Modelling Volatility Effects in Interrupted Time Series

MARK VILLAR

University of New South Wales

WILLIAM T.M. DUNSMUIR

University of New South Wales

ABSTRACT

This paper develops a practical approach to modelling the dynamically changing variability associated with public policy interventions and its impact on public health outcomes. An assessment of the deregulation in U.S. alcohol sales is made using Generalised Autoregressive Conditional Heteroskedasticity (GARCH) models that characterise the dynamic nature of the volatility. While these models have received little attention outside financial and econometric settings, this study reveals its potential applicability to the public health arena, specifically its extensions to Autoregressive Integrated Moving Average (ARIMA) intervention models. By capturing the time-varying stochastic volatility associated with discrete changes in alcohol control policies, we found that 'volatility effects' significantly change our inferences about the impact of privatisation on consumption levels. In general, the adjustment for volatility effects: (i) reduces the magnitude of the estimated permanent change in alcohol consumption, and (ii) increases estimation efficiency through lower standard error estimates, compared to standard interrupted time series analysis. This study not only highlights the efficacy of GARCH models in identifying potential biases that can arise from misspecification of the volatility structure but also demonstrates the potential use of such models in more diverse settings where similar variance anomalies have appeared. The performance and intuitive appeal of these models make their use compelling for future research and development.

1 Introduction

1.1 Overview

Many time series encountered in practice require the assessment of impacts from covariates including intervention variables. During the period for which a time series is observed, it is sometimes the case that a change has affected the level or expected path of the series, such as a public policy change. Modelling intervention variables is relevant as it enables meaningful estimation and interpretation of changes that interrupt the normal evolution of the observed series. In recent years there has been a growing literature examining the impact of industry deregulation on alcohol sales and consumption. Numerous articles have appeared in this vein, for example Macdonald (1986), Smart (1986), Mulford and Fitzgerald (1988), Wagenaar and Langley (1995) and Wagenaar and Holder (1991, 1995). However, the sizeable body of evidence determining whether privatisation delivered long-term or permanent changes to consumption levels have produced some widely disparate results.

Despite the disparity in findings and the ongoing debate over methodology amongst researchers, there is one feature common in all these studies. To date prior research has not explicitly addressed the change in variability associated with discrete changes in alcohol control policies. While recognising the complexity of contemporary alcohol control systems in their assessment of privatisation effects, standard intervention analysis using Autoregressive Integrated Moving Average (ARIMA) models, ignore the dynamic nature of the volatility associated with the intervention. In light of previous studies failing to explicitly account for conditional heteroskedasticity, the misspecification of the volatility structure could possibly have led to erroneous conclusions.

This study focuses on modelling the dynamically changing variability associated with a change in public policy and its effects on inferences about the intervention. Generalised Autoregressive Conditional Heteroskedasticity (GARCH) models, commonly used in financial and econometric applications, are used to characterise the dynamic nature of the volatility. These models provide an additional dimension to standard approaches by accounting for volatility effects and thereby facilitating a means to assess their impact on inference about the intervention. A separate working paper will document the results of all seven states that adopted public policy changes to alcohol distribution. The purpose of this current paper is to present a rigorous analysis of data from wine sales over the last thirty years in a single state. To illustrate the issues, concepts and theory behind GARCH intervention modelling, Alabama was selected because of its two-stage relaxation of monopoly control in retail wine sales.

1.2 Outline of paper

The focus of this paper is to assess the impact of persistent volatility dynamics associated with public policy interventions by comparing two types of univariate time series models: ARIMA and GARCH. ARIMA models consist of three components: lagged values of the variable of interest (the autoregressive component), lagged values of the error term (the moving average component) and the degree of integration (the number of differences required to make the series stationary). GARCH models capture time-varying but persistent volatility and can also include autoregressive and moving average lags in the level of the variable. The key difference between these types of models relate to the specification of the variance of the series. ARIMA models have a constant variance while GARCH models have time-varying variance. By uncovering changes in volatility after an intervention, GARCH models effectively capture 'volatility effects'. This paper addresses whether information about the variance structure of the series improves our inference about public policy interventions, particularly the hypothesised permanent change in alcohol consumption patterns. Essentially, we are interested in answering three key questions: (i) is there evidence of a change in the volatility structure after intervention? (ii) does misspecification of the volatility structure lead to incorrect inferences regarding the intervention? (iii) what impact does modelling the changing volatility structure have on the parameter estimates? Section 2 presents the data used in our study. Section 3 introduces the two intervention models we wish to compare, beginning with a review of traditional ARIMA methods and transfer function modelling. This is followed by the exposition of the main technical innovation used in this paper, GARCH intervention models. Section 4 gives the results of our analysis and presents a comparison of findings obtained under each modelling framework. Section 5 provides concluding remarks.

2 Data

2.1 Dependent variable

The data used in this research are quantities of absolute ethanol sold measured as shipments from wholesalers to retail suppliers. Monthly data on wine sales in Alabama from January 1968 to December 1997 is obtained from the School of Public Health, Division of Epidemiology at the University of Minnesota. Data for the seven intervention states are described in detail in Wagenaar and Holder (1991, 1995). In general, they observe that monthly sales of wine in all states are serially correlated and exhibit clear seasonal patterns. The seasonal pattern is characterised with one major peak and several minor peaks during the year. There is also a suggestion of an increasing trend in time. The amount of seasonal variation also appears to increase with the overall level of the series. Consequently, appropriate pre-differencing transformations are required to control for these seasonal behavioural features and other possible sources of bias. Since the data exhibits increasing seasonal variation with the level of the series, a log-transformation is useful in 'equalising' variability and thereby reducing heteroskedasticity. Additionally, the logarithm metric improves the interpretability of parameter estimates because of its straightforward conversion to percentage changes using the formula $(e^{\omega} - 1)*100$, where ω is the parameter estimate from the time series model.

Per capita measures of wine consumption are also useful for controlling bias arising from changes to legal age requirements within each state throughout the thirty- year period. For instance, some states that originally imposed a drinking age limit of 21 may have lowered its limit to 18 at some stage and then reverted back to 21 at a later date. The use of the drinking age population, instead of total population, ensures that any statutory changes to legal drinking age that may have occurred during the period are accounted for and bias to our measure of wine consumption is minimised. The relevant quantity or dependent variable we wish to analyse after appropriate data transformations becomes the *log of wine sales per capita*. This quantity is measured by dividing the monthly litres of ethanol sold from wine (multiplied by a scale factor of 10) by the drinking age population, and then taking its natural logarithm.

