Causality tests and conditional heteroskedasticity: Monte Carlo evidence

Jon Vilasuso*

Department of Economics, College of Business and Economics, West Virginia University, P.O. Box 6025, Morgantown, WV 26506-6025, USA

Received 1 January 1997; received in revised form 1 March 2000; accepted 5 July 2000

Abstract

This paper investigates the reliability of causality tests based on least squares when conditional heteroskedasticity exists. Monte Carlo evidence documents considerable size distortion if the conditional variances are correlated. Inference based on a heteroskedasticity and autocorrelation consistent covariance matrix offers little improvement. This size distortion traces to an inability to discriminate between causality in mean and causality in variance. As a result, this paper endorses conducting causality tests based on an empirical specification that explicitly models the conditional means and conditional variances. The relationship between money and prices serves as an illustrative example. © 2001 Elsevier Science S.A. All rights reserved.

JEL classification: Classification: C12; C15

Keywords: Simulation; ARCH

1. Introduction

Causality tests are routinely used to determine whether changes in one variable help explain movements in another variable. One popular approach is based on the work of Weiner (1956) and Granger (1969) and involves evaluating
a zero restriction in a vector autoregression (VAR).\textsuperscript{1} Unfortunately, conclusions are often sensitive to alternative specifications of the time-series properties of the data. In the case of money and output, for example, researchers have shown that inference is sensitive to the lag specification (Feige and Pearce, 1979), the method of time aggregation (Christiano and Eichenbaum, 1987), and the long-run properties of the data (Christiano and Ljungqvist, 1988; Stock and Watson, 1989). This paper examines the relationship between the time series properties of the data and statistical inference, focusing on whether conclusions based on causality tests using least-squares estimation are reliable if conditional heteroskedasticity is present.

Results of Monte Carlo simulations suggest that conclusions drawn from least squares causality tests may lead to an erroneous claim that a statistically significant causal relation exists. Misleading inference associated with least squares may be traced to two explanations. First, because the set of regressors in a VAR includes lagged-dependent variables, least-squares standard errors are not consistent and may not support correct statistical inference (Engle et al., 1985). Consequently, a logical way to proceed is to base inference on a heteroskedasticity and autocorrelation consistent (HAC) covariance matrix. The use of a HAC covariance matrix, however, may not improve inference if there is 'considerable' temporal dependence because kernel-based HAC covariance estimators have test statistic values larger (in absolute terms) than that implied by the limiting distribution (Andrews and Monahan, 1992).\textsuperscript{2} Hence, the null hypothesis of no causality is rejected too often.

A second explanation for the unsatisfactory performance of least-squares causality tests traces to a failure to adequately differentiate between causality in mean and causality in variance. This could prove particularly troublesome in practice, especially if interest is directed at asset pricing issues as Engle et al. (1990), Cheung and Ng (1996), and Lin (1997) document instances of volatility spillovers or causality in variance. Thus, if the poor performance of least-squares causality tests traces to an inability to differentiate between causality in mean and causality in variance, then causality tests are best based on a complete empirical specification of the conditional means and conditional variances.

Monte Carlo simulations suggest that if the conditional variances are unrelated, then there is only slight size distortion associated with least-squares tests, and the inconsistency of the least-squares standard errors is unlikely to be

---

\textsuperscript{1} For comprehensive treatments of the definition of causality, see the papers included in the Journal of Econometrics Annals Issue (1988) entitled "Causality", edited by Dennis Aigner and Arnold Zellner.

\textsuperscript{2} Andrews and Monahan (1992) and Cheung and Lai (1997) present Monte Carlo evidence that prewhitening is effective in reducing over-rejection. As a result, this paper uses prewhitened kernel estimators in the empirical work to follow.
Cheung and Ng (1996, p. 39) point out that causality in mean may affect tests for causality in variance. Causality in mean affects the structure of the disturbance terms, and in an autoregressive conditional heteroskedasticity model, the conditional variance is a linear function of the squared disturbances. On the other hand, causality in variance may also have a possible, but smaller effect, on causality in mean tests because the conditional mean does not depend on the second moment (the ARCH-M model being an exception). In this paper, we numerically examine this latter effect.

