ANALYSIS OF VECTOR AUTOREGRESSIONS IN THE PRESENCE OF SHIFTS IN MEAN

Serena Ng[*]   Timothy J. Vogelsang[†]

July 1997

Abstract

This paper considers the implications of omitted mean shifts for estimation and inference in VARs. It is shown that the least squares estimates are inconsistent, and the F test for Granger causality diverges. While model selection rules have the tendency to incorrectly select a lag length that is too high, this over-parameterization can reduce size distortions in tests involving the inconsistent estimates. The practical issue of how to remove the breaks is shown to depend on whether the mean shifts are of the additive or innovational type in a multivariate setting. Under the additive outlier specification, the intercept in each equation of the VAR will be subject to multiple shifts when the break dates of the mean shifts to the univariate series do not coincide. Conversely, under the innovative outlier specification, the unconditional means of the univariate time series are subject to multiple shifts when mean shifts to the innovation processes occur at different dates. Techniques designed to detect multiple shifts are recommended when break dates do not coincide.

Keywords: trend shift, structural change, causality tests, lag length selection.
JEL Classification: C2,C3,C5

[*]Department of Economics, Boston College.

[†]Department of Economics, Cornell University.

Correspondence: Tim Vogelsang, Dept. of Economics, Uris Hall, Cornell University, Ithaca, N.Y. 14583-7601 Tel. (607) 255-5108 Fax. (607) 255-2818. Email: tjv2@cornell.edu, serena.ng@bc.edu

1 Introduction

In recent years, many univariate statistics have been developed to test for the presence of structural breaks in stationary and non-stationary time series. When applied to macroeconomic data, the evidence suggests that breaks in the form of a shift in mean and/or the trend function have occurred in many series. Clearly, if there are breaks in the univariate series and these breaks are not accounted for in multivariate work, misspecification will result. While it is fairly well known how an unstable mean affects univariate analysis, much less is known about how unstable means affect multivariate analysis. The purpose of this paper is to explore the ramifications of unstable means in a vector of time series on estimation and inference using vector autoregressions (VARs). We focus on whether shifts in the mean of one or more series will induce bias in the OLS estimates of the VAR and consider the properties of tests for Granger causality and lag length selection in the presence of unstable means.

Vector autoregressions are widely used in macroeconomic analyses. Any model misspecification will affect the impulse response functions and the decomposition of variances and hence mislead our understanding of macroeconomic dynamics. This point has long been recognized. For example, Stock and Watson (1989) suggested that because inflation has a time trend, the money-income causal relationship can be shown to be significant when the neglected trend is properly included in the VAR. Evidence that mean shifts affect analyses of VARs also exists in the literature. For example, Gambe and Joutz (1993) find that if the mean shift in the growth rate of output in 1973 and the trend in the unemployment rate were not taken into account, the importance of labor-demand shocks in variations in real wages dropped, while those of aggregate demand shocks rose. On a larger scale, Stock and Watson (1996) found that in a sample of 76 representative U.S. monthly postwar macroeconomic time series there is substantial structural instability in bivariate relationships. Some of that structural instability is no doubt due to unstable means. Although simulations studies by Lutkepohl (1989) and Bianchi (1995) have shown that Granger causality tests have excessive size in finite samples when mean shifts are ignored, there does not exist a systematic theoretical analysis of how omitted mean shifts affect VARs in the literature.

This paper provides a formal analysis for the properties of VARs when shifts in the function of deterministic variables are ignored. In particular we examine the consistency properties of estimates of reduced form VARs in the presence of mean shifts and discuss implications for Granger causality tests and lag length selection in the associated regressions.

Since structural VAR parameters are functions of reduced form parameters, implications for structural estimation (e.g. impulse responses and variance decompositions) follow directly from our results. However, we do not explicitly consider structural estimation as this subject is beyond the scope of the paper.

Effects of omitted mean shifts are similar in spirit to the effects of omitted time trends, and both can prove equally treacherous in applied work. However, there is a fundamental difference between the two problems. For data of a given span, the time trend for the sample in question is defined up to scale. That is to say, the trend is sample specific rather than variable specific. It follows from the Frisch-Waugh theorem that univariate detrending or including a trend in the VAR has the same effect on the parameter estimates. However, mean shifts are variable specific to the extent that the break dates in a n-vector model may not coincide. Therefore, even if each variable experiences just one break, the number of omitted breaks in the VAR will be larger than one unless all variables experience mean shifts at the same time.

The issue of how to remove breaks in a univariate series has been analyzed under a variety of conditions. In general, the coefficients associated with the deterministic terms can be consistently estimated by projecting the observed series on a constant and/or trend, and the break dummies, and based on these estimated coefficients, a mean zero stochastic process can be constructed. Alternatively, break dummies can be added directly to the autoregressive representation of the series. Essentially, there is little practical difference between these two approaches. In the multivariate setting things are more complicated, and consideration needs to be given to dates of the breaks and how the breaks filter through the system. Similar to the univariate setting one approach is to identify and remove breaks series by series before estimating the VAR. The second approach is to place break dummies directly into the VAR representation. But, depending on how mean shifts are modeled, mean shifts at different break dates can induce multiple intercept shifts in each equation of the VAR or multiple mean shifts in the univariate representations of the series. Therefore, techniques that can identify multiple shifts such as Bai and Perron (1995) may be required when estimating VARs subject to mean shifts.

The rest of the paper proceeds as follows. Section 2 considers the properties of the OLS estimates for a bivariate VAR when mean shifts are omitted. Implications for inference are considered in Section 3. Model selection in the context of a misspecified model is also considered. Section 4 considers the appropriate methods for removing the breaks in multivariate models. An empirical example is presented in Section 5. Section 6 concludes. Although the

focus of this paper is on mean shifts, qualitatively similar results are found to hold if there are breaks in the trend instead of the mean. The generality of the results stated in this paper are also be discussed.

2 The Model

There are two common approaches to modeling mean shifts in a univariate time series which are based time series models with outliers (e.g. Tiao (1985) and Box and Tiao (1975)). The first is the "additive outlier" (AO) approach which specifies a series y_t as the sum of a deterministic component $\mu + \delta DU_t$ and a stochastic component, $z_t = az_{t-1} + e_t$, where $DU_t = 1(t > T_B)$, $1(t > T_B)$ is the indicator function, T_B is the break date and e_t is a martingale difference sequence. The second is the "innovational outlier" (IO) approach which models the break as occurring to the mean of the innovation series. More formally, y_t is specified as $y_t = \mu + z_t$, where $z_t = az_{t-1} + \delta DU_t + e_t$. Both approaches were used by Perron (1990), among others, in the context of univariate unit root tests that allow a shift in mean. The IO model has the advantage of permitting mean shifts to occur gradually over time, but has the disadvantage that the dynamic path of the break is constrained to follow the dynamics of the stochastic part of the series.

The AO and IO approach to modeling mean shifts in univariate processes provides a natural starting point to modeling mean shifts in multivariate time series. To simplify the analysis and notation, we only focus on the bivariate case. Consider the following model for the bivariate series $\{y_t\} = (y_{1t}, y_{2t})'$ for $t = 1, \dots, T$. A bivariate additive outlier mean shift model can be specified as follow:

$$y_t = \mu + D_t \delta + z_t \tag{1}$$

$$z_t = A z_{t-1} + e_t \tag{2}$$

where $z_t = (z_{1t}, z_{2t})'$, $e_t = (e_{1t}, e_{2t})'$, $\delta = (\delta_1, \delta_2)'$, D_t is a 2×2 diagonal matrix with $diag(D_t) = (DU_{1t}, DU_{2t})$, $DU_{it} = 1(t > T_{Bi})$, $i = 1, 2$ where T_{Bi} is the break date for the series y_i and A is a 2×2 matrix with elements a_{ij}.

By transforming the above model in standard fashion, (1) and (2) can be written as,

$$y_t = \mu^* + D_t^* \delta + A \Delta D_t \delta + A y_{t-1} + e_t, \tag{3}$$

where $\mu^* = (I_2 - A)\mu$, $D_t^* = (I_2 - A)D_t$ and ΔD_t is a 2×2 diagonal matrix with $diag(\Delta D_t) = (DTB_{1t}, DTB_{2t})$, where Δ is the first difference operator, $DTB_{it} = 1(t = T_{Bi} + 1), i = 1, 2$.

3

Model (3) is the VAR representation of (1) and (2). Note that the unconditional mean for the univariate and the vector representation of the series are invariant to the parameters pertaining to the dynamics of the model. Since $D_t^*\delta$ is a 2×1 vector with elements that are linear combinations of DU_{1t} and DU_{2t}, both equations of the VAR potentially have two intercepts shifts depending on the timing of the breaks and values of the dynamics parameters, a_{ij}. Therefore, the VAR described by (3) inherits the shifts in the means of the two series as shifts in the intercepts of the autoregressive equations.

A bivariate innovational outlier mean shift model can be specified as:

$$y_t = \mu + v_t \tag{4}$$

$$v_t = Av_{t-1} + e_t + D_t\delta. \tag{5}$$

By transforming the model in the usual way gives the VAR

$$y_t = \mu^* + D_t\delta + Ay_{t-1} + e_t. \tag{6}$$

Notice that the mean shifts now affect the innovation series e_{1t} and e_{2t}. Unlike the additive outlier model, the unconditional mean of y_t depends on the dynamic parameters. More importantly, each equation of the VAR is subject to a single shift in intercept whereas the univariate time series are possibly subject to multiple shifts. The opposite is true of the additive outlier model. This distinction is crucial to how mean shifts should be dealt with in a multivariate setting, an issue to which we will return.

Throughout, we assume that e_t is iid with $E(e_{1t}^2) = \sigma_1^2$, $E(e_{2t}^2) = \sigma_2^2$, $E(e_{1t}e_{2t}) = \sigma_{12}$, and $y_0 = (0,0)'$. We only consider cases where the roots of $|I_2 - AL|$ lie outside the unit circle. Therefore z_t and v_t are first order stationary bivariate time series. Both the AO and the IO model can also be used to characterize shifts in the mean of the first difference of two $I(1)$ but not cointegrated series. Cases in which the DGP has a unit root and cointegration exists are deliberately ruled out as the unit root asymptotics complicate the algebra without adding intuition to the problem at hand. We assume for now that the breaks in the two series coincide, so that $DU_{1t} = DU_{2t} = DU_t$. This assumption simplifies the algebra and analytic results while allowing us to illustrate important results. We discuss later in the context of trend removal additional issues that arise when the break dates differ.

The effects of AO mean shifts on the VAR can best be seen by considering one of the equations, say, y_{2t}, which can be written as

$$y_{2t} = \mu_2^* + \delta_2^* DU_t + \gamma_2^* DTB_t + a_{21}y_{1t-1} + a_{22}y_{2t-1} + e_{2t}, \tag{7}$$

with $\mu_2^* = (1 - a_{22})\mu_2 - a_{21}\mu_1$, $\delta_2^* = (1 - a_{22})\delta_2 - a_{21}\delta_1$, and $\gamma_2^* = a_{21}\delta_1 + a_{22}\delta_2$. Equation (7) reveals that the effects of mean breaks on y_{2t} will depend on whether or not $a_{21} = 0$. If y_1 does not Granger cause y_2, i.e. $a_{21} = 0$, y_1 is weakly exogenous for y_2. When $a_{21} = 0$, δ_2^* and γ_2^* do not depend on δ_1, and a mean break in y_1 will not appear in the equation for y_2. Under Granger non-causality, a mean break dummy will appear in the equation for y_2 only if there is a break in the mean function for y_2 itself, i.e. $\delta_2 \neq 0$. However, when y_1 Granger causes y_2, the mean shift in y_1 will, in general, appear in the conditional model for y_2 whether we have mean shifts in y_1, y_2, or both.