2.2 Privatisation in Alabama

Alabama eliminated its monopoly on wine sales in two phases. In October 1973, three of the most populated counties (Jefferson, Tuscaloosa and Mobile) were permitted to have private outlets sell wine for off-premise consumption. The sale of wine for on-premise consumption was also permitted in Montgomery County, which includes the state capital. This 'partial' privatisation policy represented 31% of the total Alabama population. In October 1980, full privatisation occurred with privatised wine sales permitted in all Alabama counties. Figure 1 displays the log of wine sales per capita in Alabama. The dotted vertical lines represent the two intervention points.

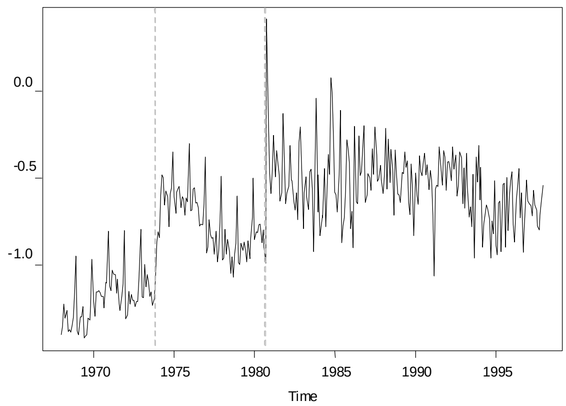

Figure 1: Log of ethanol litres from wine sales per capita

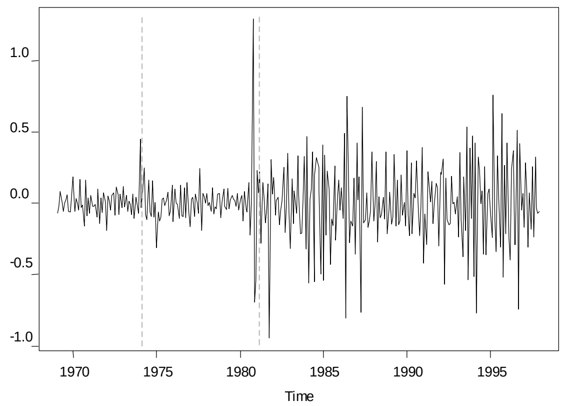

Figure 2: Simple and seasonally differenced values

2.3 Stationary processes

Stationary time series possess mean and covariance functions that are time invariant. By observation alone, the dependent variable shown in Figure 1 can be clearly characterised as non-stationary. Differencing transformations are then required to induce stationarity, a fundamental property in all time series applications. Wagenaar and Holder (1995) found that the appropriate degree of differencing is *simple and seasonal differencing*. Simple and seasonal differencing requires the variable of interest to be double differenced at lags 1 and 12, and is shown in Figure 2. Despite the mean and covariance functions failing to exhibit time invariance, we found no reason to consider higher order differencing. The vertical lines in Figure 1 mark the breaks in variability observed at the two intervention points, with variability increasing after each intervention. The corresponding breaks in the differenced series can also be detected in Figure 2, suggesting that the second public policy change made a more significant impact on observed variability than its predecessor. These two graphs present *prima facie* evidence of a changing volatility structure in the log-transformed series after intervention. While we failed to remove heteroskedasticity completely, the appropriate *order of integration* for ARIMA modelling is established by way of these differencing transformations. Section 3.1 provides further details.

2.4 Design

Intervention studies that have appeared in the literature have typically followed the Box-Jenkins (1976) interrupted time series design. This design treats the elimination of the state retail monopoly on wine as a natural experiment. It involves an examination of observations through time that are hypothesised to be affected by the intervention, and the utilisation of a comparison time series that was expected not to be affected. This comparison series serves to mitigate the effects of nationwide trends and fluctuations that could influence the behaviour of the dependent variable, independent of changes to public policy. Aggregate wine sales from all non-intervention states can be used for this purpose as the comparison series. This quantity is essentially a control variable to be included as one of the predictors in the model. Figure 3 shows the log of per capita wine sales from non-intervention states, hereinafter referred to as the *control covariate*.

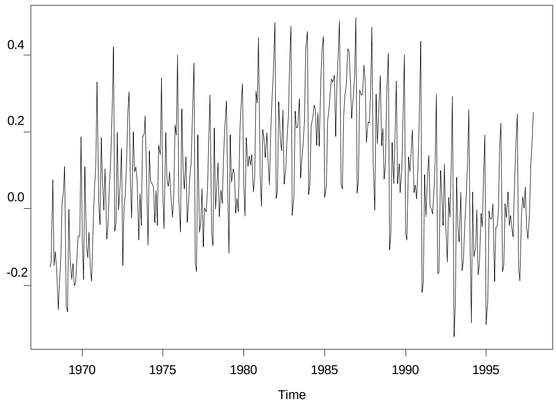

Figure 3: Log of ethanol litres per capita from control states

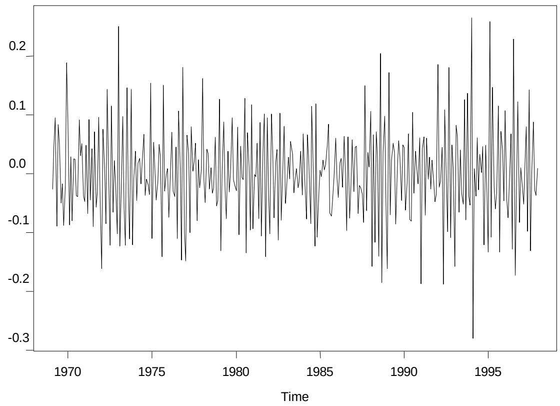

Figure 4: Simple and seasonally differenced values

Since we require the dependent variable to be double differenced for modelling purposes, the control covariate must also be differenced accordingly. Figure 4 shows the simple and seasonally differenced values of the comparison series. There are two striking features that differentiate the comparison series from the response variable. Firstly, the absence of a structural break at either intervention point and secondly, observed variability in Figures 3 and 4 appear approximately constant.

3 Methodology

3.1 ARIMA Models

Multiplicative seasonal ARIMA models are used to account for the dominant seasonal cycles and serial correlation that characterise monthly sales of alcoholic beverages. The general ARIMA$(p,d,q)(P,D,Q)_s$ model for the outcome variable y_t is given by

$$y_t = \frac{(1-\Theta_1 B^s - ... - \Theta_Q B^{sQ})(1-\theta_1 B - ... - \theta_q B^q)u_t + \alpha}{(1-\Phi_1 B^s - ... - \Phi_P B^{sP})(1-\phi_1 B - ... - \phi_p B^p)(1-B^s)^D (1-B)^d} \qquad (1)$$

where p is the order of the autoregressive process, d is the degree of non-seasonal differencing, q is the order of the moving average process, P is the order of the seasonal autoregressive process, D is the degree of seasonal differencing, Q is the order of the seasonal moving average process, s is the seasonal span, Θ_1 to Θ_Q are seasonal moving average parameters, θ_1 to θ_q are regular moving average parameters, Φ_1 to Φ_P are seasonal autoregressive parameters, ϕ_1 to ϕ_p are regular autoregressive parameters, u_t is a random white-noise error component, α is a constant and B is the backshift operator such that $B(X_t)$ equals X_{t-1}.