Note that a linear conditional mean model with ARCH disturbances can be described by a nonlinear specification without ARCH, i.e. the bilinear model. In this paper, we assume that the conditional mean is linear and is correctly specified in the simulations to follow.

problematic. However, if the conditional variances are correlated, there is considerable size distortion and a HAC covariance matrix offers only slight improvement. An apparent inability to differentiate between causality in mean and causality in variance is likely to present serious challenges in practice, and as a result, this paper recommends basing inference on an empirical model that explicitly models the conditional means and conditional variances. The relationship between money and prices serves as an illustrative example that statistical inference based on the least-squares framework and a fully specified model differs widely.

2. Monte Carlo experiments

Monte Carlo experiments are used to assess the accuracy of statistical inference based on least squares causality tests when conditional heteroskedasticity is present. Consider the mean-zero time series \( y_{1,t} \) and \( y_{2,t} \) for \( t = 1, 2, \ldots, T \) generated by a stationary VAR(1)

\[
\begin{bmatrix}
  y_{1,t} \\
  y_{2,t}
\end{bmatrix} =
\begin{bmatrix}
  \pi_{11} & \pi_{12} \\
  \pi_{21} & \pi_{22}
\end{bmatrix}
\begin{bmatrix}
  y_{1,t-1} \\
  y_{2,t-1}
\end{bmatrix} +
\begin{bmatrix}
  e_{1,t} \\
  e_{2,t}
\end{bmatrix},
\]

where \([\pi_{ij}]\) is a matrix of coefficients and \([e_{1,t}, e_{2,t}]\)' is a vector of mean-zero serially uncorrelated disturbance terms with conditional covariance matrix

\[
\begin{bmatrix}
  h_{11,t} & h_{12,t} \\
  h_{21,t} & h_{22,t}
\end{bmatrix} =
\begin{bmatrix}
  k_{11} & k_{12} \\
  k_{21} & k_{22}
\end{bmatrix} +
\begin{bmatrix}
  a_{11} & a_{12} & e_{1,t-1} \\
  a_{21} & a_{22} & e_{2,t-1}
\end{bmatrix}
\begin{bmatrix}
  e_{1,t-1} \\
  e_{2,t-1}
\end{bmatrix} +
\begin{bmatrix}
  a_{11} & a_{12} \\
  a_{21} & a_{22}
\end{bmatrix},
\]

where \([k_{ij}]\) and \([a_{ij}]\) are matrices of coefficients. The regression model specifies a linear function for the conditional means where the disturbance terms are assumed to follow a first-order multivariate autoregressive conditional heteroskedasticity (ARCH) model.

The Monte Carlo experiments are designed such that \([\pi_{ij}]\) is a lower triangular matrix which implies that \( y_{1,t} \) causes \( y_{2,t} \) in mean. In the experiments to follow,

\(^3\)Cheung and Ng (1996, p. 39) point out that causality in mean may affect tests for causality in variance. Causality in mean affects the structure of the disturbance terms, and in an autoregressive conditional heteroskedasticity model, the conditional variance is a linear function of the squared disturbances. On the other hand, causality in variance may also 'have a possible, but smaller effect' on causality in mean tests because the conditional mean does not depend on the second moment (the ARCH-M model being an exception). In this paper, we numerically examine this latter effect.

\(^4\)Note that a linear conditional mean model with ARCH disturbances can be described by a nonlinear specification without ARCH, i.e. the bilinear model. In this paper, we assume that the conditional mean is linear and is correctly specified in the simulations to follow.
we set $\pi_{11} = 0.50, \pi_{12} = 0, \pi_{21} = 0.50$, and $\pi_{22} = 0.25$. Also, each element of $[k_{ij}]$ is set equal to one. The data generating process for $e_{i,t}$ is given by $\eta_{i,t}\sqrt{h_{ii,t}}$ where $\eta_{i,t}$ is i.i.d. N(0,1). The simulations consider a sample size of 100 and 500, using 5000 replications.