In the IO mean shift model we have for y_{2t}:

$$y_{2t} = \mu_2^* + \delta_2 DU_t + a_{21}y_{1t-1} + a_{22}y_{2t-1} + e_{2t}. \tag{8}$$

In this case, the intercept shift in y_{2t} is the same magnitude as the mean shift to e_{2t} and does not depend on the mean shift to e_{1t}. Thus, the IO mean shift to e_{1t} has no direct effect on the y_{2t} equation of the VAR. Conversely, the unconditional means of y_{1t} and y_{2t} are functions of both δ_1 and δ_2. This is easily seen be rewriting (4) as

$$
\begin{aligned}
y_t &= \mu + (I - AL)^{-1}D_t\delta + z_t \tag{9} \\
z_t &= Az_{t-1} + e_t.
\end{aligned}
$$

From (9) we have

$$
\begin{aligned}
y_{1t} &= \mu_1 + g_1(L)DU_t + z_{1t} \tag{10} \\
y_{2t} &= \mu_2 + g_2(L)DU_t + z_{2t}, \tag{11}
\end{aligned}
$$

where $g_1(L) = (a_{12}\delta_2 L + (1 - a_{11}L)\delta_1)/d(L)$, $g_2(L) = (a_{21}\delta_1 L + (1 - a_{22}L)\delta_2)/d(L)$ and $d(L) = (1 - a_{11}L)(1 - a_{22}L) - a_{21}a_{12}L^2$. The long run impact on the unconditional means from the IO mean shifts are given by $g_1(1)$ and $g_2(1)$. Whether or not the IO mean shift from one series filters into the unconditional mean of the other series depends on the Granger causality relationship between the series. For example, if y_1 does not Granger cause y_2, then $a_{21} = 0$ and the unconditional mean of y_{2t} is unaffected by δ_1.

Detecting a break in univariate time series is by now standard practice. There is abundant evidence for breaks in the trend function of many macroeconomic series. For example, Perron (1989) analyzed the Nelson-Plosser data set and found that many series are stationary around segmented means and/or trends. Structural change in many of these series was confirmed by Vogelsang (1997a) and Chu and White (1992) using direct tests for shifts in

5

trend. Vogelsang (1997b) found evidence of a mean shift in the unemployment rate. Series for international output were also found to have segmented trends by Banerjee, Lumsdaine and Stock (1992), Ben-David and Papell (1995) and Perron (1991). In spite of these findings, many VARs continue to be estimated as though there were no breaks in the series.

For a VAR in which the regressors are the same for each equation, Kruskal's theorem applies and each equation can be estimated consistently by OLS. The OLS estimates are $\sqrt{T}$ consistent, asymptotically normal, and (if the errors are normally distributed) efficient. But suppose one omits mean shifts in the regression model. Without loss of generality we consider the regression for equation y_2:

$$y_{2t} = \hat{\mu}_2^* + \hat{a}_{21}y_{1t-1} + \hat{a}_{22}y_{2t-1} + \hat{e}_{2t}. \tag{12}$$

Regression (12) is a misspecified model. The misspecification is obvious upon rewriting the DGP as:

$$y_{2t} = \mu_2^* + a_{21}y_{1t-1} + a_{22}y_{2t-1} + e_{2t}^*, \tag{13}$$

with $e_{2t}^* = e_{2t} + \delta_2^* DU_t + \gamma_2^* DTB_t$ in the AO model and $e_{2t}^* = e_{2t} + \delta_2 DU_t$ in the IO model. Regressions based upon (12) suffer from omitted variables bias since the regressor DU_t is excluded in the AO model when $\delta_2 \neq 0$ and/or when $\delta_1 \neq 0$ and $a_{21} \neq 0$ and in the IO model when $\delta_2 \neq 0$. It remains to be precise about the asymptotic effects of omitted breaks. The results are summarized in the following theorems.

Theorem 1 *Let $y_t = (y_{1t}, y_{2t})'$, $t = 1, \ldots, T$ be generated by the AO model (1) and (2). Let the parameters $a_2 = (a_{21}, a_{22})'$ be estimated from (12) by OLS to yield $\hat{a}_2 = (\hat{a}_{21}, \hat{a}_{22})'$. Let C be a 2×2 matrix with elements $c_{ij} = \delta_i \delta_j \lambda(1 - \lambda) + \Gamma_{ij}$ where $\Gamma_{ij} = cov(z_{it}, z_{jt})$ and $\lambda = T_B/T$ remains constant as T increases.*

A. Under Granger non-causality with $a_{21} = 0$,

1. $plim(\hat{a}_2 - a_2) = 0$ if $\delta_2 = 0$;

2. $plim(\hat{a}_2 - a_2) = C^{-1}\delta\lambda(1 - \lambda)\delta_2^*$ if $\delta_2 \neq 0$;

3. $plim(\hat{a}_{21} - a_{21}) = 0$ if $\delta_1 = 0$, $a_{12} = 0$ and $\sigma_{12} = 0$;

B. Under Granger causality with $a_{21} \neq 0$, $plim(\hat{a}_2 - a_2) = C^{-1}\delta\lambda(1 - \lambda)\delta_2^*$ if $\delta_1 \neq 0$ and/or $\delta_2 \neq 0$ unless $(1 - a_{22})\delta_2 = a_{21}\delta_1$ in which case $plim(\hat{a}_2 - a_2) = 0$.

Theorem 2 *Let $y_t = (y_{1t}, y_{2t})'$, $t = 1, \ldots, T$ be generated by the IO model (4) and (5). Let the parameters $a_2 = (a_{21}, a_{22})'$ be estimated from (12) by OLS to yield $\hat{a}_2 = (\hat{a}_{21}, \hat{a}_{22})'$. Let*

Q be a 2×2 matrix with elements $q_{ij} = g_i(1)g_j(1)\lambda(1 - \lambda) + \Gamma_{ij}$ where $\Gamma_{ij} = cov(z_{it}, z_{jt})$ and $\lambda = T_B/T$ remains constant as T increases. Define the 2×1 vector $g(1) = (g_1(1), g_2(1))'$.

   A. Under Granger non-causality with $a_{21} = 0$,

   1. $plim(\hat{a}_2 - a_2) = 0$ if $\delta_2 = 0$;

   2. $plim(\hat{a}_2 - a_2) = Q^{-1}g(1)\lambda(1 - \lambda)\delta_2$ if $\delta_2 \neq 0$;

   3. $plim(\hat{a}_{21} - a_{21}) = 0$ if $\delta_1 = 0$, $a_{12} = 0$ and $\sigma_{12} = 0$;

   B. Under Granger causality with $a_{21} \neq 0$, $plim(\hat{a}_2 - a_2) = Q^{-1}g(1)\lambda(1 - \lambda)\delta_2$ if $\delta_1 \neq 0$ and/or $\delta_2 \neq 0$ unless $g_1(1) = g_2(1) = 0$ in which case $plim(\hat{a}_2 - a_2) = 0$.

It can easily be seen that the theorems also apply when y_1 is the dependent variable. For that equation, the properties of the estimates will depend on whether $a_{12} = 0$, with the impact of the mean break measured by δ_1 and $\delta_1^* = (1 - a_{11})\delta_1 - a_{12}\delta_2$.

The theorems establish that when there are breaks in both y_1 and y_2, the estimates associated with the autoregression which omits the intercept shifts will, in general, be inconsistent. This is due to the fact that $plim(T^{-1}\sum y_{1t-1}e_{2t}^*)$ and $plim(T^{-1}\sum y_{2t-1}e_{2t}^*)$ can be nonzero. Whether or not the OLS estimates are consistent depend on these probability limits.

Consider the AO model. For this simple first-order VAR, it follows from Theorem 1 that, under Granger non-causality,

$$\begin{aligned}
\hat{a}_{21} - a_{21} &\propto (\delta_1\Gamma_{22} - \delta_2\Gamma_{12})\delta_2^*\lambda(1 - \lambda), \\
\hat{a}_{22} - a_{22} &\propto (-\delta_1\Gamma_{12} + \delta_2\Gamma_{11})\delta_2^*\lambda(1 - \lambda),
\end{aligned} \tag{14}$$

Under the assumption that $a_{21} = 0$ we have: $\Gamma_{22} = \sigma_2^2/(1 - a_{22}^2)$, $\Gamma_{11} = \sigma_1^2/(1 - a_{11}^2) + 2\sigma_{12}a_{11}a_{12}/[(1 - a_{11}^2)(1 - a_{11}a_{12})] + \sigma_2^2 a_{12}^2(1 + a_{11}a_{22})/[(1 - a_{11}^2)(1 - a_{22}^2)(1 - a_{11}a_{22})]$ and $\Gamma_{12} = \sigma_{12}/(1 - a_{11}a_{22}) + \sigma_2^2 a_{12}a_{22}/[(1 - a_{22}^2)(1 - a_{11}a_{22})]$. Since $\delta_2^* = (1 - a_{22})\delta_2$ under Granger non-causality, it will be zero when $\delta_2 = 0$ whatever the value of δ_1 might be, and leads to A1 of the theorem. In general, $\hat{a}_{21} - a_{21} = O_p(1)$ when $\delta_2 \neq 0$, as shown in A2 of the theorem. As in classical regression analysis, the least squares estimates of a misspecified model with omitted regressors are shifted from their true values by quantities determined by the correlation between the regression error and the included regressors. In the present case of omitted mean shifts, the estimates will also depend on the timing of the breaks as determined by λ.

7

There is an exception when $\delta_2 \neq 0$ and yet one of the estimated coefficients is consistent. The situation arises when $\delta_1 = 0$, $a_{12} = 0$ and $\sigma_{12} = 0$. In this case, $\hat{a}_{21}$ is consistent as suggested by A3 of Theorem 1. Formally, the result arises because $\Gamma_{12} = 0$ when $a_{12} = 0$ and $\sigma_{12} = 0$. Hence with $\delta_1 = 0$, the effect of the break on $\hat{a}_{21}$ is nil. This is a special case because with $a_{21} = a_{12} = \sigma_{12} = 0$, y_1 and y_2 are uncorrelated. Therefore, although the regression based on (12) is misspecified and has a non-zero mean in the error term, y_{1t} is uncorrelated with the regression error and is asymptotically orthogonal to y_{2t}. Hence, the misspecification does not contaminate the estimate for a_{21}. However, $\hat{a}_{22}$ is still inconsistent even if $a_{12} = 0$, and hypothesis testing based on combinations of $\hat{a}_{21}$ and $\hat{a}_{22}$ will still be invalid.