As most statistical results in time series analysis are conditional upon stationarity properties of the fundamental process, appropriate differencing transformations need to be determined from the outset. Preliminary analysis in Section 2.3 revealed that differencing at lags 1 and 12 is the most appropriate for the Alabama series. The analogous specification of this result in terms of nonseasonal differencing d and seasonal differencing D from Equation (1) is $d = 1$ and $D = 1$.

3.2 Transfer Functions

Intervention modelling requires the application of traditional time series methods such as rational transfer functions. Transfer functions representing the hypothesised effects of an intervention form the essence of our analysis. The need to measure two specific effects of privatisation determined the functional form of these intervention components. While a step function is used to estimate the long-run or permanent change in the mean level associated with privatisation, a pulse function is used to control for transitory initial stocking effects. These temporary stocking effects represent the expected immediate surge in wholesale distribution as new private outlets throughout each state were stocked. The general form of the transfer function is given by

$$y_t = \frac{(\omega_0 - \omega_1 B - \ldots - \omega_s B^s)}{(1 - \delta_1 B - \ldots - \delta_r B^r)} (I_{t-c}) \qquad (2)$$

where ω_0 to ω_s and δ_1 to δ_r specify the manner in which the input or intervention variable I_t influences the output or dependent variable y_t . To test for effects of the intervention, I_t is either a step function with the value zero before the intervention and one thereafter, or a pulse function with the value one for the month in which the intervention begins and zero otherwise; c is a delay parameter indicating the lag between the intervention and the initial effects of the intervention.

The question of how to model the timing and extent of privatisation is critical to our assessment of intervention effects. For ease of exposition, we shall replace the term I_t in Equation (2) with the following notation, s_t for a step function and p_t for a pulse function. Since Alabama implemented two separate interventions, the preliminary ARIMA$(p,d,q)(P,D,Q)_s$ model will typically require two sets of step and pulse functions, denoted $\{s_1, s_2\}$ and $\{p_1, p_2\}$ respectively. In deciding which combination of step and pulse functions are required to reflect the behavioural features of the Alabama series, the following issues are of crucial importance to model construction.

The issue of how to account for partial interventions, as witnessed in the 1973 policy change, must be considered carefully since only 31% of the Alabama population was said to be affected the intervention. While there may be a case for setting $\{s_1 < 1, s_2 = 1\}$ this paper adopts the method of inserting two separate unit step functions at each intervention time, irrespective of the degree of privatisation. This method is consistent with Wagenaar and Holder (1995).

Choosing the appropriate ‘mix’ of transfer functions to be included in the model also depends on timing the true impact of the intervention. For example, a closer inspection of the data reveals that the effects of Alabama's partial intervention on wine sales were not truly realised until three months later. Modelling intervention effects must therefore take this 'delay' into consideration. In Equation (2) this delay parameter is defined as c. The preliminary model will therefore require the analogous specification, $c_1 = 3$ to represent the delay factor of three time periods for s_1 and p_1, should these transfer functions warrant inclusion. Ultimately, the appropriate mix of step and pulse functions, after extent and timing considerations have been taken into account, will rest on the overall model fit and the significance of parameter estimates.

Another important issue to consider in transfer function modelling is the interpretation of parameter coefficients. Equation (2) requires both s_t and p_t to be differenced in accordance with the stationarity requirements of the dependent variable. While the interpretation of the step function is straightforward, with the expression ωs_t effectively capturing the long-term or permanent effects of privatisation, the pulse function is complicated by an additional parameter.

Temporary stocking effects are best modelled with the inclusion of a decay parameter δ. This serves as a method of spreading or decaying the sudden surge in sales across more than one time period. It is necessary because the increase from stocking up inventory levels, in anticipation or reaction to the intervention, is expected to 'spill over' to subsequent months and dissipate uniformly over that period. This approach to modelling stocking effects is again consistent with Wagenaar and Holder (1995). Essentially, the pulse function requires both a numerator and a denominator coefficient, which we denote as υ and δ respectively. In short, pulse effects are best captured by the expression $\left[\upsilon/(1-\delta)\right]p_t$.

3.3 ARIMA Estimation

The SAS software package is used to select the appropriate ARIMA$(p,d,q)(P,D,Q)_s$ model. The model fitting procedure involves several iterations of the identification-estimation-diagnosis process. This model building process ensures that all conventional criteria for ARIMA model adequacy have been met, including statistically significant parameter estimates and the removal of all significant dimensions of the autocorrelation structure in the residuals. The functional form of a typical intervention model can be expressed as a linear combination of Equation (1) plus a specified number of transfer functions in the form of Equation (2) plus any control variables. The parameters of our intervention model, which includes the ARIMA components, an appropriate mix of step and pulse functions, and the control covariate, are estimated simultaneously using maximum likelihood methods.

3.4 GARCH Review

GARCH models were originally proposed to model financial time series. It is observed in many financial time series that although the residuals obtained after modelling the mean process are uncorrelated, the squared residuals are highly correlated. This implies that the variance of the time series is dependent on past observations. To capture this phenomenon of persistent volatility, Engle (1982) developed a new class of stochastic processes called autoregressive conditional heteroskedasticity (ARCH). These are zero mean, serially uncorrelated processes with non-constant variances conditional on the past, but constant unconditional variances. For such processes the recent past offers information about the 'one-period ahead' volatility forecast. The ARCH model was extended to the generalised autoregressive conditional heteroskedasticity (GARCH) model by Bollerslev (1986) to allow the current conditional variance to be a function of past conditional variances as well as the past error terms derived from the mean process. These models facilitate the capture of time-dependent volatility through a conditional variance equation, while modelling the mean process simultaneously.

3.5 Event-Study Literature

Although the public health literature has witnessed the extensive use of ARIMA models, we are not aware of any empirical work specific to the area of alcohol consumption that has combined GARCH methodology with intervention analysis. In the finance literature however, the GARCH modelling strategy still receives widespread attention, and more recently *event study* approaches have come to the fore. Boehmer et al. (1991) cite that classical event-study methodology, which ignores event-induced volatility, tends to "reject the null hypothesis of zero abnormal returns too frequently when it is true". In response to this bias towards detecting 'effects' regardless of whether such effects actually exist, Brockett et al. (1999) developed an alternative approach that explicitly incorporates stochastic behaviours of the market, assumed away by classical methods. Batchelor and Orakcioglu (2003), Wang et al. (2002) and Corhay and Rad (1996) apply similar GARCH modelling techniques to estimate not only the effects of a particular economic event on stock market returns, but also to uncover the impact of event-induced volatility on inferences about that event.