The experiments consider three different parameterizations of $[a_{ij}]$. Model 1 serves as a benchmark and $[a_{ij}]$ is set equal to the null matrix. In this case, the disturbance terms are homoskedastic, and least squares is consistent and efficient.

The remaining models maintain that the disturbances exhibit conditional heteroskedasticity. The conditional covariance matrix for Model 2 is based on

$$
\text{Model 2} \quad a = \begin{bmatrix} 0.2 & 0 \\ 0 & 0.9 \end{bmatrix}.
$$

Under Model 2, the conditional variances of the disturbance terms exhibit conditional heteroskedasticity but are unrelated. Because the set of regressors and the disturbances are correlated, least-squares standard errors are not consistent. The reliability of statistical inference is then potentially improved by constructing a HAC covariance matrix. The heteroskedasticity consistent estimator (HCE) developed by White (1980) is considered along with kernel-based HAC estimators popularized by Newey and West (1987) and Andrews (1991). We consider the Bartlett and Quadratic Spectral (QS) kernels where the bandwidth parameter is set according to the guidelines presented by Newey and West (1994) where a VAR(1) is used to prewhiten prior to testing for causality in mean. A comparison of the results for Models 1 and 2 then examines whether least-squares causality tests are adversely affected by the inconsistency of the standard errors, and if so, whether a consistent estimator of the covariance matrix offers improvement.

Model 3 maintains that there is a causality in variance relation. The matrix $[a_{ij}]$ is an upper triangular matrix which implies that the conditional variance of $y_{2,t}$ causes the conditional variance of $y_{1,t}$, but given that $[\pi_{ij}]$ is lower triangular, $y_{2,t}$ does not cause $y_{1,t}$ in mean. The conditional covariance matrix for Model 3 is based on

$$
\text{Model 3} \quad a = \begin{bmatrix} 0.2 & 0.7 \\ 0 & 0.9 \end{bmatrix}.
$$

Comparing results for Models 2 and 3 sheds light on whether the performance of least-squares causality tests in mean are affected by causality in variance.


Table 1
Empirical size of nominal 5% causality in mean tests

<table>
<thead>
<tr>
<th>Model description</th>
<th>Covariance matrix estimation method</th>
<th>Empirical size, $T = 100$</th>
<th>Empirical size, $T = 500$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model 1</td>
<td>LS</td>
<td>5.94</td>
<td>4.92</td>
</tr>
<tr>
<td></td>
<td>HCE</td>
<td>5.90</td>
<td>5.12</td>
</tr>
<tr>
<td></td>
<td>Bartlett</td>
<td>5.60</td>
<td>5.42</td>
</tr>
<tr>
<td></td>
<td>QS</td>
<td>5.12</td>
<td>5.40</td>
</tr>
<tr>
<td>Model 2</td>
<td>LS</td>
<td>6.36</td>
<td>5.78</td>
</tr>
<tr>
<td></td>
<td>HCE</td>
<td>6.60</td>
<td>5.12</td>
</tr>
<tr>
<td></td>
<td>Bartlett</td>
<td>6.82</td>
<td>5.20</td>
</tr>
<tr>
<td></td>
<td>QS</td>
<td>6.72</td>
<td>5.18</td>
</tr>
<tr>
<td>Model 3</td>
<td>LS</td>
<td>13.60</td>
<td>12.16</td>
</tr>
<tr>
<td></td>
<td>HCE</td>
<td>10.62</td>
<td>10.14</td>
</tr>
<tr>
<td></td>
<td>Bartlett</td>
<td>11.40</td>
<td>10.46</td>
</tr>
<tr>
<td></td>
<td>QS</td>
<td>11.48</td>
<td>10.04</td>
</tr>
</tbody>
</table>