Given the definition of $\delta_2^* = (1 - a_{22})\delta_2 - a_{21}\delta_1$, the properties of the OLS estimates improve the closer is a_{22} to +1 and deteriorate the closer is a_{22} to -1. The reason, as in the univariate additive outlier case, is that as a_{22} increases, the size of any mean shift is reduced relative to the unconditional variance of y_2, as defined by $\sigma_2^2/(1 - a_{22})$. When a_{22} is unity with $a_{21} = 0$, y_2 is a random walk, and a mean shift only induces a one time outlier in the first differences of the data. On the other hand, negative serial correlation in y_{2t} reduces the unconditional variance of the series and effectively increases the relative magnitude of the break. These results also highlight the point that the intercepts of the VAR [i.e. (3)] are tied to the dynamics in the additive outlier model. One cannot tell if a shift in μ^* is due to variations in the parameters pertaining to the dynamics, or to a shift in the unconditional mean of the data. To the extent that our ultimate interest is in the impact of the mean shift holding the dynamics fixed, it is desirable to be explicit about the source of the mean shift. This makes clear that a mean shift in the AO case is specifically a break to the unconditional mean of the model.

Differences between parts A and B of the Theorem 1 formalize our earlier observation that the extent by which $\hat{a}_2$ deviates from a_2 will depend on the value of a_{21}. Under Granger causality, the least squares estimates will be inconsistent so long as there is a mean shift. Furthermore, the effects of the omitted break in y_1 will no longer be isolated to the equation for y_1 alone. The regression error for y_2 which contains the omitted break is now correlated with lags of y_1. Thus, under Granger causality, no configuration of δ and A can resurrect consistency of the least squares estimates.

The consistency results for the IO model are very similar to the results for the AO model. Interestingly, part A3 continues to hold for the IO model. This can be seen by writing the

equivalent of (14) for the IO model

$$
\begin{aligned}
\hat{a}_{21} - a_{21} &\propto (g_1(1)\Gamma_{22} - g_2(1)\Gamma_{12})\delta_2\lambda(1 - \lambda), \\
\hat{a}_{22} - a_{22} &\propto (-g_1(1)\Gamma_{12} + g_2(1)\Gamma_{11})\delta_2\lambda(1 - \lambda).
\end{aligned}
\tag{15}
$$

When $a_{12} = 0$ and $\delta_1 = 0$, it follows that $g_1(1) = 0$, and since it is still true that $\Gamma_{12} = 0$ when $a_{12} = 0$ and $\sigma_{12} = 0$, we see that $\hat{a}_{21} - a_{21}$ is proportional to zero even when $\delta_2 \neq 0$. Thus, $\hat{a}_{21}$ is a consistent estimate in spite of the omitted regressor DU_t when y_{1t} has no IO mean shift and y_1 and y_2 are asymptotically orthogonal.

There are some subtle differences between the AO and IO consistency results that are useful to highlight. First, as a_{22} approaches ± 1, it is no longer clear how the magnitude of the bias of the OLS estimates changes in the IO model. As a_{22} approaches one, it is still true that the unconditional variance of y_2 increases. But, unlike in the AO model, the magnitude of the unconditional shift in mean is not independent of a_{22} and increases as a_{22} approaches one. When $a_{22} = 1$ the mean shift becomes a trend shift and y_{2t} becomes a unit root series with a slope shift. Therefore, it is difficult to determine the net effect on bias[1]. Second, in the AO model the exception to part B of Theorem 1 arises when the mean shifts and dynamics are such that the intercept to the y_2 equation of the VAR is not affected by the mean shifts. Thus, regression (12) is not misspecified in this case. In the IO model, regression (12) is always misspecified as long as $\delta_2 \neq 0$. However, when $g_1(1) = g_2(1) = 0$, the unconditional means of both series are unaffected by the IO mean shifts as the dynamics are such that the means shifts cancel each other. In this case the regressors y_{1t-1} and y_{2t-1} and the omitted regressor DU_t are asymptotically orthogonal and OLS remains consistent.

A more general parameterization of the DGPs would allow the mean breaks to occur at possibly different dates across series, but this will not change the qualitative aspects of the theorems. Repeating the analysis, it is easy to see that omitting one of the breaks will still lead to inconsistent least squares estimates even under Granger non-causality. The discrepancy between the OLS estimates and their true value can be quite large if the two break dates are sufficiently far apart. Although it would be practical to include one break dummy for the most likely break date in a VAR, this may not be adequate when there are genuinely multiple breaks in multivariate time series. We defer the issue of appropriate detrending of VARs to Section 4.

When the models are expanded to allow for time trends with a possible break, the additive outlier DGP, for example, becomes $y_{it} = \mu_i + \beta_i t + \delta_i DU_{it} + \theta_i DT_{it} + z_{it}(i = 1, 2)$

---

[1]This illustrates a discontinuity in the functional form of the deterministic components of the series as autoregressive roots approach one in the IO model.

9

where $DT_t = 1(t > T_B)(t - T_B)$. Then y_2 takes the form

$$y_{2t} = \bar{\mu}_2^* + \bar{\delta}_2^* DU_t + \bar{\gamma}_2^* DT B_t + \bar{\beta} t + \bar{\theta}_2^* DT_t + a_{21} y_{1t-1} + a_{22} y_{2t-1} + e_{2t},$$

where $\bar{\mu}_2, \bar{\delta}_2^*, \bar{\gamma}_2^*, \bar{\theta}_2^*$ are functions of μ, δ, and θ. Regressions of the form

$$y_{2t} = \hat{\mu}_2^* + \hat{\beta}_2^* t + a_{21} y_{1t-1} + a_{22} y_{2t-1} + e_{2t},$$

are misspecified because of the omitted shifts in mean and/or the trend. When there is an omitted mean shift (but not in the trend), the least squares estimates have properties as stated in Theorem 1. When there is an omitted shift in the trend function, consistency of the estimates in the y_2 equation still depend on the presence or absence of Granger causality. By analogy to the preceding analysis, we now have $\bar{\theta}_2^* = (1 - a_{22}\theta_2 - a_{21}\theta_1)$ which controls the impact of the omitted trend shifts on the regression model. When both series have a slope shift, y_{1t-1} and y_{2t-1} are asymptotically collinear. Using transformations similar to those suggested by Sims, Stock and Watson (1990), it is possible to establish the inconsistency of of $\hat{a}_2$. In fact, a theorem nearly identical to Theorem 1 holds when there are slope shifts. Details are available upon request. Finally, when z_{1t} and/or z_{2t} are $I(1)$ and are not cointegrated, Theorem 1 obviously applies to Δz_t. Similar developments apply to the IO model.

Since Theorems 1 and 2 establish that the least squares estimates are inconsistent, functions of these estimates will also be inconsistent. The consequence from a practical point of view is that the conditional forecasts and impulse response functions based upon the OLS estimates will also be inconsistent. This easily can be seen by forming the companion matrix (B) for the vector process $y(t) = (y_{1t}, y_{1t-1}, y_{2t}, y_{2t-1})'$, with $y(t) = By(t-1) + E(t)$. Since $\hat{B}$ depends on $\hat{a}_{ij}$, $\hat{B}$ is inconsistent, from which it follows that $\hat{B}^i$ is also inconsistent, where i is the horizon of the forecast.

2.1 Finite Sample Properties of the Estimates in the AO Model

In this subsection we use a simulation study to assess the empirical implications of the consistency theorems. For brevity, we focus on the AO model and Theorem 1. Similar results hold for the IO model . We generated series according to (1) and (2) using $T = 500$ and $T_B = 250$ with e_t *iid* draws from a standard bivariate normal distribution. Since the variances of the e_{it}'s are unity, the magnitude of the mean shift is measured in terms of the standard deviations of the e_{it}'s. Four possible combinations of (δ_1, δ_2) were considered: $(0,0)$, $(1,0), (0,1), (1,1)$. This allows for the possibility that a break occurs in none, one, or both series. For the base case, we considered three values of a_{11} (-0.5,0.0,0.5) and three

values for a_{21} (-0.3, 0.0, 0.3). Values for a_{12} and a_{22} are taken from the parameter set $-0.6, -0.4, -0.2, 0.0, 0.2, 0.4, 0.6$ for each value of a_{11}. All parameterizations have eigenvalues of A that lie inside the unit circle with unequal roots. This rules out explosive models, and models whose moment matrix of regressors are singular. In all cases, 1000 replications were used. For each replication, (12) was estimated by OLS. For analysis in a subsequent section, we focus on the properties of $\hat{a}_{21}$. Reported are the averages of the estimates over simulations.

When there is no break in either series, the coefficients are precisely estimated. When $\delta' = (1,0)$, the estimates are still very close to the true value in accordance with A1 of the theorem. These results are not reported to conserve space. When $\delta' = (0,1)$ there are significant deviations between $\hat{a}_{21}$ from its true value of zero. A synopsis of the results is given in Table 1a for $a_{21} = 0$, i.e. Granger non-causality is assumed. There are several features of note. The properties of the estimates are sensitive to the true value of a_{12}, even though the omitted variable is not a function of a_{12}. From (14), it can be seen that when $\delta_1 = 0$, the direction of the bias is completely determined by the size of $-\Gamma_{12}$. Accordingly, the bias is negative when $a_{12}a_{22} > 0$ and positive otherwise. When $a_{12} = 0$, the two series are independent. The bias is small as predicted by A3 of Theorem 1.

The errors in estimating $\hat{a}_{21}$ are orders of magnitude larger when $\delta' = (1,1)$ than when $\delta' = (0,1)$. In such cases, a_{12} affects $\hat{a}_{21}$ in a complex way. From (14), it is clear that for breaks of the same sign, $\hat{a}_{21} - a_{21}$ will be larger when Γ_{12} is negative. Although for a given a_{12}, $\hat{a}_{21} - a_{21}$ is smaller when a_{22} is closer to one, $\hat{a}_{21}$ is still far from its true value. An average estimate for $\hat{a}_{21}$ of 0.10 is common, and could be as large as 0.22 when the true value is zero.

When y_1 Granger causes y_2 and hence that $a_{21} \neq 0$, mean shifts in y_1 and/or y_2 will always affect the estimates for the VAR because of feedback between the two series. A representative set of results for $a_{21} = 0.3$ is given in Table 1b and for $a_{21} = -0.3$ in Table 1c. Recall that in the Granger non-causality case, $\hat{a}_{21}$ was precisely estimated when only y_1 has a mean shift. Under Granger causality, even $\hat{a}_{21}$ is imprecisely estimated under these conditions. In general, $\hat{a}_{21}$ tends to be over-estimated when $a_{21} > 0$ and underestimated (in absolute terms) when $a_{21} < 0$ under Granger causality. It is also interesting to note that the biases are generally larger when $a_{21} = -0.3$ than when $a_{21} = 0.3$. This is because $\hat{a}_{21} - a_{21}$ is proportional to $\delta_2^* = (1 - a_{22})\delta_2 - a_{21}\delta_1$, and a negative a_{21} magnifies δ_2^*. This, of course, is a consequence of the fact that δ_1 and δ_2 are both positive in the simulation experiments. However, the point to be made is that the least squares estimates will be less affected by the omission of mean shifts if these shifts can be offset in the sense of reducing δ_2^*. In the

extreme when $(1 - a_{22})\delta_2 = a_{21}\delta_1$, the estimates are unaffected by the omission of mean shifts as given by Part B of Theorem 1. This special case corresponds to a restriction on the dynamics and magnitudes of the mean shifts. Engle and Kozicki (1993) have analyzed the presence of co-features in stationary models. Our analysis provides an example of such a phenomenon.