These studies found that the inclusion of GARCH effects can produce considerably different parameter estimates in comparison to classical approaches, which may very well cause researchers to draw inappropriate conclusions about the economic event in question. In fact, Brockett et al. found that explicitly modelling event-induced volatility in relation to a well-known regulatory change (specifically, the ratification of California's Proposition 103, mandating substantial changes to the state's automobile insurance market) resulted in exactly the opposite conclusion to three previous empirical studies that used standard event study techniques, including Fields et al. (1990). They suggest that the source of inconsistency is due to the inclusion of volatility effects. While event studies examine the impact of 'events' and not interventions per se, recent GARCH extensions to event-study approaches emulate our proposed intervention techniques. This provides the impetus behind our choice of modelling strategy despite apparent lack of precedence in the public health literature.

3.6 GARCH Models

The conditional variance equation characterised by a GARCH model of orders *m* and *n*, denoted GARCH*(m,n)* is generated by the following recursion

$$\sigma_t^2 = \gamma + \sum_{j=1}^{m} a_j \varepsilon_{t-j}^2 + \sum_{k=1}^{n} b_k \sigma_{t-k}^2 \tag{3}$$

where $\sigma_t \mid \Omega_t \sim N(0,\sigma_t^2)$ and Ω_t is the information set at time *t* on which the distribution of the errors is assumed to be conditioned, γ is a constant, $a_1,\ldots,a_m$ and $b_1,\ldots,b_n$ are parameters to be estimated. Equation (3) implies that the conditional variance σ_t^2 at time *t* depends on both past conditional variances σ^2_{t-k} and recent squared innovations ε^2_{t-j}. In other words, the conditional variance is determined by its previous realisations and the squared error terms from the conditional mean equation. The ARCH parameters a_j for $j = 1,\ldots,m$ measure the short-term persistence of a shock, while the GARCH parameters b_k for $k = 1,\ldots,n$ capture longer-term effects.

However it might be expected that some other variables z_i are influential on the behaviour of conditional variance, for example, the intervention itself. The conditional variance equation must therefore adjust for possible influences arising from exogenous variables in the mean process such as the control variable and/or the transfer functions. To reflect this 'noise' component, Equation (3) can be modified such that

$$\sigma_t^2 = \gamma + \sum_{j=1}^{m} a_j \varepsilon_{t-j}^2 + \sum_{k=1}^{n} b_k \sigma_{t-k}^2 + \sum_{i=1}^{h} \lambda_i z_{it} \tag{4}$$

where each λ_i are parameter coefficients corresponding to each exogenous variable z_i included in the recursion. The choice of exogenous variables to be included in the conditional variance equation is based on a stepwise examination of each variable's contribution to improving model performance.

3.7 GARCH Estimation

A GARCH intervention model that allows for volatility effects is simply an ARIMA intervention model with an additional conditional variance component. The intervention model retains its mean equation components but is augmented by a conditional variance equation characterised by a GARCH*(m,n)* model. In determining the appropriate orders of ARCH and GARCH terms, *m* and *n* respectively, we begin with parsimonious models such as the ARCH(1) or GARCH(1,1) process. Higher order models are then estimated and compared to lower order specifications using standard model selection criteria such as Akaike information criteria (AIC) and Bayesian information criteria (BIC), as well as other diagnostic checks to be discussed in the following subsection. The optimal GARCH*(m,n)* model is then chosen among a variety of different specifications.

GARCH intervention models are estimated using the iterative BHHH algorithm, a nonlinear optimisation technique in the S-Plus FinMetrics module used to compute maximum likelihood estimates. Convergence is reached within the specified maximum number of iterations (50) and reliable coefficient estimates and their corresponding standard errors are obtained.

3.8 Diagnostic Tests

Apart from checking autocorrelation function plots for the absence of serial correlation, formal diagnostic tests based on standardised residuals need to be performed in order to test the validity of the models. Let the standardised residuals be defined as

$$e_t = (\hat{\sigma}_t)^{-1}\varepsilon_t \tag{5}$$

where $\hat{\sigma}_t$ is the estimated conditional standard deviation given by Equation (4) with parameter estimates replacing model parameters. If the GARCH model is correctly specified, then these standardised residuals should be uncorrelated and should follow an independent and identically distributed (IID) sequence. The Ljung-Box (1978) portmanteau test for serial correlation is used to assess the adequacy of these models. The Ljung-Box Q-statistic is given by

$$Q_{LB} = N(N+2)\sum_{l=1}^{r}(N-l)^{-1}\hat{\rho}^2(l) \tag{6}$$

where $\hat{\rho}$ is the sample autocorrelation at lag l computed from the residuals and N is the number of observations. Under the null hypothesis that the model is correctly specified, Q_{LB} is asymptotically distributed as χ^2 with $(r-k)$ degrees of freedom, where k is the number of independent parameters. A small value of Q suggests that the sample autocorrelations of the standardised residuals are small enough to support the IID hypothesis. The adequacy of our model thereby rests on the 'acceptance' of the null hypothesis at the 5% level of significance.

Another portmanteau test, formulated by McLeod and Li (1983), is used to check that the squared standardised residuals are uncorrelated. The Q_{ML} test is based on the same statistic used for the Ljung-Box test except that the sample autocorrelations of the standardised residuals are replaced by the sample autocorrelations of the squared standardised residuals.

To check the assumption of conditional normality of the residuals, a visual inspection of the quantile-quantile plots is carried out. The Shapiro-Wilk test is also used to formally check the distributional properties of the residuals. The test statistic for this normality test is given by

$$W = \frac{\left(\sum_{i=1}^{N} h_i e_{(i)} \right)^2}{\sum_{i=1}^{N} (e_{(i)} - \bar{e})^2} \qquad (7)$$

where $e_{(i)}$ is the i^{th} order statistic of the sample, $\bar{e}$ is the sample mean and N is the sample size. Royston (1982) provide approximations and the tabled values which may be used to compute the coefficients h_i, $i = 1,\ldots,N$ and obtain the p-value of the W-statistic. The adequacy of our model again rests on the acceptance of the null hypothesis of normally distributed standardised errors.

