Another Look at Models of the Short-Term Interest Rate

Author(s): Robin J. Brenner, Richard H. Harjes and Kenneth F. Kroner

Source: *The Journal of Financial and Quantitative Analysis*, Vol. 31, No. 1 (Mar., 1996), pp. 85-107

Published by: Cambridge University Press on behalf of the University of Washington School of Business Administration

Stable URL: http://www.jstor.org/stable/2331388

Accessed: 18-08-2017 14:49 UTC

Another Look at Models of the Short-Term Interest Rate

Robin J. Brenner, Richard H. Harjes, and Kenneth F. Kroner*

Abstract

The short-term rate of interest is fundamental to much of theoretical and empirical finance, yet no consensus has emerged on the dynamics of its volatility. We show that models which parameterize volatility only as a function of interest rate levels tend to over emphasize the sensitivity of volatility to levels and fail to model adequately the serial correlation in conditional variances. On the other hand, serial correlation based models like GARCH models fail to capture adequately the relationship between interest rate levels and volatility. We introduce and test a new class of models for the dynamics of short-term interest rate volatility, which allows volatility to depend on both interest rate levels and information shocks. Two important conclusions emerge. First, the sensitivity of interest rate volatility to interest rate levels has been overstated in the literature. While this relationship *is* important, adequately modeling volatility as a function of unexpected information shocks is also important. Second, we conclude that the volatility processes in many existing theoretical models of interest rates are misspecified, and suggest new paths toward improving the theory.

I. Introduction

The short-term riskless rate of interest (r) is fundamental to much of theoretical and empirical finance. Yet no consensus has emerged on the dynamics of either the level or the volatility of the interest rate. In this paper, we compare and evaluate two popular classes of empirical models for short-term interest rate volatility, and propose an alternative class that overcomes some of the empirical weaknesses of the existing classes.

The first models we evaluate belong to the class of continuous time models in which volatility is parameterized only as a function of interest rate levels. We will refer to these as LEVELS models. Several continuous time models, such

*Brenner, Merrill Lynch, World Financial Center, North Tower, 15th Floor, New York, NY 10281; Harjes, Advanced Strategies and Research Group, Wells Fargo Nikko Investment Advisors, 45 Fremont Street, San Francisco, CA 94105; and Kroner, both the University of Arizona (departments of economics and finance) and Advanced Strategies and Research Group, Wells Fargo Nikko Investment Advisors, 45 Fremont Street, San Francisco, CA 94105. The authors acknowledge insightful comments from Torben Andersen, Dave Backus, Peter Bossaerts, Frank Diebold, Ed Kane, Andrew Karolyi, Ieuan Morgan, Mark Wohar, Jeff Wooldridge, Wayne Ferson (JFQA associate editor and referee), seminar participants at Bentley College, University of Arizona, UC Davis, the 1994 FMAs, and the UC San Diego Conference on Multivariate Financial Modeling. The authors also thank Steve Pince and Craig Stephenson for research assistance. Of course, all errors are the authors' responsibility.

as Merton (1973), Brennan and Schwartz (1980), and Cox, Ingersoll, and Ross (1985), are examples of LEVELS models. The second class of models we evaluate, the GARCH models, parameterizes volatility only as a function of unexpected shocks to the interest rate market. In these models, originally proposed by Engle (1982) and Bollerslev (1986), volatility is not a function of interest rate levels. In this paper, we discuss some of the weaknesses of both LEVELS and GARCH models of the short-term interest rate, and propose a new class of models that treats volatility as a function of both the interest rate level and unexpected interest rate shocks.

Which of the three classes of models best captures the dynamics of short-term rates is important for several reasons. To illustrate, the short rate is often used as a proxy for the latent state variable, driving the changes in the entire term structure. This makes the choice of a model for short-term rates crucial to pricing bonds, pricing interest rate derivatives, and hedging interest rate risk. For example, to price long-dated derivatives such as swaps, swaptions, or embedded bond options correctly, we need to model both the instantaneous and time series properties of interest rate volatility. This follows because both the current level of volatility and the stochastic properties of volatility will affect dynamic hedge ratios and the distribution of future interest rate levels, which determine a derivative's price. The use of an incorrect model could lead to incorrect inferences, mishedged or unhedged risks, or pricing errors.