The conclusion to be drawn from the preceding analysis is that the direction and magnitude of $\hat{a}_2 - a_2$ depends in important ways on the presence or absence of Granger causality. Omitting the break dummies in the regression when there are mean shifts in the data will generally lead to OLS estimates that deviate from the true value by 10 percent, if not more. The poor properties of the least squares estimates will likely affect inferences. We now turn to these issues.

3   Properties of the F tests for Granger Causality and Lag Length Selection

This section consists of three parts. The first examines the Granger causality test when the lag length of the regression model corresponds to the true order of the autoregression. The next subsection examines the properties of the test when the regression is over-parameterized. The third subsection considers the the choice of the lag length using the AIC. Again, we focus on the additive outlier DGP with similar results holding for the innovational outlier DGP. Results for Granger causality tests based on the innovational outlier DGP can be found in Bianchi (1995) and Lutkepohl (1989).

3.1   Granger Causality Tests at the Correct Lag Length

Suppose we are interested in testing whether y_1 Granger causes y_2, but we omit the mean shift in one or both series. Assuming that one knows the correct lag length corresponding to the DGP given by (1) and (2), the statistic is simply the F-test for testing the hypothesis that $a_{21} = 0$ in (12). More precisely, we have

$$GC = \frac{(\hat{a}_{21} - a_{21})^2}{s^2 (\widetilde{X}'\widetilde{X})_{11}^{-1}} = \frac{T(\hat{a}_{21} - a_{21})^2}{s^2 (T^{-1}\widetilde{X}'\widetilde{X})_{11}^{-1}}, \tag{16}$$

where X is the $(T-1)\times 2$ matrix of values of $X_t = (y_{1t-1}, y_{2t-1})$ and for a variable x, $\tilde{x}$ denotes the demeaned variable. Under the null hypothesis of no Granger causality, $a_{21} = 0$ and the test statistic asymptotically has a χ_1^2 distribution in the absence of mean shifts. When $\delta_2 \neq 0$ and the mean shift is omitted, $\hat{a}_{21}$ is not consistently estimated and $T(\hat{a}_{21} - a_{21})^2$ diverges

to ∞. Since, $s^2 = T^{-1}\hat{e}_2'\hat{e}_2$ and $T^{-1}(\widetilde{X}'\widetilde{X})_{11}^{-1}$ are $O_p(1)$ for any δ_2, the GC statistic diverges to ∞ when $\delta_2 \neq 0$. We have the following formal result.

Theorem 3 *Suppose the data is generated under the AO model (1) and (2), and (12) is used to test whether y_1 Granger Causes y_2. Then under the null hypothesis of no Granger Causality $(a_{21} = 0)$ as $T \to \infty$,*

1. *If $\delta_2 = 0$ then $GC \Rightarrow \chi_1^2$;*

2. *If $\delta_2 \neq 0$ then $T^{-1}GC = O_p(1) > 0$, unless $\delta_1 = 0$, $a_{12} = 0$ and $\sigma_{12} = 0$ in which case $GC = O_p(1)$;*

If the mean shift is to y_1 alone, then $\delta_2^* = 0$ and regression (12) is correctly specified. That y_1 has a shift in mean is immaterial for the regression and standard OLS results apply asymptotically. Therefore GC has the usual χ_1^2 limiting distribution. When $\delta_2 \neq 0$, the statistic for testing Granger non-causality has properties that can affect inference in important ways. Since GC diverges as T increases, the size of the test is unity asymptotically. In finite samples we should expect size to be biased upwards. Therefore, GC tends to reject non-causality even when y_1 does not Granger cause y_2. The exception is when y_1 and y_2 are independent because $\hat{a}_{21}$ is consistent for a_{21}. Although Theorem 3 is stated in terms of the AO model, similar conclusions can be drawn for the IO model.

We now use simulations to assess the implications of Theorem 3. Since we are interested in size distortions of the GC test, we set $a_{21} = 0$ in all cases. For $\delta' = (0,0)$ and $(1,0)$, the GC test has an exact size close to the nominal size of 5% following part 1 of Theorem 3. These results are shown in the top panels of Figure 1 which graph the size of the test with $a_{11} = 0$ for various values of a_{12} and a_{22} (qualitatively similar results hold for $a_{11} \neq 0$). Of interest are the two cases with $\delta_2 \neq 0$. The results are presented in the lower panels of Figure 1. When $\delta_1 = 0$ but $\delta_2 = 1$, size distortions are larger the closer is $|a_{12}|$ to unity and/or a_{22} to -1, and are smaller the closer is a_{22} to unity. When $\delta' = (1,1)$, the size distortions are magnified relative to the $\delta' = (0,1)$ case. When the autoregressive coefficients are negative, there can be a rejection of Granger non-causality with probability one even if there is no causal relationship in the data.

The issue of how shifts in mean affect the F-test for Granger causality have also been studied by Lutkepohl (1989) and Bianchi (1995) using simulations. Although not explicit, both studies used the innovational outlier model as the data generating process and hence the unconditional mean of each series depends on the dynamic parameters of the model.

13

Since the simulations presented in this analysis are based on an additive outlier framework, the unconditional mean is invariant to the $a'_{ij}s$. Nonetheless, all results show that when the regression model is misspecified in the sense of omitting the break dummy, we will be misled into finding evidence of Granger causality too often, and size distortions are larger when a_{22} is negative than when it is positive. Our analysis shows that these size distortions are due to inconsistency of the OLS estimates, and we make precise how the dynamics parameters interact with the least squares bias.

3.2 Granger Causality Tests using Higher Order Autoregressions

We have thus far assumed that the true order of the autoregression (i.e. one) is known. In practice, this information is rarely available, and the lag length is usually selected according to criteria such as the Akaike Information Criteria (AIC). There are two issues that need to be addressed. The first is the performance of these information based criteria in the case of a misspecified model. The second is the implications for the Granger causality test when the order of the autoregression does not coincide with the true model. We begin with the latter.

Because the properties of the Granger causality F test depend on the consistency or lack thereof of the least squares estimates, it is instructive to first compare the properties of the estimates from a kth order autoregression versus the estimates derived from a lower order model. Let $W_{1t} = (y_{1t-1}, y_{2t-1}, \ldots, y_{1t-k+1}, y_{2t-k+1})$, and $W_{2t} = (y_{1t-k}, y_{2t-k})$ for $k \geq 2$. Hence the DGP is

$$y_{2t} = \mu_2^* + \delta_2^* DU_t + \gamma_2 DTB_t + W_{1t}\beta_1 + W_{2t}\beta_2 + e_{2t}, \tag{17}$$

with $\beta_1 = (a_{21}, a_{22}, 0_{1 \times 2(k-2)})'$, and $\beta_2 = (0, 0)'$.

The regression model is:

$$\begin{aligned} y_{2t} &= \hat{\mu}_2^* + \sum_{i=1}^{k} \hat{a}_{21}(i)y_{1t-i} + \sum_{i=1}^{k} \hat{a}_{22}(i)y_{2t-i} + \hat{e}_{2t} \\ &= \hat{\mu}_2^* + W_{1t}\hat{\beta}_1 + W_{2t}\hat{\beta}_2 + \hat{e}_{2t}. \end{aligned} \tag{18}$$

Let W_1, W_2 and Y_2 denote the $(T-k) \times (2k-2)$, $(T-k) \times 2$ and $(T-k) \times 1$ matrices of values of W_{1t}, W_{2t} and y_{2t}. By partitioned regression, $\hat{\beta}_2 = (\widetilde{W}_2' M_1 \widetilde{W}_2)^{-1} \widetilde{W}_2' M_1 \widetilde{y}_2$ and $\hat{\beta}_2 - \beta_2 = (\widetilde{W}_2' M_1 \widetilde{W}_2)^{-1} \widetilde{W}_2' M_1 e_2^*$, where $e_{2t}^* = e_{2t} + \delta_2^* DU_t + \gamma_2 DTB_t$, $M_1 = I_T - \widetilde{W}_1(\widetilde{W}_1'\widetilde{W}_1)^{-1}\widetilde{W}_1'$. Since the elements of $\widetilde{W}_2$ inherit properties of the mean shifts, it is correlated with e_2^*. Hence $\hat{\beta}_2$ will deviate from its population value of zero.

It is well-known that including irrelevant regressors in a well-specified model will lead to inefficient but unbiased estimates. However, the present setting of including more lags

14

than the true model is non-standard because the model with the correct number of lags is misspecified. The elements of W_2 are no longer irrelevant in the sense that each lagged regressor is correlated with the omitted mean shift. The additional lags will unambiguously increase the variance of the estimates, though it could reduce the bias depending on the sign of the partial correlation between e_2^* and W_2 conditional on W_1. The latter in turn depends on whether one or both series have breaks, and on the dynamics of the model. The overall effect of including additional lags on the properties of the estimates is ambiguous a priori.

Simulations reveal that the point estimate for $a_{21}(1)$ tend to be more precise as k increases. The reduction in bias is monotonic in k when $a_{12} > 0$, and non-monotonic when $a_{12} < 0$. The results are reported in Table 2a and 2b for $a_{12} = 0.4$ and -0.4 respectively. For example, when $a_{12} = 0.4$, the average point estimate for $\hat{a}_{21}$ at lag one is 0.21 at $k = 1$, and improved from 0.1872 at $k = 2$ to 0.1069 at $k = 4$ and eventually to 0.0617 at $k = 8$. The intuition is that each additional set of lags contain relevant information which help purge the correlation between the regression error and the included regressors. When $a_{12} < 0$, a small increase in k could actually make the estimates less precise, but with a sufficiently large enough increase in k, an improvement in the precision of $\hat{a}_{21}$ is still possible.

In hypothesis testing such as testing for Granger causality, it is the properties of a set of estimates that matter. Although the point estimates are more precise in a kth order than in a $k - 1^{th}$ order autoregression, the number of inconsistent coefficients also increases as the lag length increases. Not surprisingly, the desirability of increasing k in Granger causality tests critically depend on whether increasing the number of lags can reduce the bias of the estimates in the autoregression by enough to offset the increase in variability in the estimates induced by the additional lags. More formally, the Granger Causality test is now the F-test for the joint hypothesis $\hat{a}_{21}(1) = \hat{a}_{21}(2) \ldots = \hat{a}_{21}(k) = 0$. It is straightforward to show that the GC test can be written as

$$GC = \frac{\widetilde{Y}_2' M_1 \widetilde{W}_2 (\widetilde{W}_2 M_1 \widetilde{W}_2)^{-1} \widetilde{W}_2' M_1 \widetilde{Y}_2}{k s_k^2}, \tag{19}$$

where $s_k^2 = T^{-1} \sum_{t=1}^{T} \hat{e}_{2t}^2$. It is clear that k has two effects on the test: it influences the bias of the estimates and the variability in the moment matrix of regressors. More precisely, we can show that the results of Theorem 3 continue to hold when $k > 1$ (we do not state a formal theorem for brevity). In the case where $\delta_2 \neq 0$, though, it is useful to make results of Theorem 3 more explicit. Let ι be a $k \times 1$ vector of ones. Define the $k \times k$ matrix $\Gamma_{ij}(k) = E[(z_{it-1}, z_{it-2}, \ldots, z_{it-k})'(z_{jt-1}, z_{jt-2}, \ldots, z_{jt-k})]$. As long as $a_{21} \neq 0$ it can be shown

that:

$$plim T^{-1} GC = \frac{1}{k} \frac{K'L^{-1}K}{(\sigma_2^2/\delta_2^{*2})(M - K'L^{-1}K)} > 0, \qquad (20)$$

where $K = \delta_1 \lambda (1 - \lambda) \iota - [\delta_1 \delta_2 \lambda (1 - \lambda) \iota\iota' + \Gamma_{21}(k)'][\delta_2^2 \lambda (1 - \lambda) \iota\iota' + \Gamma_{22}(k)]^{-1} \iota \lambda (1 - \lambda) \delta_2$,

$\quad L = \delta_1^2 \lambda (1 - \lambda) \iota\iota' - [\delta_1 \delta_2 \lambda (1 - \lambda) \iota\iota' + \Gamma_{21}(k)'][\delta_2^2 \lambda (1 - \lambda) \iota\iota' + \Gamma_{22}(k)]^{-1} [\delta_1 \delta_2 \lambda (1 - \lambda) \iota\iota' + \Gamma_{21}(k)]$,

$\quad M = \lambda (1 - \lambda) - \delta_2^2 \lambda^2 (1 - \lambda)^2 \iota'[\delta_2^2 \lambda (1 - \lambda) \iota\iota' + \Gamma_{22}(k)]^{-1} \iota$.