4 Results

4.1 ARIMA

Identification

Preliminary model runs reveal that the inclusion of a pulse function at the first intervention point is unnecessary. Estimates of both numerator and denominator coefficients, υ_1 and $\delta_{1,}$ are invariably insignificant regardless of which ARIMA specification is used. This suggests that there were no observable stocking effects from the 1973 partial policy change and therefore only the step function s_1 is required to explain the associated rise in wine consumption at this intervention point.

Estimates of the decay parameter at the second intervention point δ_2 are also consistently insignificant. This tends to suggest that the sudden surge in sales that was observed in October 1980 did not 'spill over' to subsequent months as expected. In other words, stocking effects from the 1980 intervention failed to dissipate over time and the surge was realised only for that particular month when the policy change was enforced. However, the insignificance of this parameter could also be attributed to possible non-uniformities in the decay pattern of stocking effects. The decay mechanism in our model may not be capable of explaining such irregularities. For modelling purposes δ_2 will also be excluded from the ARIMA model.

The preliminary ARIMA model therefore includes two step functions $\{s_1, s_2\}$, one pulse function without decay p_2, and the control covariate denoted as x. Note that the constant term α from Equation (1) has also been omitted due to insignificance. Several model specifications are identified as possible candidates for selection, but ultimately the ARIMA(0,1,1)$(0,1,1)_{12}$ is chosen as the superior model.

Estimation

Recall that y_t is the log-transformed series. Define $Y_t = \nabla_{12}\nabla y_t$, where ∇ denotes the difference operator. By letting $\{U_t\} \sim WN(0, \sigma^2)$, the ARIMA(0,1,1)$(0,1,1)_{12}$ model for Alabama can then be expressed as follows

$$Y_t = U_t + \theta_1 U_{t-1} + \Theta_1 U_{t-12} + \theta_1 \Theta_1 U_{t-13} + \beta X_t + \omega_1 S_t(T_1 + 3) + \omega_2 S_t(T_2) + \upsilon_2 P_t(T_2) + \varepsilon_t \quad (8)$$

where $X_t = \nabla_{12}\nabla x_t$, $S_t = \nabla_{12}\nabla s_t$, $P_t = \nabla_{12}\nabla p_t$; T_1 and T_2 are the two intervention time points; θ_1 and Θ_1 are the moving average coefficients where $\theta_1 \in [-1,1]$ and $\Theta_1 \in [-1,1]$; ω_1 and ω_2 are intervention parameters for each long-term privatisation effect; υ_2 controls for stocking effects associated with the second intervention point; β controls for variance in wine sales from non-intervention states and ε_t are the error terms. Table 1 shows the parameter estimates for the ARIMA(0,1,1)(0,1,1)$_{12}$ model.

Parameter	Estimate	Std. Error	t-statistic	p-value
θ_1	-0.845	0.029	29.10	<.001
Θ_1	-0.732	0.041	18.04	<.001
β	0.450	0.132	3.40	0.001
ω_1	0.510	0.072	7.06	<.001
ω_2	0.379	0.075	5.03	<.001
υ_2	0.704	0.132	5.32	<.001

Table 1: Maximum likelihood estimates for the ARIMA intervention model

The model equation is given by

$$\hat{Y}_t = U_t - .84U_{t-1} - .73U_{t-12} + .62U_{t-13} + .45X_t + .51S_t(T_1+3) + .38S_t(T_2) + .70P_t(T_2) \quad (9)$$

Diagnostics

The parameter estimates for the ARIMA model are all highly significant and are therefore considered reliable. The model appears to have performed adequately in terms of predictive power, with a very small variance estimate, $\hat{\sigma} = 0.021$. Residuals analysis however reveals that the variability is not constant through time.

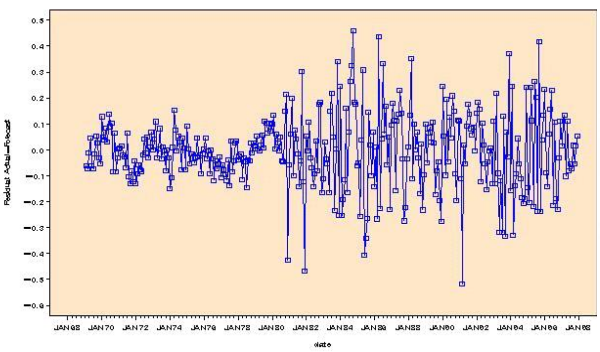

Figure 5: Plot of the residuals extracted from ARIMA model

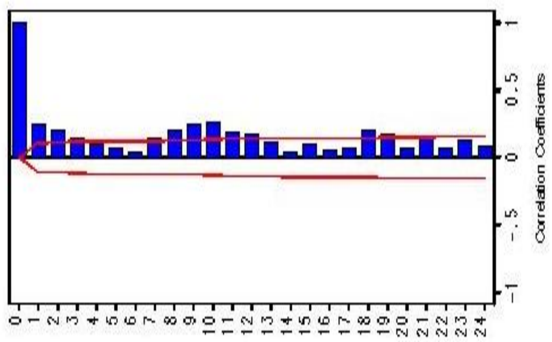

Figure 6: ACF plot of ARIMA model squared residuals

Figure 5 clearly shows the lack of homogeneity of variance in the residuals. The observed break in the variability of the residuals, which occurs around the second intervention point, highlights the dramatic difference in variability levels between the two periods. The presence of these heteroskedastic errors is a clear deficiency of the ARIMA approach. However, the case for modelling volatility effects by way of GARCH methods is strengthened further by the analysis of squared residuals. Figure 6 displays the autocorrelation function (ACF) plot of the squared residuals extracted from the ARIMA model. This plot provides overwhelming evidence that ARIMA modelling fails to remove serial dependence, with significant correlations observed at lower lags from 1 to 4, and at higher seasonal lags from 7 to 12. This is a serious shortcoming of the ARIMA approach and questions regarding the accuracy of any inference made from such methods must be raised.

4.2 GARCH

Identification

GARCH intervention models that incorporate the volatility structure of the underlying series extend the scope of standard models by way of a conditional variance equation. These models combine observed volatility effects with the mean process specifications obtained under ARIMA modelling. The GARCH modelling procedure thereby combines the optimal ARIMA model identified in Equation (8) with the conditional variance formula given in Equation (4). The choice of appropriate ARCH and GARCH terms to include in the conditional variance equation thus becomes our primary objective. However, the appropriate mix of noise variables from the mean process to be included in the variance equation must also be determined. Preliminary model runs reveal that only the step function at the second intervention point s_2 improves overall model fit, and thus only one exogenous variable z_t is required. To illustrate the technical aspects of our modelling strategy, we identify the following models, namely the GARCH(1,1), ARCH(2) and GARCH(2,1) as possible candidates for selection. A comparison of overall model fits, significance of parameter estimates and the relevant diagnostic checks show that the ARCH(2) model best characterises the changing volatility structure observed in the Alabama series. Details of the two alternative GARCH models and model selection criteria can be found in the Appendix.