We demonstrate that if both interest rate levels and unexpected shocks are allowed to affect volatility, the sensitivity of volatility to levels diminishes substantially. Therefore, we conclude that the extreme sensitivity of volatility to levels found in the existing literature is a result of model misspecification. We also show that our proposed models perform much better statistically than either the LEVELS or the GARCH models. These results call into question the validity of existing empirical modeling techniques for the short-term interest rate, and suggest a strategy for developing a new generation of models.

The rest of the paper is organized as follows. Section II reviews the two classes of models that we consider and introduces our proposed empirical generalization. Section III discusses the robust conditional moment test statistics used in this paper. Section IV describes the data, and Section V presents our results. Section VI concludes.

II. Models of Short-Term Interest Rates

A. Two Existing Classes of Volatility Models

To analyze the variety of models for the short rate and its volatility, Chan, Karolyi, Longstaff, and Sanders (1992) (hereafter CKLS), uses a generalized continuous time short rate specification,

$$dr = (\alpha + \beta r)\,dt + \psi r^{\gamma}\,dW, \tag{1}$$

where r is the interest rate level, t is time, W is a Brownian motion, and α, β, ψ, and γ are parameters. CKLS shows that with appropriate restrictions on α, β, and γ, many popular interest rate models can be obtained. For example, setting

$\gamma = ½$ gives the Cox, Ingersoll, and Ross (1985) square root model, and setting $\alpha = 0$ gives the Cox (1975) constant elasticity of variance model. In these models, $\alpha + \beta r$ is the drift and $\psi^2 r^{2\gamma}$ is the variance of unexpected interest rate changes. Also, rewriting $\alpha + \beta r$ as $\beta(r - \alpha^*)$ reveals that β can be viewed as a measure of the speed of mean reversion in rate levels. The more negative β is, the faster r responds to deviations from α^*. The volatility parameter, ψ^2, is simply a scale factor for the variance of unexpected interest rate changes. If ψ^2 doubles, then the variance doubles. We will argue that allowing ψ^2 to be a time-varying function of the information set results in models that perform better than the existing models. The parameter γ allows the volatility of interest rates to depend on the level of the interest rate. At higher γs, the volatility is more sensitive to interest rate levels.

Following the lead of Marsh and Rosenfeld (1983) (hereafter MR), CKLS and Dietrich-Campbell and Schwartz (1986), we consider the following Euler-Maruyama discrete time approximation to the continuous time model (1),[1]

$$\begin{aligned} r_t - r_{t-1} &= \alpha + \beta r_{t-1} + \epsilon_t, \\ \mathrm{E}\left(\epsilon_t | \Im_{t-1}\right) &= 0, \quad \mathrm{E}\left(\epsilon_t^2 | \Im_{t-1}\right) \equiv \sigma_t^2 = \psi^2 r_{t-1}^{2\gamma}. \end{aligned} \tag{2}$$

In (2), $\Im_{t-1}$ is the information set at time $t-1$, and σ_t^2 is the (conditional) variance of interest rate changes. In this model, heteroskedasticity enters solely through the squared level of the interest rate. We therefore refer to this empirical model as the LEVELS model. A significant contribution of CKLS was the finding that, for one-month Treasury bill yields, the best models for capturing the dynamics of the short rate allow the volatility of unexpected interest rate changes to depend positively on the rate level (i.e., $\gamma > 0$). CKLS estimated γ to be about 1.50 in the LEVELS model, and demonstrated that models with $\gamma \geq 1$ outperformed those with $\gamma < 1$. MR also estimated a flexible functional form that nests several different interest rate models. MR's best models had $\gamma = 1$, but they did not examine models with $\gamma > 1$.