*Notes: The empirical size is based on 5000 replications of each model and $T$ denotes the sample size. Inference is based on least squares (LS), White’s (1980) heteroskedasticity consistent estimator (HCE), and kernel-based HAC estimators of the covariance matrix. Kernels examined include the Bartlett and Quadratic Spectral (QS) kernels where the bandwidth parameter is set according to the guidelines offered by Newey and West (1994).*

3. Simulation results

Table 1 summarizes the empirical test size associated with the null hypothesis that $y_{2,t}$ does not cause $y_{1,t}$ in mean, which is true by construction. The nominal test size is 5%. Inference is based on the least-squares (LS) standard errors, White’s (1980) heteroskedasticity consistent estimator (HCE), and a HAC covariance matrix using a Bartlett and Quadratic Spectral kernel computed under the least-squares framework.

For Model 1, $e_{1,t}$ and $e_{2,t}$ are uncorrelated and are distributed i.i.d. N(0,1). Thus, least-squares is consistent and efficient, and the empirical test size should not differ from the nominal test size. As shown in Table 1, this is the case as the empirical test size is approximately equal to the nominal test size.

The empirical test size based on the least-squares standard errors for Model 2 is approximately equal to the nominal test size for $T = 500$. This suggests that if $[a_{ij}]$ is diagonal, inference based on the least-squares standard errors appears robust for larger samples. That is, if the conditional variances are independent, then inference based on least squares is reliable even though the least-squares standard errors are not consistent. There is some gain from using a HAC covariance for $T = 500$. Results for the Bartlett and QS kernels differ little, which is consistent with the conclusion of Newey and West (1994) that the
choice of kernel is of secondary importance compared to the choice of the bandwidth parameter. Comparing results for Models 1 and 2 suggests that the inconsistency of the least-squares standard errors is not particularly harmful if the conditional variances are unrelated, although a HAC estimator of the covariance matrix does improve inference for larger samples.

Tests results based on the least-squares standard errors for Model 3, however, are markedly different. There is considerable size distortion, and constructing a HAC covariance matrix offers only slight improvement. This finding likely traces to an inability to discriminate between causality in mean and causality in variance. That is, when there is causality in variance, inference based on least squares and a HAC covariance matrix is apt to erroneously claim that there is causality in mean.

4. The money–price relation

This section explores the relationship between money and prices in an effort to gauge the performance of least-squares causality tests in practice. Friedman and Schwartz (1963) espouse the Monetarist claim that money causes prices in mean via the quantity equation. However, it is certainly plausible that causality runs from prices to money. If the monetary authorities condition policy at least in part on past inflation, then lagged inflation helps explain movements in intermediate or indicator variables which include money (see, e.g. Fuhrer and Moore, 1995).

The causal structure linking money and prices is examined for the period 1964:I–1997:III. Seasonally adjusted, quarterly data for money and prices were obtained from the St. Louis Federal Reserve data files. M1 serves as a measure of money and the CPI represents the price level. As noted by Hoover (1991), M1 is an appropriate measure of money because M1 served as the primary monetary aggregate used by the monetary authorities for much of the sample period.

Because causality tests are sensitive to the long-run characteristics of the data, we examine the trend properties of the data expressed in log levels. The sequential unit root testing procedure (not shown) developed by Dickey and Pantula (1987) for up to two unit roots finds that money and prices have a single unit root. In an influential study, Stock and Watson (1989) report that log-differenced M1 is stationary around a linear time trend, and prices are best described as an I(1) process with drift for the 1960–1985 period. We follow this work and regress the log-differenced data on a constant, a linear time trend, and six of its own lags (not shown). Results generally agree with Stock and Watson’s work, but we find much less evidence that M1 growth follows a time trend when more recent observations are included. One explanation for this finding is that M1 growth exhibits a trend break. We explore this possibility using the SupF
We also tabulated a critical value of the SupF test based on bootstrap simulations because the testing procedure developed by Bai and Perron (1997) is asymptotic, and in some cases dividing the full sample results in small subsamples. In any event, the asymptotic and bootstrap values both suggest a significant trend break at the 5% level.