Estimation

An ARCH(2) model that includes the noise parameter can be represented by

$$\sigma_t^2 = \gamma + a_1\varepsilon_{t-1}^2 + a_2\varepsilon_{t-2}^2 + \lambda_1 z_{1t} \tag{10}$$

Equation (10) implies that conditional variance is a deterministic function only of past error terms derived from the mean process, and the noise variable. Previous realisations of conditional variance have no influence on the current variance equation in an ARCH(2) process. The absence of GARCH parameters suggests that the shock generated only short-term persistence with no lingering effects on volatility. Table 2 presents the parameter estimation results for our final model.

Parameter	Estimate	Std. Error	t-statistic	p-value
θ_1	-0.686	0.043	-15.84	<.001
Θ_1	-0.792	0.036	-22.03	<.001
β	0.455	0.076	5.96	<.001
ω_1	0.427	0.040	10.73	<.001
ω_2	0.298	0.038	7.92	<.001
υ_2	0.762	0.035	21.53	<.001
γ	0.003	0.001	4.50	<.001
a_1	0.299	0.087	3.45	0.001
a_2	0.237	0.077	3.08	0.002
λ_1	0.015	0.003	4.31	<.001

Fit Statistics	**Value**	
AIC	-477.56	
BIC	-435.22	
Test Statistics	**Value**	**p-value**
W test for normality	0.992	0.976
Q_{LB} test for standardised residuals	19.47	0.078
Q_{ML} test for squared std. residuals	17.79	0.122

Table 2: Parameter estimation output for the GARCH intervention model

The final model is given by

$$\hat{Y}_t = U_t - .69U_{t-1} - .79U_{t-12} + .54U_{t-13} + .45X_t + .43S_t(T_1+3) + .30S_t(T_2) + .76P_t(T_2)$$

$$\hat{\sigma}_t^2 = .003 + .299\varepsilon_{t-1}^2 + .237\varepsilon_{t-2}^2 + .015z_{1t} \tag{12}$$

Diagnostics

Figure 7 depicts the conditional standard deviations from the ARCH(2) model. It is not surprising that a substantial 'jump' in conditional variance is observed after the second intervention point. This is consistent with an earlier observation from Figure 5 where ARIMA modelling failed to adequately explain the substantial break in variability observed in the residuals series. The conditional standard deviations appear to follow a 'step' pattern of their own, with the post-October 1980 period clearly exhibiting higher levels of variability. However, the true worth of GARCH intervention models over current methods is made apparent by the residuals analysis. Figure 8 shows the standardised residuals extracted from the final model. Unlike Figure 5, this residuals plot displays homoskedastic or constant variance errors.

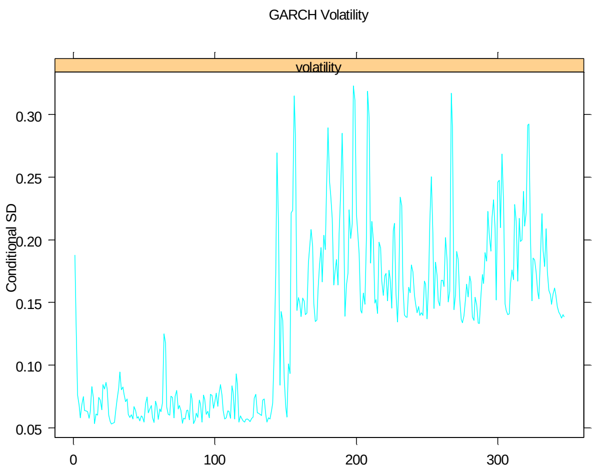

Figure 7: Conditional standard deviations from final model

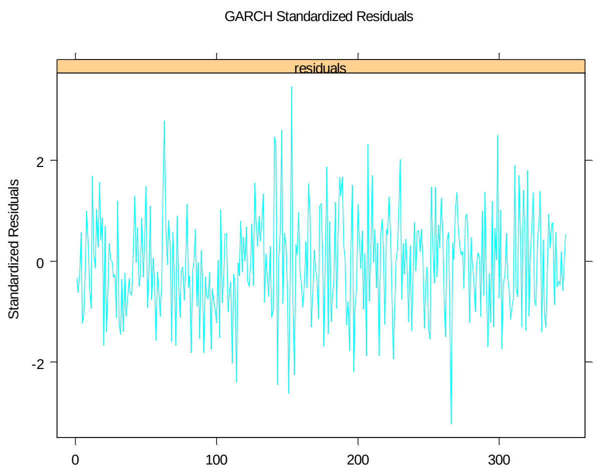

Figure 8: Standardised residuals extracted from final model

We also find that the Ljung-Box test (Q_{LB} = 19.47, p-value = 0.08) is passed comfortably at the 5% level of significance, suggesting that the standardised residuals are not serially correlated. In other words, we accept the null hypothesis that the standardised residuals follow an IID sequence. However, the true efficacy of our approach to modelling interrupted time series is best encapsulated by the ACF plot shown in Figure 9. This plot suggests that the final model has delivered uncorrelated squared standardised residuals. This result is in stark contrast to the corresponding ACF plot obtained under the ARIMA model shown earlier in Figure 6. The McLeod-Li test (Q_{ML} = 17.79, p-value = 0.12) formally confirms the validity of our model. That is, there is no sufficient evidence to suggest that serial correlation persists in the squared standardised residuals after applying GARCH techniques.

