boldsymbol{\Lambda}), \tag{35}$$

we obtain:

$$\mathbf{D}=\Sigma_{22}(0)^{-1}\Sigma_{22}(1) \quad \boldsymbol{\Lambda}=\boldsymbol{\Sigma}_{22}(0)-\Sigma_{22}(1)\Sigma_{22}(0)^{-1}\Sigma_{22}(1).$$

Thus, we can rewrite the above expressions for the VAR parameters $(\mathbf{A},\boldsymbol{\Omega})$, based on $D(\mathbf{Z}_t|\mathbf{Z}_{t-1};\boldsymbol{\psi})$ as follows:

$$\mathbf{A}=\left(\begin{array}{cc} \rho & \mathbf{0} \\ (\mathbf{D}-\boldsymbol{\rho}\mathbf{I}_k)\boldsymbol{\beta} & \mathbf{D} \end{array}\right), \qquad \boldsymbol{\Omega}=\left(\begin{array}{cc} \sigma^2+\boldsymbol{\beta}^{\top}\boldsymbol{\Lambda}\boldsymbol{\beta} & \boldsymbol{\beta}^{\top}\boldsymbol{\Lambda} \\ \boldsymbol{\Lambda}\boldsymbol{\beta} & \boldsymbol{\Lambda} \end{array}\right),$$

where $\mathbf{I}_k$ is a $k\times k$ identity matrix.