Causality in mean test results are summarized in Table 2. Causality tests constructed under the least squares framework evaluate a zero restriction in a estimated VAR which includes four lagged terms as suggested by Akaike’s Information Criterion. Panel A examines whether money causes prices, and Panel B focuses on whether prices cause money. Beginning in Panel A, statistical inference based on least squares standard errors and a HAC covariance matrix support the claim that money causes prices at the 5% significance level. Turning to Panel B, there is little evidence that prices cause money. Thus, causality test results constructed under the least-squares framework support the Monetarist claim that money causes prices.

There is, however, evidence of conditional heteroskedasticity. The Ljung and Box (1978) portmanteau test for up to sixth-order serial correlation applied to the squared disturbances is significant at the 1% level in the price equation for both descriptions of money (not shown). Engle’s (1982) Lagrange multiplier test (not shown) rejects the homoskedastic null hypothesis at the 5% significance level for the squared disturbances from the price and money equations, regardless of whether a trend break in money growth is included. Consequently, the least-squares framework may not support accurate statistical inference, particularly if the time-varying conditional variances are correlated.

Because the causal structure linking money and prices may encompass the second moments, maximum likelihood estimation is better suited to differentiate between causality in mean and causality in variance. Two maximum likelihood techniques are considered: a multivariate ARCH(1) specification denoted ‘multivariate’ and the univariate two-stage procedure\(^8\) developed by Cheung and Ng (1996) labeled ‘univariate’ in Table 2. The two-stage procedure is based on the cross correlations of the standardized residuals from an AR(4) model with ARCH(1) disturbances. Maximum likelihood hypothesis test statistics are based on the standard errors suggested by Bollerslev and Wooldridge (1992) that are robust to non-normality. Contrary to conclusions based on the least squares

\(^5\) We also tabulated a critical value of the SupF test based on bootstrap simulations because the testing procedure developed by Bai and Perron (1997) is asymptotic, and in some cases dividing the full sample results in small subsamples. In any event, the asymptotic and bootstrap values both suggest a significant trend break at the 5% level.

\(^6\) Consistent with Stock and Watson (1989), we find no evidence that prices and money are cointegrated.

\(^7\) Ball and Cecchetti (1990), among others, also detect conditional heteroskedasticity for inflation.

\(^8\) Monte Carlo results reported by Cheung and Ng (1996) indicate that the two-stage maximum likelihood test possesses good empirical size and power properties.
Table 2

<table>
<thead>
<tr>
<th>Covariance matrix estimation method</th>
<th>Measure of money</th>
<th>Detrended M1</th>
<th>Detrended M1 with trend break</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>A. p-values for the hypothesis that all coefficients on lagged money in the price equation are zero.</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Least-squares framework</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LS</td>
<td>0.05</td>
<td>0.05</td>
<td></td>
</tr>
<tr>
<td>HCE</td>
<td>0.04</td>
<td>0.02</td>
<td></td>
</tr>
<tr>
<td>Bartlett</td>
<td>0.02</td>
<td>0.02</td>
<td></td>
</tr>
<tr>
<td>QS</td>
<td>0.02</td>
<td>0.02</td>
<td></td>
</tr>
<tr>
<td>Maximum likelihood</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Multivariate</td>
<td>0.35</td>
<td>0.17</td>
<td></td>
</tr>
<tr>
<td>Univariate</td>
<td>0.44</td>
<td>0.22</td>
<td></td>
</tr>
<tr>
<td><strong>B. p-values for the hypothesis that all coefficients on lagged prices in the money equation are zero.</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Least-squares framework</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LS</td>
<td>0.14</td>
<td>0.19</td>
<td></td>
</tr>
<tr>
<td>HCE</td>
<td>0.16</td>
<td>0.17</td>
<td></td>
</tr>
<tr>
<td>Bartlett</td>
<td>0.24</td>
<td>0.23</td>
<td></td>
</tr>
<tr>
<td>QS</td>
<td>0.18</td>
<td>0.20</td>
<td></td>
</tr>
<tr>
<td>Maximum likelihood</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Multivariate</td>
<td>0.08</td>
<td>0.04</td>
<td></td>
</tr>
<tr>
<td>Univariate</td>
<td>0.10</td>
<td>0.05</td>
<td></td>
</tr>
</tbody>
</table>