Clearly, increasing k has a non-linear effect on GC asymptotically. The interesting question is whether the additional regressors reduce or magnify finite sample size distortions of the GC statistic. The simulation results are reported in Tables 3a and 3b for $a_{12} = 0.4$ and -0.4 respectively for $k = 2,4,8$. When the break is to y_2 alone, increasing the lag length can aggravate size distortions with a small increase in k but eventually improve the size as k increases further. However, when there are breaks to both series, the size improvement from increasing k can be quite dramatic. For example, with $a_{22} = 0$ and $a_{12} = .4$, size distortions are reduced from .86 at one lag to .33 at four lags and to .12 at eight lags for a nominal size of 5%. These results essentially mirror those for $\hat{a}_{21}(1)$ reported in Table 2, showing that when the point estimates for $a_{21}(L)$ can be made more precise by increasing k, the size of the F test will also improve. The lesson to be learned from these results is that if the potential for misspecification of a VAR from omitted shifts in means remains (such as if some, but not all, shifts in mean have been identified), then VARs with long lag lengths tend to reduce potential size distortions of Granger causality tests compared to more parsimonious lag lengths.

3.3 Lag Length Selection in the Presence of Mean Shifts

In the standard framework when the regression model is assumed to be correctly specified, it has been shown that information based model selection criteria such as the AIC will choose an order, k, that is no smaller than the true order, k^*, at least asymptotically [see, for example, Lutkepohl (1993)]. To understand the implications of omitted mean shifts for model selection, consider the case of testing whether y_1 Granger causes y_2. Let $\hat{\sigma}_2^2(k)$ be the conditional MLE estimate of σ_2^2 in a regression of y_2 on k lags of y_1 and y_2 (and a constant) assuming Gaussian errors. The AIC for this regression is given by $AIC(k) = log(\hat{\sigma}_2^2(k)) + 2k/T$. When there are omitted mean shifts in the series, it is easy to establish that $plim AIC(k)$ is decreasing in k. This leads to the following observation: *If AIC is used to select $k \in [1, kmax]$ when testing whether y_1 Granger causes y_2, AIC will choose kmax asymptotically.*

16

To show that $plim AIC(k)$ is decreasing in k it is sufficient to show that $plim[AIC(k) - AIC(k-1)] < 0$. First note that $AIC(k) - AIC(k-1) = log[\hat{\sigma}_2^2(k)/\hat{\sigma}_2^2(k-1)] + 2/T$. It is straightforward to show that $plim(\hat{\sigma}_2^2(k-1)) > plim(\hat{\sigma}_2^2(k)) > 0$ when there are omitted mean shifts since $\hat{\sigma}_2^2(k)$ is decreasing in k. This implies that $plim(\hat{\sigma}_2^2(k)/\hat{\sigma}_2^2(k-1)) < 1$. Therefore, $log[plim(\hat{\sigma}_2^2(k)/\hat{\sigma}_2^2(k-1))] < 0$ which gives the result since $plim(2/T) = 0$. Note that the Schwarz (1978) criterion will also choose $kmax$ asymptotically since $plim(log(T)/T) = 0$.

Intuitively, the omitted mean shifts cause the kth lagged variables (which depend on the shifts) to be correlated with the implied regression error (which also depends on the shifts) conditional on the other $k-1$ lagged variables, and this correlation persists as T increases. This in turn results in $\hat{\sigma}_2^2(k-1) > \hat{\sigma}_2^2(k)$ asymptotically. Therefore, $log(\hat{\sigma}_2^2(k))$ is decreasing in k and dominates the penalty term $2k/T$ in large samples. (In a correctly specified model $plim\hat{\sigma}_2^2(k-1) = plim\hat{\sigma}_2^2(k)$ and $log(\hat{\sigma}_2^2(k))$ is not decreasing in k in the limit.) In finite samples, even a small correlation between the kth lag and the omitted break will make the AIC choose a k over $k-1$. Of course, when working with a VAR, k should be chosen to be optimal for the system as a whole. The criterion is based on the determinant of the estimated error variance matrix. However, the problem remains that when there is a genuine correlation between the kth lagged variables and the regression error conditional on the $k-1$ lags, the AIC will tend to over-parameterize the model.

In finite sample the frequency with which the AIC chooses $k \in [1, kmax]$ for $kmax = 4$ is reported in the last four columns of Table 4 for selected parameter values. Clearly, the presence of a break increases the probability that the AIC will choose a higher k. In unreported results for $kmax=8$, the probability mass is piled up at eight, leaving the probability of choosing six or less to nil.

The first column of Table 4 reports the size of the Granger causality test when k is set to one (the true order of the autoregression). The next two columns give the size of the test and when k is chosen according to the AIC with $kmax$ set to four and eight respectively. When only y_1 has a mean shift, using the AIC to select k leads to a test that is slightly oversized. When the break is to y_2 only, the distortions are smaller the larger is k, as we saw in Table 3, or the larger is $kmax$ as Table 4 indicates. Although the data dependent rule selects $k = kmax$ frequently, it is sometimes selected to be below $kmax$, giving size distortions that are larger on average than fixing k at $kmax$. When there are breaks in both series, using a data dependent rule to select k generally yields a better size for the Granger causality test than when the $k = 1$. Indeed, when size distortions from omitting the mean shifts are severe as in the $a_{22} = -0.2$ case, overparameterizing the regression model can yield

substantial reduction in size distortions.

4 Treatment of Mean Shifts in a VAR in Practice

Structural shifts in a multivariate setting have been dealt with in a number of ways in the empirical literature. For example, Canova (1995) performed residual diagnostics to verify that the recursive residuals of the VAR are stable over time but he found evidence of deviations from normality in one of the series. He therefore included a dummy for 1974:1 in the VAR. Because the variables in Canova's VAR are I(1), this corresponds to $a_{22} \to 1$ in our framework. However, in this case of a mean shift in the level of a non-stationary series, the one time outlier would not have affected consistency of the parameter estimates even if the outlier was left unattended. Other researchers often check for sub-sample parameter instability by running the VAR for different samples. See, for example, Stock and Watson (1989). This is an extreme form of a break as it allows all parameters in the VAR to take a discrete shift after a certain period. In Blanchard and Quah (1989) and Gambe and Joutz (1993), a mean shift in output growth in 1973 was suspected, and the break was removed from the series prior to estimating the VAR. These studies report results with the breaks removed from the series one by one along with results which assumed breaks do not exist.

How should mean shifts be dealt with in a VAR? Suppose we adopt the additive outlier model as the DGP but the mean shifts in the two series do not coincide. That is, $DU_{1t} \neq DU_{2t}$. From (1) and (2) we have that y_{1t} and y_{2t} are simple regression models with stationary but serially correlated errors. Since z_t follows a bivariate VAR(1) process it is easy to show that z_{1t} and z_{2t} can written as stationary $ARMA(2,1)$ processes.[2] Because z_{1t} and z_{2t} are stationary ARMA processes it follows from the classic results of Grenander and Rosenblatt (1957) that if the break dates are known a priori, the OLS estimates equation by equation of μ_1, μ_2, δ_1 and δ_2 are asymptotically equivalent to GLS and are optimal. Therefore, the mean shifts can be removed by OLS and the VAR analysis applied to the residuals.

Of course, in practice it is often unknown whether individual series have stable means. A first step in the VAR analysis would be to first test the individual series for stable means using one or more of the tests available in the literature e.g. Andrews (1993), Andrews

---

[2]For example, we have $z_{2t} = (a_{11} + a_{22})z_{2t-1} + (a_{12}a_{21} - a_{11}a_{22})z_{2t-2} + \eta_t + \phi\eta_{t-1}$ where η_t is $i.i.d.$ with $var(\eta_t) = (\sigma_2^2(1 + a_{11}^2) - 2\sigma_{12}a_{11}a_{21} + \sigma_1^2 a_{21}^2)/(1 + \phi^2)$ and ϕ solves $\sigma_2^2 a_{11}\phi^2 + (\sigma_2^2(1 + a_{11}^2) - 2\sigma_{12}a_{11}a_{21} + \sigma_1^2 a_{21}^2)\phi + a_{11}\sigma_2^2 = 0$.

and Ploberger (1994), Perron (1991), Vogelsang (1997a) or Vogelsang (1997b). If there is evidence of a shift in mean, then the date of the shift can be consistently estimated by least squares following Bai (1994). Using the estimated break dates, dummy variables can be formed and the series demeaned. Because $y_{it} = \mu_i + \delta_i DU_{it} + z_{it}$, there is only one mean shift in the univariate representation of each series. Hence, techniques which remove one break are generally adequate unless multiple shifts are suspected in which case procedures of Bai and Perron (1995) would be required.

An alternate approach to demeaning series by series would be to work directly with the VAR and test for stability of the intercepts in each equation. At first glance, this would appear to be a similar approach and perhaps equivalent asymptotically. If each series has at most a single shift in mean and the break dates are all the same, then working directly with the VAR is asymptotically equivalent to demeaning series by series. However, if the break dates are different, working directly with the VAR is potentially much more complicated. Recall equation (3) where the intercepts in the VAR were given by $\mu^* + D_t^* \delta$. The intercept for the y_{1t} equation is given by $(1 - a_{11})\mu_1 - a_{12}\mu_2 + (1 - a_{11})\delta_1 DU_{1t} - a_{12}\delta_2 DU_{2t}$ and analogously for the y_{2t} equation. Therefore, depending on the values of the dynamic parameters in the A matrix, both equations have multiple intercept shifts. In a VAR with n equations there is the possibility of each equation having n intercept shifts. So, if the VAR is analyzed directly, the techniques of Bai and Perron (1995) would be required to identify the multiple intercept shifts whereas single shifts in means need only be identified series by series. Naturally, if the individual series were subject to multiple mean shifts, then the intercepts in the VAR would be subject to additional intercept shifts. In the end, a simpler analysis is obtained if the series are first demeaned before estimating the VAR.