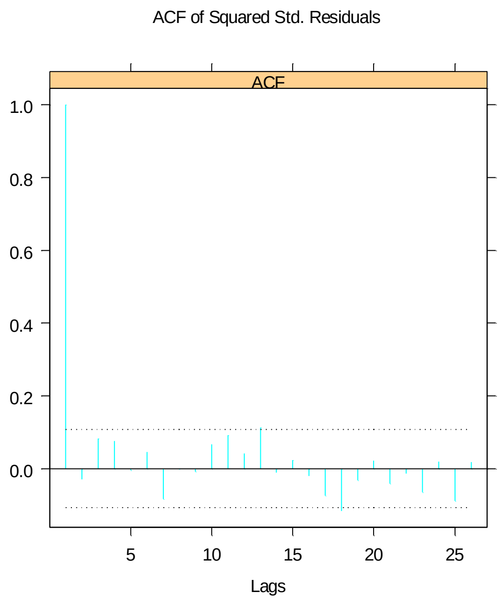

Figure 9: ACF plot of squared standardised residuals

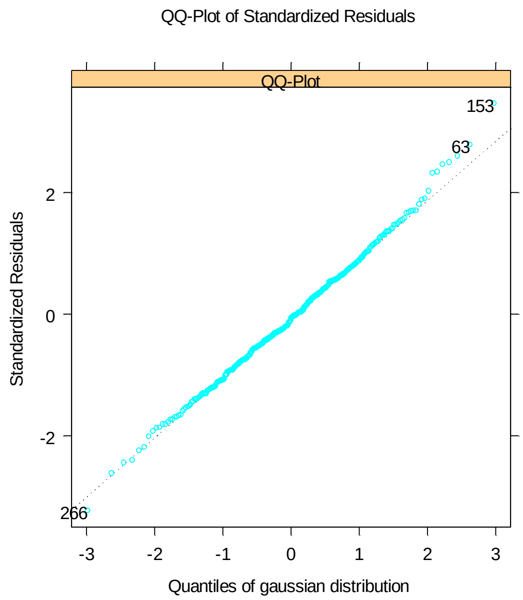

Figure 10: Quantile-quantile plot of standardised residuals

Figure 10 shows the QQ-plot of the standardised residuals. This plot suggests that the residuals have satisfied the normality assumption, with the greater majority of residuals behaving linearly or lying along the normal probability line. The Shapiro-Wilk statistic (W = 0.992, p-value = 0.976) verifies that we cannot reject the null hypothesis of normally distributed standardised errors at all conventional levels of significance. Non-Gaussian models such as the t-distribution, typically used to account for kurtosis or heavy-tailedness in the residuals, are therefore unnecessary for the Alabama series.

4.3 Comparative Analysis

Parameter estimates and standard errors

Table 3 compares parameter estimates and standard errors obtained under ARIMA and GARCH models. To determine the impact of volatility effects, we need to examine whether coefficient estimates, particularly for the intervention variables, differ significantly between the two approaches. The ARIMA model tends to overestimate the long-term permanent effects of both policy changes ω_1 and ω_2. On the other hand, the temporary stocking effect υ_2 is underestimated when volatility effects are ignored. Additionally, the GARCH approach produced smaller standard errors in general and thus possessed greater estimation efficiency than its counterpart.

	Coefficient Estimate			Standard Errors	
Parameter	***ARIMA***	***GARCH***		***ARIMA***	***GARCH***
θ_1	-0.845	-0.686		0.029	0.043
Θ_1	-0.732	-0.792		0.041	0.036
β	0.450	0.455		0.132	0.076
ω_1	0.510	0.427		0.072	0.04
ω_2	0.379	0.298		0.075	0.038
υ_2	0.704	0.762		0.132	0.035
γ		0.003			0.001
a_1		0.299			0.087
a_2		0.237			0.077
λ_1		0.015			0.003

Table 3: Comparison of parameter estimation results between ARIMA and GARCH models

Permanent effects of privatisation

Tables 4 summarises the estimated permanent change in wine consumption under different modelling frameworks. Results from Wagenaar and Holder (1995) are also reported to highlight estimation differences arising from increased availability of data. Their study adopted the ARIMA(0,1,8)(0,1,1)$_{12}$ model to analyse data from 1968 to 1991. Our study improves the balance of the design with an additional six years of post-intervention data, thereby extending the long-term follow-up of privatisation effects.

1973 change					
Model	**Estimate**	**Standard Error**	**Percentage Change**	**95% Confidence Interval**	
				Lower	**Upper**
W & H	0.144	0.107	15.5	-6.2	42.4
ARIMA	0.510	0.072	66.5	44.5	91.8
GARCH	0.427	0.040	53.3	41.8	65.7
1980 change					
Model	**Estimate**	**Standard Error**	**Percentage Change**	**95% Confidence Interval**	
				Lower	**Upper**
W & H	0.350	0.115	42.0	13.4	77.7
ARIMA	0.379	0.075	46.0	26.0	69.2
GARCH	0.298	0.038	34.7	25.1	45.0

Table 4: Comparison of estimated permanent effects using alternative methods

At the 1973 intervention point, Wagenaar and Holder report a 15.5% permanent increase in consumption levels. This estimate however is statistically insignificant at the 5% level, implying the absence of an effect. This is evidenced by the inclusion of the null result within the limits of the 95% confidence interval. In contrast, our estimates for the first step function parameter are both highly significant. The ARIMA model estimates a 66.5% long-term rise in wine consumption associated with the policy change. By adjusting for volatility effects, there is a smaller resultant increase. The GARCH model estimates a sustained increase of 53.3%, a substantial difference of 13.2% in comparison to the ARIMA model. Similarly, results from the 1980 intervention suggest that ARIMA approaches tend to overestimate the long-term change in alcohol consumption patterns. Wagenaar and Holder's estimate of 42.0% is considerably close to our ARIMA estimate of 46.0%. The correction for GARCH properties results in a smaller permanent increase of 34.7%, a difference of 11.3%.

5 Conclusion

Significantly increased wine sales following privatisation is an empirical result that has been acknowledged by many impact assessment studies found in the literature. However previous analyses have yet to explicitly consider the changing volatility patterns observed in these interrupted time series experiments. Modelling strategies that aim to capture the true impact of an intervention should therefore reflect the dynamically changing and time-dependent behaviour of volatility. By allowing for nonlinear intertemporal dependence in the residuals by way of GARCH methods, impact assessment of public policy changes serves only to benefit from the additional information derived from modelling volatility effects. Our analysis of the Alabama wine series shows that volatility effects significantly change our inferences about the impact of privatisation on alcohol consumption. Specifically, GARCH intervention models reduce the magnitude of the estimated permanent increase in consumption. Conversely, standard intervention analysis tends to overestimate the hypothesised long-term change. Results also show that our modelling strategy produces lower standard errors in general, thereby improving estimation efficiency in comparison to the ARIMA approach. These findings emphasise the relevance of GARCH methods in identifying and quantifying potential biases that can arise from misspecification of the volatility structure.

While the use of GARCH models has received limited attention outside the financial domain, this study demonstrates its potential utility to public health settings where similar variance anomalies have appeared, particularly in the area of alcohol consumption. By augmenting standard intervention models to accommodate for volatility effects, we are able to draw higher quality conclusions from a more rigorous analytical framework and offset some of the inadequacies of traditional approaches. Consequently, we have shown that the reliability of our inferences about intervention effects improves significantly, in terms of both accuracy and efficiency. Policymakers concerned about the many social and health repercussions of high drinking levels in society should have at their disposal the most reliable estimates of public policy outcomes. The methodological innovation used in this paper advances our understanding of these outcomes in the alcohol industry, but aims to ultimately deliver similar gains to other industries or regulatory bodies planning to implement public policy change.