The link between volatility and interest rate levels seems intuitive. As levels increase, we would expect volatility to also increase, if only for scale reasons. Thus, models in which short rate volatility is sensitive to the level of interest rates (i.e., models with $\gamma > 0$) should perform well. However, $\gamma = 1.00$ or $\gamma = 1.50$ implies an unreasonable sensitivity of volatility to levels. For example, in several historical periods (such as 1983–1984), rates were high but stable, and in several periods (such as late 1992 and early 1993), rates were low but volatility was high. The LEVELS model also suffers from other problems. For example, the estimated relationship between volatility and levels is sensitive to whether the October 1987 crash observation is included in the dataset (Bliss and Smith (1994)). It is also sensitive to shifts in the Fed's operating procedures (Ball and Torous (1994)).

[1]While other more sophisticated and perhaps more accurate approximations are available, the Euler-Maruyama approximation is the simplest and most straightforward. Furthermore, other approximations would not allow us to nest the models considered by MR, CKLS, and Dietrich-Campbell and Schwartz (1986) inside our flexible functional form, which is a key aspect of our paper. To minimize the possible deficiencies of this simple approximation, we use relatively high frequency (weekly and monthly) data in our study. Lower frequencies, such as quarterly, are probably less appropriate for this approximation. For a discussion of Euler-Maruyama discrete time approximations, their convergence to the continuous time process, and other types of discrete time approximations, see Kloeden and Platen (1992).

Perhaps the strongest criticism of the LEVELS model is that it restricts volatility to be a function of interest rate levels only, and not of the news arrival process. An alternative model that addresses this is the GARCH model, in which this period's volatility is a function of last period's unexpected news. For example, in a GARCH(1,1) model,

$$\sigma_t^2 \quad = \quad a_0 + a_1 \epsilon_{t-1}^2 + b \sigma_{t-1}^2,$$

where ϵ_{t-1} is the unexpected shock to r. See Engle, Lilien, and Robins (1987) and Evans (1989) for applications of this model to interest rate data, or Bollerslev, Chou, and Kroner (1992) for a comprehensive survey.

At least three problems arise when using GARCH to model short-term interest rate volatility. For example, in direct contrast to much of the theoretical literature, GARCH models do not permit volatility to be a function of interest rate levels. Also, most empirical applications to interest rates find that $\hat{a}_1 + \hat{b} \approx 1$. This implies that current shocks affect volatility forecasts infinitely far into the future, i.e., volatility shocks persist forever. Finally, GARCH models permit negative interest rates.

B. An Alternative Class of Volatility Models

As indicated above, both LEVELS models and GARCH models have important weaknesses. Specifically, in LEVELS models with $\gamma > 0$, increases in interest rates necessarily lead to increased volatility, while decreases in interest rates necessarily lead to decreased volatility. On the other hand, in GARCH models, interest rate levels have no direct impact on volatility. These implications are unrealistic, but they are easy to address by allowing variance to be a function of both the level of the interest rate and unexpected shocks to the interest rate market. We now propose several models that permit this.

Consider, for example, the generalization of (2) obtained by allowing the coefficient ψ^2 to vary through time as new information arrives. An attractive candidate for the information variable is last period's forecast error, which can be viewed as unexpected news. For simplicity, we assume that ψ_t^2 follows an autoregressive-type process. In particular, we suppose

$$\psi_t^2 \quad = \quad a_0 + a_1 \epsilon_{t-1}^2 + b \psi_{t-1}^2, \tag{3}$$

where ϵ_{t-1} is the residual from (2) and a_0, a_1, and b are positive.[2] In this model, large interest rate shocks (as measured by large residuals) cause increases in rate volatility through their effect on ψ_t^2. Using (3) to generalize (2), we propose the following discrete time flexible functional form,

$$\begin{aligned} r_t - r_{t-1} &= \alpha + \beta r_{t-1} + \epsilon_t, \\ \mathrm{E}\left(\epsilon_t | \Im_{t-1}\right) &= 0, \quad \mathrm{E}\left(\epsilon_t^2 | \Im_{t-1}\right) \equiv \sigma_t^2 = \psi_t^2 r_{t-1}^{2\gamma}, \\ \psi_t^2 &= a_0 + a_1 \epsilon_{t-1}^2 + b \psi_{t-1}^2. \end{aligned} \tag{4}$$

[2]Gagnon, Morgan, and Neave (1993) suggests replacing the information variable ϵ_{t-1}^2 with $z_t^2 = \epsilon_{t-1}^2 / \sigma_{t-1}^2$ in equation (3). Experimentation revealed that our results are not sensitive to whether we use ϵ_t^2 or z_t^2 as the information variable.