It is important to keep in mind that the above recommendations are contingent upon the model being of the additive outlier form. The additive outlier approach implicitly models the mean shifts as affecting each series individually independent of the correlation between series. This is in contrast to the innovative outlier approach in which it is the means of the innovations that are subject to shifts. A closer examination of IO model reveals that a shift in the mean of e_{1t} initially only affects the mean of y_{1t}. Over time, the shift will also affect the mean of y_{2t} unless y_1 does not granger cause y_2. Therefore, in general, the means of y_{1t} and y_{2t} will have two shifts each if the dates of the mean shifts to e_{1t} and e_{2t} occur at different dates. In this case direct analysis of the VAR requires identification of single shifts in intercepts equation by equation whereas analysis of the individual time series in y_t requires identification of potential multiple shifts. Thus, analysis of the VAR directly is simpler than

analyzing the individual series if the structural change is the result of innovational mean shifts.

In practice applied researchers should be aware of the model underlying the intercept shifts (and other shifts in trend) in their VARs because the methods used to deal with these shifts do depend on what is the model being assumed. Removing mean shifts series by series should be based on the additive outlier approach. Placing intercept dummy variables in a VAR should be based on the innovational mean shift approach. When the break dates are the same in each equation, the two approaches are for practical purposes identical and differ only with respect to the the one-time dummy variables, DTB_t that appear. Asymptotically there will be no difference between the two approaches in terms of estimation and hypothesis testing. But, when the break dates are different across series, the underlying model of the mean shifts is important since using an inappropriate method of removing the shifts may lead to problems if the inherent multiplicity of shifts are not fully identified in each series or equation.

5 An Empirical Example

Many researchers have found evidence for shifts in the mean and/or the trend function in macroeconomic time series. As an example, we consider a three-variable VAR. The data are quarterly output growth for Germany, France, and Italy, analyzed in a recent paper by Bai, Lumsdaine and Stock (1997). The authors reported that for the trivariate system, a shift in mean in the growth rate of GDP in 1973:4 was estimated with a tight confidence band.[3] We estimated a three variable VAR of demeaned variables with four and eight lags. We call this Model A. We then removed the mean shift from the data. Because the break date is identical across countries, we took the residuals from a projection of output growth, country by country, on a constant, and $DUM734 = 1(t > 73 : 4)$. The residuals were then used to form a VAR, which we refer to as Model B. The p-values for the Granger causality tests are as follows:

---

[3]The data for France includes observations for the 1968 strike. A dummy for 68:2 and one for 68:3 are included to control for these outliers.

Granger Non-Causality Tests for the Output Growth of Germany, France, and Italy

| | Four Lags | | Eight Lags | |
Null Hypothesis	Model A	Model B	Model A	Model B
Germany does not GC France	0.5840	0.6889	0.6814	0.7474
France does not GC Germany	3.7E-5	0.0001	0.0007	0.0824
Italy does not GC France	0.0018	0.0782	0.0068	0.2264
France does not GC Italy	0.0637	0.1316	0.3822	0.3967
Italy does not GC Germany	0.0158	0.2872	0.0041	0.1209
Germany does not GC Italy	0.4814	0.5863	0.7485	0.8695

Note that for a given lag, the p values are uniformly larger when breaks are removed from the data. Had the break been omitted, we would have falsely concluded that Italy Granger causes Germany at the one percent level, and that France and Italy Granger cause each other at the ten percent level. When the breaks are included, such relationships either do not exist or are much weaker. Quantitative differences in the results are also reflected in the decomposition of variances for the two models. In Model A, France and Italy accounted for seven and two percent of the variance in Germany's output growth and stabilized at those levels after 5 periods. In Model B, these statistics rose to seventeen and six respectively. More dramatic are the differences in the decomposition of Italy's output growth. Without the mean shifts, France is found to account for twenty-eight percent of Italy's variance in output growth, and Germany accounts for eight. Once the mean shifts are taken into account, the numbers dropped to six and one respectively. Italy's output therefore appears considerably more exogenous in Model B.

6 Conclusion

This paper has examined the implications of omitted mean shifts in the context of a bivariate VAR. It was shown that the least squares estimates of the VAR are inconsistent, and hypothesis testing using those estimates can lead to incorrect inference. The results extend to omitted shifts in the trend function. Although we have focused on a two-variable model for analytical convenience, qualitatively similar results are expected to hold in higher dimension models. An interesting result of our study is that when the model is misspecified in the sense of omitted mean shifts, information criteria such as the AIC will tend to over-parameterize the model. Including more lags to the misspecified model can alleviate size distortions insofar as the additional lags reduce the significance of the omitted mean shift in the regression error. In simulations, this situation occurs when both series have shifts in the same direction. We also discussed how to remove the breaks in a multivariate setting. Under the additive

outlier specification, the intercept in each equation of the VAR will be subject to multiple shifts when the break dates of the mean shifts to the univariate series do not coincide. Under the innovative outlier specification, the unconditional means of the univariate time series are subject to multiple shifts when mean shifts to the innovation processes occur at different dates. Therefore, techniques designed to detect multiple shifts may be required when break dates do not coincide.

Proof of Theorem 1

Let $\widetilde{Y}_2$, $\widetilde{DU}$, $\widetilde{DTB}$, and $\widetilde{e}_2$ be $(T-1) \times 1$ vectors of demeaned y_{2t}, DU_t, DTB_t, and e_{2t} respectively. Let $\widetilde{X}$ be the $(T-1) \times 2$ matrix of demeaned (y_{1t-1}, y_{2t-1}). The model given by (7) can be written as

$$\widetilde{Y}_2 = \widetilde{DU}\delta_2^* + \widetilde{DTB}\gamma_2^* + \widetilde{X}a_2 + \widetilde{e}_2. \tag{A1}$$

Using regression (12) and plugging in for $\widetilde{Y}_2$ using (A1), we have

$$
\begin{aligned}
\hat{a}_2 &= (\widetilde{X}'\widetilde{X})^{-1}\widetilde{X}'\widetilde{Y}_2 \\
&= a_2 + (\widetilde{X}'\widetilde{X})^{-1}\widetilde{X}'(\widetilde{DU}\delta_2^* + \widetilde{DTB}\gamma_2^* + \widetilde{e}_2)
\end{aligned}
$$

We now consider the following limiting results:

$$
T^{-1}\widetilde{X}'\widetilde{X} = \begin{bmatrix} T^{-1}\widetilde{y}_{1t-1}^2 & T^{-1}\sum \widetilde{y}_{1t-1}\widetilde{y}_{2t-1} \\ T^{-1}\sum \widetilde{y}_{1t-1}\widetilde{y}_{2t-1} & T^{-1}\sum \widetilde{y}_{2t-1}^2 \end{bmatrix},
$$

where all summations are from 1 to T. From (1) we have $\widetilde{y}_{1t-1} = \delta_1\widetilde{DU}_{t-1} + \widetilde{z}_{1t-1}$ and $\widetilde{y}_{2t-1} = \delta_2\widetilde{DU}_{t-1} + \widetilde{z}_{2t-1}$, and it follows directly that

$$
\begin{aligned}
T^{-1}\sum \widetilde{y}_{1t-1}^2 &= T^{-1}\sum (\delta_1\widetilde{DU}_{t-1} + \widetilde{z}_{1t-1})^2 \\
&= \delta_1^2 T^{-1}\sum \widetilde{DU}_{t-1}^2 + T^{-1}\sum \widetilde{z}_{1t-1}^2 + o_p(1) \\
&\xrightarrow{p} \delta_1^2\lambda(1-\lambda) + \Gamma_{11}.
\end{aligned}
$$

Likewise we have $T^{-1}\sum \widetilde{y}_{2t-1}^2 \xrightarrow{p} \delta_2^2\lambda(1-\lambda) + \Gamma_{22}$ and $T^{-1}\sum \widetilde{y}_{1t-1}\widetilde{y}_{2t-1} \xrightarrow{p} \delta_1\delta_2\lambda(1-\lambda) + \Gamma_{12}$. This establishes that $T^{-1}\widetilde{X}'X \xrightarrow{p} C$. Since C is positive-definite for all values of δ_1 and δ_2, it follows that $(T^{-1}\widetilde{X}'\widetilde{X})^{-1} \xrightarrow{p} C^{-1}$. Furthermore,

$$
T^{-1}\widetilde{X}'\widetilde{DU}\delta_2^* = \begin{bmatrix} T^{-1}\sum \widetilde{y}_{1t-1}\widetilde{DU}_t\delta_2^* \\ T^{-1}\sum \widetilde{y}_{2t-1}\widetilde{DU}_t\delta_2^* \end{bmatrix} \xrightarrow{p} \begin{bmatrix} \delta_1\lambda(1-\lambda)\delta_2^* \\ \delta_2\lambda(1-\lambda)\delta_2^* \end{bmatrix} = \delta\lambda(1-\lambda)\delta_2^*.
$$

Finally, $T^{-1}\widetilde{X}'(\widetilde{DTB}\gamma_2^* + \widetilde{e}_2) = o_p(1)$. Using these convergence results it directly follows that

$$
\begin{aligned}
plim\,(\hat{a}_2 - a_2) &= plim[(T^{-1}\widetilde{X}'\widetilde{X})^{-1}T^{-1}\widetilde{X}\widetilde{DU}\delta_2^*] \\
&= C^{-1}\delta\lambda(1-\lambda)\delta_2^*
\end{aligned}
$$

which establishes parts A2 and B. Part A1 of the theorem follows since $\delta_2{=}0$ and $a_{21}{=}0$ imply that $\delta_2^* = 0$ giving $plim(\hat{a}_2 - a_2) = 0$. Part A3 follows from (14) since $\delta_1 = 0$, $a_{12} = 0$ and $\sigma_{12} = 0$ gives $\Gamma_{12} = 0$ and it follows that $(\delta_1\Gamma_{22} - \delta_2\Gamma_{12}) = 0$. The exception to part B follows since $\delta_2^* = 0$ when $(1 - a_{22})\delta_2 = a_{21}\delta_1$.

The proof of Theorem 2 follows using similar arguments and is omitted.