*Notes: Kernel-based HAC estimators of the covariance matrix use a VAR(1) to prewhiten. Multivariate (quasi) maximum likelihood causality in mean tests are based on a multivariate ARCH(1) specification, and the univariate (quasi) maximum likelihood causality in mean tests are based on an AR(4) with ARCH(1) residuals. Maximum likelihood p-values are based on the Bollerslev and Wooldridge (1992) standard errors.

framework, results shown in Table 2 indicate that causality tests based on maximum likelihood techniques offer evidence that prices cause money, consistent with the reaction function literature and findings reported by Hoover (1991).9

The contradictory causality conclusions based on maximum likelihood tests and tests constructed under the least-squares framework are likely due to the existence of a causality in variance relation. Tests for causality in variance are summarized in Table 3. In Panel A, p-values offer little support that prices cause money in variance, but turning to Panel B there is evidence that the

9The ARCH(1) specification appears to adequately fit the data. Likelihood ratio tests reject homogeneity at the 1% significance level for each specification, and Ljung–Box tests applied to the squared standardized residuals do not indicate statistically significant serial correlation.
Table 3
Money-price causality in variance tests, 1964:I–1997:IV*

<table>
<thead>
<tr>
<th>Measure of money</th>
<th>Detrended M1</th>
<th>Detrended M1 with trend break</th>
</tr>
</thead>
<tbody>
<tr>
<td>Estimation method</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A. $p$-values for the hypothesis that prices do not cause money in variance.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Maximum likelihood</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Multivariate</td>
<td>0.12</td>
<td>0.16</td>
</tr>
<tr>
<td>Univariate</td>
<td>0.23</td>
<td>0.29</td>
</tr>
<tr>
<td>B. $p$-values for the hypothesis that money does not cause prices in variance</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Maximum likelihood</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Multivariate</td>
<td>0.05</td>
<td>0.06</td>
</tr>
<tr>
<td>Univariate</td>
<td>0.03</td>
<td>0.04</td>
</tr>
</tbody>
</table>

*Notes: See notes to Table 2.

conditional variance of money causes the conditional variance of prices. Thus, the least-squares conclusion that money causes prices in mean likely traces to an inability to differentiate between causality in mean and causality in variance.

5. Conclusion

This paper demonstrates that if conditional heteroskedasticity is ignored, least squares causality tests exhibit considerable size distortion if the conditional variances are correlated. Moreover, inference based on a HAC covariance matrix constructed under the least squares framework offers only slight improvement. The poor performance of least squares causality tests traces to an inability to differentiate between causality in mean and causality in variance if the regression disturbances are incorrectly assumed to be homoskedastic. Consequently, this paper recommends that causality tests be based on an empirical specification that models both the conditional means and the conditional variances.

Maximum likelihood estimation is well suited to jointly model the conditional means and the conditional variances. Numerous studies have successfully estimated multivariate ARCH models using maximum likelihood.\textsuperscript{10} However, there is certain to be cases where maximizing a multivariate ARCH (log)

\textsuperscript{10}See Bollerslev et al. (1992) for a discussion of this literature.
likelihood function proves daunting if a large number of time series is included or if a long lag structure is used to specify the conditional means. In these cases, the univariate maximum likelihood causality tests proposed by Cheung and Ng (1996) represent an attractive alternative.

References