This model has several interesting features. For example, the sensitivity of volatility to levels is a function of information flows. Specifically, it is higher during high-information periods (when the shocks ϵ_t are larger in absolute magnitude) than during stable periods. Whether this property is good or bad is an empirical issue. Also, the specification in (4) nests both the LEVELS model and the GARCH model. Specifically, if $a_1 = b = 0$, then the time variation in ψ_t disappears and we are back to the LEVELS model. In contrast, if $\gamma = 0$, then the levels effect disappears and we are back in the GARCH framework. This is convenient econometrically because it makes model comparisons straightforward. Finally, as mentioned above, this model can be interpreted as a time-varying parameter version of the LEVELS model. We therefore refer to this model as the TVP-LEVELS model.[3]

A potential weakness of the TVP-LEVELS model is the implication that positive and negative shocks have the same impact on volatility. This weakness is easily addressed by replacing equation (3) in the TVP-LEVELS model with

$$\psi_t^2 = a_0 + a_1\epsilon_{t-1}^2 + a_2\eta_{t-1}^2 + b\psi_{t-1}^2, \tag{5}$$

where $\eta_{t-1} = \min(\epsilon_{t-1}, 0)$. In (5), if $a_2 > 0$, then bad news (negative shocks) has a larger impact on volatility than good news (positive shocks). We refer to this model as the AsymTVP model.

These time-varying parameter models demonstrate one way to allow volatility to depend on both levels and information. An alternative method is to add a levels term directly to the GARCH(1,1) model, giving

$$\sigma_t^2 = a_0 + a_1\epsilon_{t-1}^2 + b\sigma_{t-1}^2 + a_3 r_{t-1}^{2\gamma}. \tag{6}$$

In this model, which we refer to as the GARCH-X model, volatility depends on information flows, while the sensitivity of volatility to levels does not. This contrasts with the TVP models where rate levels have a bigger (smaller) impact when information flows are higher (lower). Like the TVP models, this model also nests both the LEVELS model (if $a_0 = a_1 = b = 0$) and the GARCH model (if $a_3 = 0$).

The final model we examine is an asymmetric extension of the GARCH-X model,

$$\sigma_t^2 = a_0 + a_1\epsilon_{t-1}^2 + a_2\eta_{t-1}^2 + b\sigma_{t-1}^2 + a_3 r_{t-1}^{2\gamma}. \tag{7}$$

This generalization of (6) is analagous to the extension from TVP-LEVELS to AsymTVP, in that we simply add an asymmetry term to the GARCH-X equation. As above, if $a_2 > 0$, then negative shocks have a greater impact on volatility than positive shocks. We refer to this model as the AsymG-X model.

III. Misspecification Tests

The models we estimate are evaluated with the likelihood ratio statistic and the Ljung-Box statistic. In addition, we employ the robust conditional moment

[3]This generalization focuses entirely on one-factor models. Our model is a one-factor model because only one source of uncertainty appears in the mean equation, and this same source of uncertainty drives the GARCH behavior of the parameter ψ_t^2. We could examine two-factor models as in Longstaff and Schwartz (1992) by introducing another source of uncertainty into both the GARCH variance process and the mean equation, but our aim is to examine and evaluate different one-factor models.

tests of Wooldridge (1990). The conditional moment tests allow us to identify possible sources of misspecification in the model, and are robust to our distributional assumptions. Since these tests may be unfamiliar, we provide the following brief description.