Proof of Theorem 3

When $\delta_2 = 0$ (and $a_{21} = 0$), the model given by (7) reduces to $y_{2t} = \mu_2^* + a_{22}y_{1t-1} + a_{22}y_{2t-1} + e_{2t}$. Therefore regression (12) is correctly specified and standard OLS results apply giving part 1 of the Theorem. When $\delta_2 \neq 0$, we have from Theorem 1 that $(\hat{a}_{21} - a_{21})^2 = O_p(1) > 0$ and $(T^{-1}\widetilde{X}'\widetilde{X})_{11}^{-1} = O_p(1) > 0$. Let $M = I - \widetilde{X}(\widetilde{X}'\widetilde{X})^{-1}\widetilde{X}'$. The limit of s^2 follows from

$$
\begin{aligned}
s^2 = T^{-1}\widetilde{Y}_2' M \widetilde{Y}_2 \;&=\; T^{-1}(\widetilde{DU}\delta_2^* + \widetilde{DT}B\gamma_2^* + \tilde{e}_2)'M(\widetilde{DU}\delta_2^* + \widetilde{DT}B\gamma_2^* + \tilde{e}_2) \\
&=\; T^{-1}(\widetilde{DU}\delta_2 + \tilde{e}_2)'M(\widetilde{DU}\delta_2^* + \tilde{e}_2) + o_p(1) \\
&=\; \delta_2^* T^{-1}\widetilde{DU}'M\widetilde{DU} + T^{-1}\tilde{e}_2'M\tilde{e}_2 + o_p(1) \\
&=\; (\delta_2^*)^2(T^{-1}\widetilde{DU}'\widetilde{DU} - T^{-1}\widetilde{DU}'\widetilde{X}(T^{-1}\widetilde{X}'\widetilde{X})^{-1}T^{-1}\widetilde{X}'\widetilde{DU}) + \sigma_2^2 + o_p(1) \\
&\xrightarrow{p}\; (\delta_2^*)^2(\lambda(1-\lambda) - \lambda(1-\lambda)\delta'C^{-1}\delta\lambda(1-\lambda)) + \sigma_2^2.
\end{aligned}
$$

Since $0 < plim(s^2) < \infty$, we have $0 < plim(1/s^2) < \infty$. Combining these results gives $T^{-1}GC = (\hat{a}_{21} - a_{21})^2/(s^2(T^{-1}\widetilde{X}'\widetilde{X})_{11}^{-1}) = O_p(1) > 0$. When $\delta_1 = 0$, $a_{12} = 0$ and $\sigma_{12} = 0$, it is straightforward to show that $T(\hat{a}_{21} - a_{21})^2 = O_p(1)$, and since it is still true that $(T^{-1}\widetilde{X}'\widetilde{X})_{11}^{-1} = O_p(1)$ and $0 < plim(1/s^2) < \infty$, we have $GC = O_p(1)$.

24

References

Andrews, D. W. K. (1993), Tests for Parameter Instability and Structural Change with Unknown Change Point, *Econometrica* **61**, 821–856.

Andrews, D. W. K. and Ploberger, W. (1994), Optimal Tests When a Nuisance Parameter is Present Only Under the Alternative, *Econometrica* **62**, 1383–1414.

Bai, J. S. (1994), Least Squares Estimation of a Shift in Linear Process, **15**, 453–472.

Bai, J. S. and Perron, P. (1995), Estimating and Testing Linear Models with Multiple Structural Breaks, mimeo, University of Montreal.

Bai, J. S., Lumsdaine, R. L. and Stock, J. H. (1997), Testing for and Dating Breaks in Stationary and Nonstationary Time Series, *Review of Economic Studies, forthcoming.*

Banerjee, A., Lumsdaine, R. L. and Stock, J. H. (1992), Recursive and Sequential Tests of the Unit Root and Tread Break Hypothesis, *Journal of Business and Economic Statistics* **10**, 271–287.

Ben-David, B. and Papell, D. H. (1995), The Great Wars, the Great Crash, and Steady State Growth: Some New Evidence about an Old Stylized Fact, *Journal of Monetary Economics* **36**, 453–475.

Bianchi, M. (1995), Granger Causality in the Presence of Structural Changes, mimeo, Bank of England.

Blanchard, O. and Quah, D. (1989), The Dynamic Effects of Aggregate Demand and Supply, *American Economic Review* **79**, 655–73.

Box, G. E. P. and Tiao, G. C. (1975), Intervention Analysis With Applications to Economics and Environmental Problems, *Journal of the American Statistical Association* **70**, 70–79.

Canova, F. (1995), Vector Autoregressive Models: Specification, Estimation, Inference and Forecasting, *Handbook of Applied Econometrics*, Basil Blackwell, Oxford.

Chu, C.-S. and White, H. (1992), A Direct Test for Changing Trend, *Journal of Business and Economic Statistics* **10**, 289–300.

Engle, R. F. and Kozicki, S. (1993), Testing for Common Features, *Journal of Business and Economic Statistics* **11**, 369–379.

Gambe, E. N. and Joutz, F. (1993), The Dynamic Effects of Aggregate Demand and Supply Disturbances: Comment, *American Economic Review* **83**, 1387–1393.

Grenander, U. and Rosenblatt, M. (1957), *Statistical Analysis of Stationary Time Series*, Wiley, New York.

Lutkepohl, H. (1989), The stability assumption in tests of causality between money and income, *in* W. Kramer (ed.), *Econometrics of Structural Change*, Springer-Verlag, New York, pp. 75–86.

Lutkepohl, H. (1993), *Introduction to Multiple Time Series*, Springer Verlag, Berlin.

Perron, P. (1989), The great crash, the oil price shock and the unit root hypothesis, *Econometrica* **57**, 1361–1401.

Perron, P. (1990), Testing for a Unit Root in a Time Series with a Changing Mean, *Journal of Business and Economic Statistics* **8**, 153–162.

Perron, P. (1991), A Test for Changes in a Polynomial Trend Function for a Dynamic Time Series, manuscript, Princeton University.

Schwarz, G. (1978), Estimating the Dimension of a Model, *The Annals of Statistics* **6**, 461–464.

Sims, C., Stock, J. H. and Watson, M. W. (1990), Inference in Linear Time Series Models with Some Unit Roots, *Econometrica* **58**, 113–144.

Stock, J. H. and Watson, M. (1989), New Indexes of Coincident and Leading Economic Indications, *in* O. J. Blanchard and S. Fischer (eds), *NBER Macroeconomics Annual 1989*, M. I. T. Press, Cambridge.

Stock, J. H. and Watson, M. W. (1996), Evidence on Structural Instability in Macroeconomic Time Series Relations, *Journal of Business and Economic Statistics* **14**, 11–30.

Tiao, G. C. (1985), Autoregressive Moving Average Models, Intervention Problems and Outlier Detection in Time Series, *Handbook of Statistics 5*, North Holland, Amsterdam, pp. 85–118.

Vogelsang, T. J. (1997a), Wald-Type Tests for Detecting Shifts in the Trend Function of a Dynamic Time Series, *Econometric Theory, forthcoming.*

Vogelsang, T. J. (1997b), Testing for a Shift in Mean Without Having to Estimate Serial Correlation Parameters, *Journal of Business and Economic Statistics, forthcoming.*

Table 1a. Average Estimates for $\hat{a}_{21}$ when $a_{21} = 0$

a_{22}	-0.6	-0.4	-0.2	0.0	0.2	0.4	0.6
			$\delta = (0,1)$				
$a_{12} = -0.6$	-0.0745	-0.0409	-0.0163	0.0004	0.0105	0.0151	0.0148
-0.4	-0.0661	-0.0348	-0.0137	-0.0002	0.0075	0.0108	0.0105
-0.2	-0.0419	-0.0215	-0.0087	-0.0010	0.0033	0.0051	0.0049
0.0	-0.0021	-0.0020	-0.0019	-0.0018	-0.0017	-0.0016	-0.0016
0.2	0.0380	0.0177	0.0051	-0.0024	-0.0066	-0.0082	-0.0080
0.4	0.0630	0.0316	0.0106	-0.0028	-0.0105	-0.0138	-0.0134
0.6	0.0722	0.0384	0.0138	-0.0030	-0.0131	-0.0177	-0.0174
			$\delta = (1,1)$				
-0.6	0.0717	0.0931	0.1012	0.0994	0.0904	0.0756	0.0556
-0.4	0.1221	0.1285	0.1246	0.1141	0.0988	0.0798	0.0572
-0.2	0.1793	0.1636	0.1451	0.1250	0.1038	0.0812	0.0568
0.0	0.2214	0.1864	0.1561	0.1288	0.1034	0.0788	0.0538
0.2	0.2307	0.1890	0.1541	0.1240	0.0972	0.0724	0.0484
0.4	0.2128	0.1747	0.1413	0.1121	0.0863	0.0630	0.0411
0.6	0.1843	0.1526	0.1232	0.0968	0.0733	0.0523	0.0330

Table 1b. Average Estimates for $\hat{a}_{21}$ Under Granger Causality: $a_{21} = 0.3$.

	δ	$a_{22}=$ -0.6	-0.4	-0.2	0.0	0.2	0.4	0.6
$a_{12}=0.0$	(1,0)	0.3599	0.3298	0.3085	0.2947	0.2878	0.2265	0.2583
0.0	(0,1)	0.3395	0.3393	0.3339	0.3263	0.3182	0.3105	0.3041
0.0	(1,1)	0.5077	0.4723	0.4377	0.4047	0.3732	0.3427	0.3126
-0.4	(1,0)	0.3443	0.3273	0.3139	0.3039	0.2975	0.4325	0.3869
-0.4	(0,1)	0.2793	0.3028	0.3174	0.3243	0.3254	0.3224	0.3164
-0.4	(1,1)	0.4304	0.4288	0.4165	0.3971	0.3729	0.3454	0.3149
0.4	(1,0)	0.2853	0.3004	0.3103	0.3155	0.3157	0.2617	0.2783
0.4	(0,1)	0.4325	0.3869	0.3533	0.3284	0.3105	0.2985	0.2919
0.4	(1,1)	0.5099	0.4626	0.4226	0.3882	0.3582	0.3316	0.3077

Table 1c. Average Estimates for $\hat{a}_{21}$ Under Granger Causality: $a_{21} = -0.3$.

	δ	$a_{22}=$ -0.6	-0.4	-0.2	0.0	0.2	0.4	0.6
$a_{12}=0.0$	(1,0)	-0.2530	-0.2528	-0.2528	-0.2529	-0.2530	-0.2532	-0.2534
0.0	(0,1)	-0.3435	-0.3431	-0.3375	-0.3297	-0.3214	-0.3137	-0.3072
0.0	(1,1)	-0.0547	-0.0941	-0.1246	-0.1492	-0.1703	-0.1898	-0.2092
-0.4	(1,0)	-0.2622	-0.2616	-0.2611	-0.2607	-0.2603	-0.2601	-0.2599
-0.4	(0,1)	-0.4349	-0.3896	-0.3560	-0.3312	-0.3132	-0.3012	-0.2946
-0.4	(1,1)	-0.2147	-0.1976	-0.1895	-0.1878	-0.1910	-0.1987	-0.2108
0.4	(1,0)	-0.2645	-0.2615	-0.2592	-0.2575	-0.2562	-0.2553	-0.2546
0.4	(0,1)	-0.2823	-0.3058	-0.3203	-0.3272	-0.3283	-0.3252	-0.3192
0.4	(1,1)	-0.0766	-0.1068	-0.1358	-0.1622	-0.1856	-0.2066	-0.2261

Table 2a. Average Estimates for $\hat{a}_{21}(1)$ at Different Lag Lengths, k.
$a_{21} = 0.0$, $a_{12} = 0.4$.

k	δ	$a_{22}=$ -0.6	-0.4	-0.2	0.0	0.2	0.4	0.6
1	(1,0)	-0.0021	-0.0023	-0.0024	-0.0025	-0.0026	-0.0026	-0.0027
2	(1,0)	-0.0015	-0.0016	-0.0016	-0.0016	-0.0016	-0.0017	-0.0017
4	(1,0)	-0.0014	-0.0014	-0.0014	-0.0013	-0.0013	-0.0013	-0.0012
8	(1,0)	-0.0018	-0.0018	-0.0018	-0.0018	-0.0018	-0.0018	-0.0018
1	(0,1)	0.0630	0.0316	0.0106	-0.0028	-0.0105	-0.0138	-0.0134
2	(0,1)	-0.0871	-0.0811	-0.0741	-0.0660	-0.0566	-0.0457	-0.0331
4	(0,1)	-0.0514	-0.0515	-0.0507	-0.0487	-0.0449	-0.0388	-0.0297
8	(0,1)	-0.0305	-0.0318	-0.0329	-0.0335	-0.0330	-0.0307	-0.0253
1	(1,1)	0.2128	0.1747	0.1413	0.1121	0.0863	0.0630	0.0411
2	(1,1)	0.1872	0.1739	0.1582	0.1399	0.1189	0.0947	0.0670
4	(1,1)	0.1069	0.1040	0.0989	0.0912	0.0805	0.0661	0.0478
8	(1,1)	0.0617	0.0610	0.0592	0.0559	0.0504	0.0422	0.0308