Appendix

GARCH(1,1): $\sigma_t^2 = \gamma + a_1\varepsilon_{t-1}^2 + b_1\sigma_{t-1}^2 + \lambda_1 z_{1t}$

Test Statistics	**Value**	**p-value**
Shapiro-Wilk test for normality	0.993	0.995
Q_{LB} test for standardised residuals	19.15	0.085
Q_{ML} test for squared std. residuals	38.61	0.000

Table A: Diagnostic test statistics for the GARCH(2,1) model

$$\hat{Y}_t = U_t - .68U_{t-1} - .79U_{t-12} + .54U_{t-13} + .47X_t + .43S_t(T_1+3) + .87S_t(T_2) + .18P_t(T_2)$$

$$\hat{\sigma}_t^2 = .001 + .418\varepsilon_{t-1}^2 + .396\sigma_{t-1}^2 + .007z_{1t}$$

While the GARCH(1,1) specification produces significant estimates for all parameters and obeys the fundamental stationarity condition $a_1 + b_1 < 1$, it fails to adequately remove the serial dependence in the squared standardised residuals (Q_{ML} = 38.61, p-value = 0.00). This is considered a serious violation and therefore the GARCH(1,1) model is deemed inappropriate for the Alabama case.

GARCH(2,1): $\sigma_t^2 = \gamma + a_1\varepsilon_{t-1}^2 + a_2\varepsilon_{t-2}^2 + b_1\sigma_{t-1}^2 + \lambda_1 z_{1t}$

Test Statistics	**Value**	**p-value**
Shapiro-Wilk test for normality	0.992	0.977
Q_{LB} test for standardised residuals	19.22	0.083
Q_{ML} test for squared std. residuals	20.62	0.056

Table B: Diagnostic test statistics for the GARCH(2,1) model

$$\hat{Y}_t = U_t - .70U_{t-1} - .78U_{t-12} + .54U_{t-13} + .43X_t + .46S_t(T_1+3) + .32S_t(T_2) + .74P_t(T_2)$$

$$\hat{\sigma}_t^2 = .002 + .224\varepsilon_{t-1}^2 + .242\varepsilon_{t-2}^2 + .238\sigma_{t-1}^2 + .010z_{1t}$$

The GARCH(2,1) model is considered superior to the GARCH(1,1) model because the assumption of no serial correlation in the squared standardised residuals is satisfied at the 5% level of significance, albeit marginally (Q_{ML} = 20.62, p-value = 0.06). The stationarity condition $a_1 + a_2 + b_1 < 1$ is also upheld. However, the insignificance of the GARCH coefficient ($\hat{b}_1 = 0.238$, p-value = 0.201) precludes the GARCH(2,1) from being selected the optimal conditional variance specification.

Model Selection:

Model	**AIC**	**BIC**
GARCH(1,1)	-465.88	-423.54
ARCH(2)	-477.56	-435.22
GARCH(2,1)	-480.20	-434.01

Table C: Comparison of fit statistics between different GARCH models

To confirm our choice of optimal GARCH model, a comparison of information criteria statistics needs to be performed. Models with the smallest AIC and BIC values outperform other model specifications. Table C verifies the superiority of the ARCH(2) model over competing models in terms of BIC. Even though the GARCH(2,1) model has the minimum AIC value, the significance of all ARCH(2) parameter estimates and its unequivocal removal of serial dependence in the squared residuals series ensure the validity of our final model selection.

REFERENCES

Batchelor, R., and I. Orakcioglu, “Event-related GARCH: the impact of stock dividends in Turkey,” *Applied Financial Economics* 13, 2003, pp. 295-307

Boehmer, E., J. Masumeci, and A.B. Poulsen, “Event-study methodology under conditions of event-induced variance,” *Journal of Financial Economics* 30, 1991, pp. 253-272

Bollerslev, T., “Generalized Autoregressive Conditional Heteroskedasticity,” *Journal of Econometrics* 31, 1986, pp. 307-327

Box, G.E.P, and G.M. Jenkins, *Time Series Analysis: Forecasting and Control,* 1976, revised edition, San Francisco, Holden-Day

Brockett, P.L., H.W. Chen, and J.R. Garven, “A new stochastically flexible event methodology with application to Proposition 103,” *Insurance: Mathematics and Economics* 25, 1999, pp. 197-217

Corhay A., and A.T. Rad, “Conditional Heteroskedasticity Adjusted Market Model and an Event Study,” *Quarterly Review of Economics and Finance* 36, 1996, pp. 529-538

Engle, R., “Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of United Kingdom Inflation,” *Econometrica* 50, 1982, pp. 987-1006

Fields, J.A., C. Ghosh, D.S. Kidwell and L.S. Klein, “Wealth effects of regulatory reform,” *Journal of Financial Economics* 28, 1990, pp. 233-250

Ljung, G.M., and G.E.P. Box, “On a measure of lack of fit in time series models,” *Biometrika* 65, 1978, pp. 297-303

Macdonald, S., “The impact of increased availability of wine in grocery stores on consumption: Four case histories,” *British Journal of Addiction* 81, 1986, pp. 381-387

McLeod, A.I., and W.K. Li, “Diagnostic checking ARMA time series models using squared-residual autocorrelations,” *Journal of Time Series Analysis* 4, 1983, pp. 269-273

Mulford, H.A., and J.L. Fitzgerald, “Consequences of increasing off-premise wine outlets in Iowa,” *British Journal of Addiction* 83, 1988, pp. 1271-1279

Royston, J.P., “An extension of Shapiro’s and Wilk’s W test for normality to large samples,” *Applied Statistics* 31, 1982, pp. 115-124

Smart, R.G., “The impact on consumption of selling wine in grocery stores,” *Alcohol & Alcoholism* 21, 1986, pp. 233-236

Wagenaar, A., and H. Holder, “A change from public to private sale of wine: Results from natural experiments in Iowa and West Virginia,” *Journal of Studies on Alcohol* 52, 1991, pp. 162-173

Wagenaar, A., and H. Holder, “Changes in alcohol consumption resulting from the elimination of retail wine monopolies: Results from five U.S. States,” *Journal of Studies on Alcohol* 56, 1995, pp. 566-572

Wagenaar, A., and J. Langley, “Alcohol licensing system changes and alcohol consumption: introduction of wine into New Zealand grocery stores,” *Addiction* 90, 1995, pp. 773-783

Wang, Z., V. Salin, N. Hooker, and D. Leatham, “Stock market reaction to food recalls: a GARCH application,” *Applied Economics Letters* 9, 2002, pp. 979-987