Conditional moment (CM) tests analyze a model by forming functions of the data and model parameters called generalized residuals (v_t), which are constructed to have zero conditional expectation if the model is correctly specified. The test investigates whether the generalized residuals have zero conditional expectation by testing whether they are uncorrelated with selected elements of the information set. These elements of the information set, or, as they are referred to in this context, misspecification indicators ($\lambda_{i,t-1}$), are chosen by considering dimensions along which the model is likely to fail. The essence of the CM test is whether the conditional cross-moments, formed as the product of the misspecification indicators and the model's generalized residuals ($m_{it} = v_t \lambda_{i,t-1}$), are far enough from zero to warrant rejection of the model. CM tests are especially useful because they can be tailored to evaluate specific forms of misspecification simply by selecting appropriate misspecification indicators and generalized residuals.

The CM test is constructed from a series of two auxiliary regressions after the null model has been estimated. First, the misspecification indicator, $\lambda_{i,t-1}$, is regressed on the expected gradient of the generalized residual, $\mathrm{E}_{t-1}(\nabla v_t)$. If there are k parameters in the null model, the gradient ∇v_t will be a $k \times 1$ vector, which can be calculated by either numerically or analytically differentiating the generalized residual with respect to the null model's k parameters. The expectation is taken conditional on information available at time $t-1$. For most applications (including ours), this conditional expectation is computed analytically. The residuals from this first stage regression are

$$\hat{z}_t = \lambda_{i,t-1} - \hat{\delta}' \mathrm{E}_{t-1}(\nabla v_t),$$

where δ is a $k \times 1$ vector of regression coefficients. In the second stage regression, a vector of ones is regressed on the product of the first stage residuals $\hat{z}_t$ and the generalized residual v_t. Denoting the sample size as T and the coefficient of determination from the second stage regression as R^2, Wooldridge (1990) shows that TR^2 is asymptotically distributed χ_1^2 under the null hypothesis that the indicator is not correlated with the generalized residual. In the case of multiple indicators, the first step is to run a multivariate regression of the indicators on the expected value of the gradient, and the second step is unchanged. TR^2 from this second stage regression is distributed χ_p^2, where p is the number of indicator variables being tested.

A simple example makes the nature of the test clear. Consider testing the standard OLS simple regression model, $y_t = \hat{\beta} X_t + \hat{\epsilon}_t$, for misspecification. Here, we can use $\epsilon_t = y_t - \beta X_t$ as the generalized residual since its conditional expectation is zero. Denoting $\Im_{t-1}$ as the time $t-1$ information set, the CM test provides evidence on whether $\mathrm{E}(\epsilon_t \lambda_{t-1}) = 0$, where λ_{t-1} is any variable (or function of variables and parameters) contained in $X_t \cup \Im_{t-1}$. In this example, one choice of misspecification indicator might be $\lambda_{t-1} = \epsilon_{t-1}$, in which case, the CM test provides evidence on whether $\mathrm{E}(\epsilon_t \epsilon_{t-1}) = 0$, i.e., evidence on first order serial correlation. In this example, ϵ_{t-1} is the indicator and $m_t = \epsilon_t \epsilon_{t-1}$ is the conditional moment being tested. The gradient of the generalized residual is $\nabla\epsilon_t = \nabla(y_t - \beta X_t) = -X_t$, so the expected gradient is $\mathrm{E}_{t-1}(\nabla\epsilon_t) = -X_t$. As discussed above, the CM test is constructed from two auxiliary regressions. The first regresses the misspecification indicators (in this illustration, the lagged OLS residuals $\hat{\epsilon}_{t-1}$) on the expected gradient of the generalized residuals (in this case, the negative of the independent variable, $-X_t$). The residuals from this regression are $\hat{z}_t$. The second regresses a vector of ones on $\hat{z}_t\hat{\epsilon}_t$ (product of the first stage residuals and the generalized residual). The TR^2 from the second stage regression is the test statistic. This example makes clear the intuition behind the CM test: the first step purges from the misspecification indicator all the information already in the model. The second step tests whether the unexplained part of the indicator is correlated with the generalized residual.