Table 2b. Average Estimates for $\hat{a}_{21}(1)$ at Different Lag Lengths, k.
$a_{21} = 0.0$, $a_{12} = -0.4$.

k	δ	$a_{22}=$ -0.6	-0.4	-0.2	0.0	0.2	0.4	0.6
1	(1,0)	-0.0006	-0.0005	-0.0005	-0.0005	-0.0005	-0.0005	-0.0005
2	(1,0)	-0.0015	-0.0015	-0.0016	-0.0016	-0.0016	-0.0016	-0.0016
4	(1,0)	-0.0014	-0.0014	-0.0014	-0.0014	-0.0013	-0.0013	-0.0013
8	(1,0)	-0.0018	-0.0018	-0.0018	-0.0018	-0.0018	-0.0018	-0.0018
1	(0,1)	-0.0661	-0.0348	-0.0137	-0.0002	0.0075	0.0108	0.0105
2	(0,1)	0.0854	0.0794	0.0723	0.0640	0.0545	0.0435	0.0308
4	(0,1)	0.0498	0.0499	0.0490	0.0469	0.0431	0.0368	0.0275
8	(0,1)	0.0279	0.0291	0.0302	0.0308	0.0304	0.0281	0.0227
1	(1,1)	0.1221	0.1285	0.1246	0.1141	0.0988	0.0798	0.0572
2	(1,1)	0.2751	0.2464	0.2164	0.1850	0.1520	0.1172	0.0803
4	(1,1)	0.1334	0.1242	0.1129	0.0994	0.0836	0.0655	0.0453
8	(1,1)	0.0733	0.0688	0.0631	0.0560	0.0473	0.0371	0.0254

Table 3a. Size of the Granger Causality Test at Different Lag Lengths, k.
$a_{21} = 0.0$, $a_{12} = 0.4$.

k	δ	$a_{22}=$ -0.6	-0.4	-0.2	0.0	0.2	0.4	0.6
1	(1,0)	0.0580	0.0600	0.0590	0.0540	0.0590	0.0590	0.0630
2	(1,0)	0.0540	0.0570	0.0540	0.0550	0.0540	0.0600	0.0560
4	(1,0)	0.0470	0.0480	0.0470	0.0490	0.0520	0.0490	0.0420
8	(1,0)	0.0570	0.0560	0.0550	0.0550	0.0550	0.0570	0.0550
1	(0,1)	0.1830	0.1100	0.0870	0.0680	0.0740	0.0810	0.0870
2	(0,1)	0.3510	0.3100	0.2670	0.2130	0.1740	0.1270	0.0970
4	(0,1)	0.2650	0.2660	0.2700	0.2510	0.2190	0..1910	0.1330
8	(0,1)	0.1380	0.1530	0.1690	0.1860	0.1960	0.1920	0.1360
1	(1,1)	1.0000	0.9980	0.9610	0.8650	0.6210	0.3360	0.1470
2	(1,1)	0.9640	0.9200	0.8140	0.6590	0.4510	0.2460	0.1170
4	(1,1)	0.5010	0.4720	0.4080	0.3300	0.2510	0.1600	0.0820
8	(1,1)	0.1240	0.1310	0.1290	0.1260	0.1200	0.0950	0.0770

Table 3b. Size of the Granger Causality Test at Different Lag Lengths, k.
$a_{21} = 0.0$, $a_{12} = -0.4$.

| k	δ	$a_{22}=$ -0.6	-0.4	-0.2	0.0	0.2	0.4	0.6
1	(1,0)	0.0450	0.0430	0.0470	0.0490	0.0510	0.0510	0.0580
2	(1,0)	0.0470	0.0460	0.0450	0.0480	0.0470	0.0470	0.0450
4	(1,0)	0.0500	0.0500	0.0500	0.0500	0.0520	0.0520	0.0550
8	(1,0)	0.0590	0.0600	0.0630	0.0620	0.0600	0.0600	0.0600
1	(0,1)	0.1940	0.1250	0.0750	0.0580	0.0530	0.0660	0.0750
2	(0,1)	0.3210	0.2800	0.2390	0.2020	0.1670	0.1300	0.0990
4	(0,1)	0.2470	0.2530	0.2570	0.2470	0.2240	0.1790	0.1300
8	(0,1)	0.1410	0.1680	0.1860	0.2000	0.2030	0.1960	0.1540
1	(1,1)	0.9680	0.9560	0.9320	0.8890	0.8020	0.6530	0.4170
2	(1,1)	0.9930	0.9880	0.9780	0.9580	0.8910	0.7330	0.4560
4	(1,1)	0.9630	0.9630	0.9600	0.9460	0.9100	0.7730	0.4940
8	(1,1)	0.6960	0.7550	0.8000	0.8190	0.7960	0.6850	0.4400

Table 4a: Size and Frequency Distribution of the Lag Order Selected by AIC: $kmax = 4$.
$a_{11} = 0.5, a_{12} = 0.4$.

| $GC(1)$	$GC(k_{aic}^{kmax=4})$	$GC(k_{aic}^{kmax=8})$	δ	k=1	k=2	k=3	k=4
$a_{22} = 0.2$							
0.0640	0.1040	0.1090	(0,0)	0.3450	0.1850	0.2080	0.2620
0.0610	0.0910	0.1040	(1,0)	0.2040	0.1600	0.2050	0.4310
0.0660	0.2330	0.2190	(0,1)	0.0040	0.0250	0.1510	0.8200
0.7110	0.3210	0.1750	(1,1)	0.0290	0.0600	0.2160	0.6950
$a_{22} = -0.2$							
0.0640	0.0870	0.1000	(0,0)	0.3530	0.1850	0.1840	0.2780
0.0590	0.0760	0.0940	(1,0)	0.2150	0.1560	0.2080	0.4210
0.0770	0.2710	0.1910	(0,1)	0.0000	0.0000	0.0170	0.9830
0.9690	0.4260	0.1640	(1,1)	0.0000	0.0010	0.0570	0.9420

Table 4b: Size and Frequency Distribution of the Lag Order Selected by AIC: $kmax = 4$.
$a_{11} = 0.5, a_{12} = -0.4$.

| $GC(1)$	$GC(k_{aic}^{kmax=4})$	$GC(k_{aic}^{kmax=8})$	δ	k=1	k=2	k=3	k=4
$a_{22} = 0.2$							
0.0570	0.0800	0.0840	(0,0)	0.3560	0.1740	0.1890	0.2810
0.0480	0.0720	0.0900	(1,0)	0.1940	0.1570	0.2260	0.4230
0.0510	0.2380	0.2360	(0,1)	0.0040	0.0290	0.1390	0.8280
0.8120	0.9100	0.8160	(1,1)	0.0000	0.0010	0.0400	0.9590
$a_{22} = -0.2$							
0.0590	0.0770	0.0870	(0,0)	0.3400	0.1810	0.2050	0.2740
0.0480	0.0680	0.0770	(1,0)	0.1970	0.1660	0.2400	0.3970
0.0720	0.2570	0.2050	(0,1)	0.0000	0.0000	0.0190	0.9810
0.9610	0.9600	0.8190	(1,1)	0.0000	0.0000	0.0020	0.9980
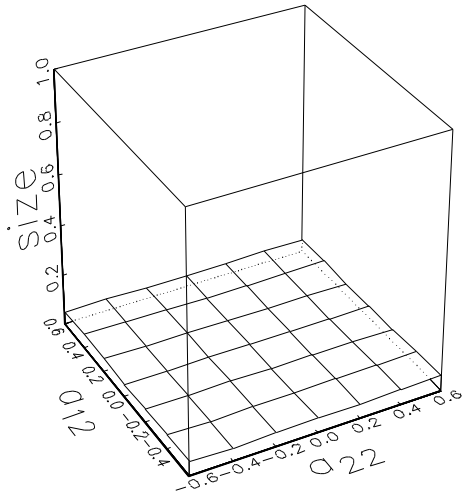
Panel (i) $\delta_1=0$, $\delta_2=0$, $a_{11}=0.0$
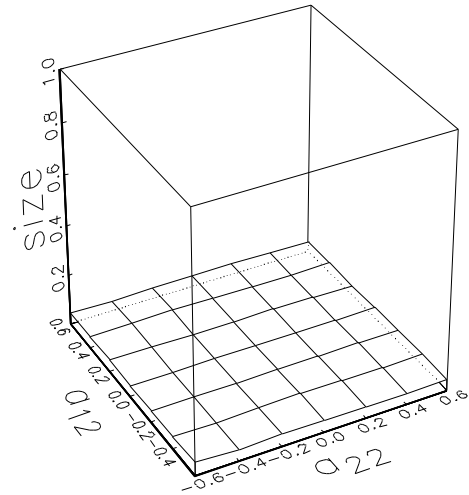
Panel (ii) $\delta_1=1$, $\delta_2=0$, $a_{11}=0.0$
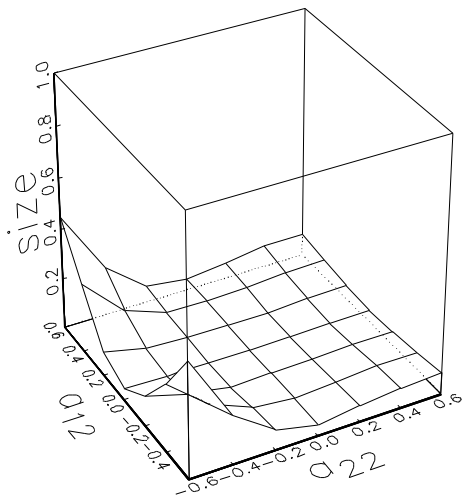
Panel (iii) $\delta_1=0$, $\delta_2=1$, $a_{11}=0.0$
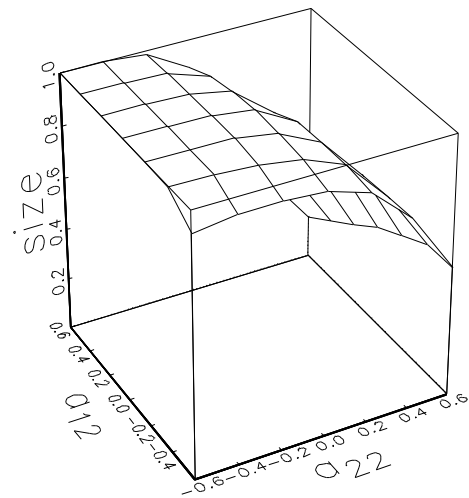
Panel (iv) $\delta_1=1$, $\delta_2=1$, $a_{11}=0.0$

Figure 1: Size of Granger Causality Test, T=100.