Turning now to interest rate volatility models, if the variance equation is correctly specified, we would expect

$$\mathrm{E}\left(\epsilon_t^2 - \sigma_t^2 \mid \Im_{t-1}\right) = 0.$$

Hence, $v_t = \epsilon_t^2 - \sigma_t^2$ is a reasonable candidate for a generalized residual to test whether the variance equation is correctly specified. Any element of $\Im_{t-1}$ is a valid candidate for a misspecification indicator. The possibilities for indicators are numerous. Selecting indicators carefully should provide important information about the deficiencies of the interest rate models we examine.

We propose several indicators to test for likely sources of misspecification in the variance equation. The first indicator is the lagged interest rate level,

$$\lambda_{1,t-1} = r_{t-1}.$$

The moment for this test is $m_{1t} = v_t\lambda_{1,t-1}$, and if the variance equation is correctly specified, $\mathrm{E}(m_{1t}) = 0$. On the other hand, if the variance model under predicts (over predicts) variance when interest rates are high, this moment will have an expected value greater than (less than) zero. Hence, this moment should identify variance models that misspecify the sensitivity of volatility to interest rate levels.

The second indicator we consider is

$$\lambda_{2,t-1} = I\left(\epsilon_{t-1} < 0\right)\epsilon_{t-1},$$

where $I(\cdot)$ equals one when the condition in parentheses is true and equals zero otherwise. If our variance equation is correctly specified, then the associated moment, $m_{2t} = v_t\lambda_{2,t-1}$, should have expected value zero. But if our model under predicts (over predicts) variance after negative shocks, m_{2t} should have a positive (negative) expected value. Hence, this moment tests for the appropriateness of asymmetric variance models, such as equation (5) above.

Our third set of misspecification indicators focuses on serial correlation in standardized squared residuals,

$$\begin{aligned}
\lambda_{3,t-1} &= v_{t-1},\\
\lambda_{4,t-1} &= v_{t-2},\\
\lambda_{5,t-1} &= v_{t-3},\\
\lambda_{6,t-1} &= v_{t-4},
\end{aligned}$$

where v_{t-k} are lagged values of the generalized residuals. The associated moments have expected value zero if our model adequately represents the dynamics in the variances. Essentially, these test for remaining GARCH effects.[4]

The final set of misspecification indicators we examine tests for a structural break in the variance process caused by the change in the Fed's operating procedures between October 1979 and October 1982. The misspecification indicator we propose is

$$\lambda_{7,t-1} \quad = \quad I(\text{Oct } 1979 < t < \text{Oct } 1982).$$

Again, if our model under predicts (over predicts) variances during the Fed's experiment, then the associated moment should have a positive (negative) expected value.

To summarize, we offer seven misspecification indicators, broken down into four sets. The first indicator tests whether the model misrepresents the dependence of variance on levels; we call this test a "rate level" test. The second tests for asymmetric variance models; we call this an "asymmetry" test. The third set tests for misspecified dynamics in the estimated variance process; we call these "GARCH" tests. The final set tests for a structural break during the Fed's monetary targeting experiment; we call these "structural break" tests.

IV. The Data

We analyze two data sets in this study. The first consists of 909 weekly observations on 13-week Treasury bill yields, from February 9, 1973, to July 6, 1990. These data were obtained from Data Resources, Inc.[5] The second data set consists of 407 monthly observations of the total return on 30-day Treasury bills, from January 1960 to December 1993. These data come from Ibbotson Associates' Stocks, Bonds, Bills and Inflation Yearbook. For expositional purposes, we refer to the weekly 13-week yield data as TB13WK, and the monthly 30-day rates as TB30DY. Portions of this study were replicated on other data sets (monthly three-month Treasury bill yields and weekly seven-day Eurodollar rates) and the results are qualitatively similar. Following the literature, we focus on nominal rates to avoid the serious data problems that would be created by attempting to define real rates. Also, throughout the ensuing analysis on TB13WK, we use a crash dummy variable that takes the value one during the week of the 1987 market crash and zero otherwise. We include the dummy because interest rates fell by more than 11 standard deviations over the week of the crash, and none of the models we evaluate is designed to capture the effects of large "one-time" exogenous shocks.[6] The dummy variable is not included in the analysis of TB30DY because these rates fell by only three standard deviations over the month of the crash.

[4]In the empirical section of this paper, we report only the joint test for whether m_{3t}, m_{4t}, m_{5t}, and m_{6t} all have expected value zero, against the alternative that at least one of them has a nonzero expected value.

[5]We wish to thank Mark Flannery for making these data available to us.

[6]Much of our ensuing analysis was also conducted without the crash dummy, and the only material difference using our proposed class of models was in the estimated degrees of freedom in our conditional t-distribution. In contrast, Bliss and Smith (1994) finds that including the crash observation in the LEVELS model causes one to overestimate the sensitivity of volatility to levels.

Our weekly data have two advantages over our monthly data. First, the discrete time approximation to the continuous time model given in (1) should hold better with higher frequency data. Second, the increased sample size makes the asymptotic χ^2 statistics of the CM tests more appropriate. However, the monthly data have the advantage that 30-day bills are closer to the true short rate that these models are designed to analyze. Our conclusions from both data sets are essentially identical, suggesting either that changing horizon and/or frequency probably does not affect our conclusions, or that the two changes somehow offset each other.

The 13-week Treasury bill yield over our sample averaged 8.1 percent (annualized) with a standard deviation of 2.59 percent, while the changes averaged 0.0022 percent with a standard deviation of 0.351 percent. The highest 13-week yield was 16.68 percent on May 22, 1981, and the lowest was 4.25 percent on December 17, 1976. The largest one-week increase in the yield was 1.81 percent during the week ending May 8, 1981, and the largest one-week decline was 2.17 percent during the week ending August 20, 1982. The monthly 30-day rates in TB30DY averaged 6.01 percent with a standard deviation of 2.80 percent, while the changes averaged -0.003 percent with a standard deviation of 0.91 percent. The highest monthly return (annualized) was 16.17 percent during June 1981 and the lowest was 1.58 percent during November 1961. The largest one-month increase was 4.20 percent over December 1980 and the largest one-month decline was 5.35 percent over the month of May 1980. The dotted lines in Figures 1 and 2 graph the changes in interest rates for TB13WK and TB30DY, respectively.

FIGURE 1
Weekly 13-Week Treasury Bill Volatility
February 1973–July 1990

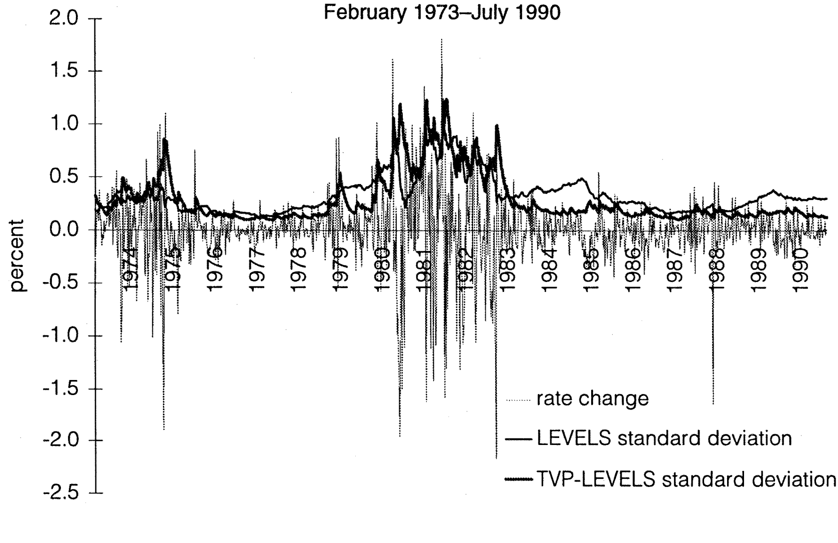
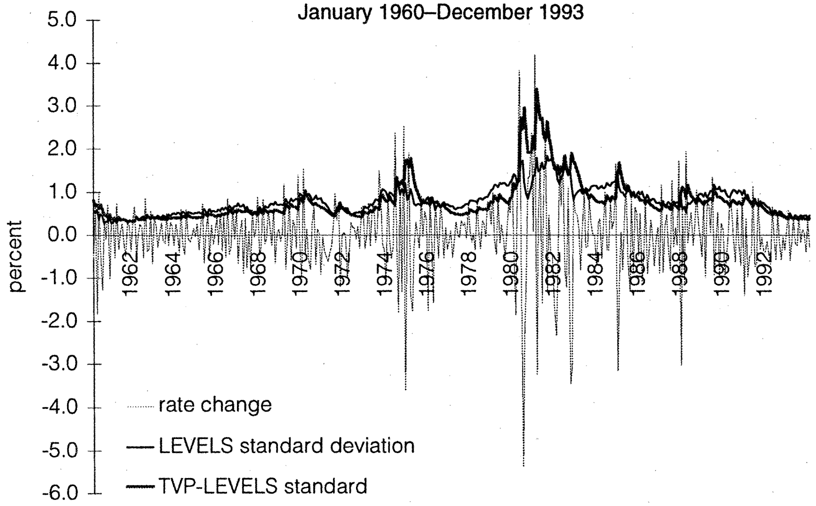


FIGURE 2
Monthly 30-Day Treasury Bill Volatility
January 1960–December 1993



V. Empirical Results

A. Model Estimates

We estimate our models with maximum likelihood, assuming a conditional student's t-distribution where the degrees of freedom, ν, are estimated from the data. The student's t-distribution gives our model enough flexibility to capture the leptokurtosis in interest rate data. Also, robust conditional moment tests require only $\sqrt{T}$-consistent estimates of the model parameters. Maximizing a t-distributed likelihood function yields $\sqrt{T}$-consistent estimates under very weak conditions if the mean and variance equations are correctly specified and the true distribution is symmetric (Newey and Steigerwald (1994)). We use maximum likelihood instead of a distribution-free estimator (such as GMM) for two reasons. First, our simulations later in the paper require a precise specification of the conditional distribution. And second, Broze, Scaillet, and Zakoian (1993) prove in their proposition 3.3 that the GMM estimator of the LEVELS model is not well behaved if $\gamma > 1$. CKLS obtains estimates of γ that exceed one, so we prefer to avoid GMM estimation.

Before estimating our flexible functional forms, we estimate and present two baseline models against which comparisons can be made: the LEVELS model and the GARCH model. These models are useful reference points because both have many applications in the literature. For example, the LEVELS-type volatility processes and various special cases of it are used by Vasicek (1977), Dothan (1978), Cox, Ingersoll, and Ross (1980), Oldfield and Rogalski (1987), Gibbons and Ramaswamy (1993), and Sanders and Unal (1988), among many others. GARCH and GARCH-type models are used by Cai (1994), Cecchetti, Cumby, and Figlewski (1988), Engle, Ng, and Rothschild (1992), Flannery, Hameed, and Harjes (1992), and Hamilton and Susmel (1992), among many others.

The LEVELS and GARCH model estimates are presented in the first two columns of Tables 1 and 2. Table 1 presents results from TB13WK while Table 2 presents results from TB30DY. For both data sets, the mean equation parameters (α and β) are similar between the LEVELS and GARCH models. There is no evidence of mean reversion in the TB13WK data, and only weak evidence in the TB30DY data.[7] This result is common to studies of short-term interest rates. The mean equation similarities between the LEVELS and GARCH models suggest that any differences between these models are caused by how they treat volatility.

The Ljung-Box $Q(\varepsilon_t/\sigma_t)$ statistics indicate that both models (and all other models evaluated in this paper) have significant serial correlation in the residuals. There are probably several sources of this misspecification, one of which is that our drift term, $\alpha + \beta r_{t-1}$, is not the correct arbitrage-free drift (see Heath, Jarrow, and Morton (1992)). Therefore, our mean equation is necessarily misspecified. We do not use the arbitrage-free drift here because, for the models and frequencies we consider, the arbitrage-free drift is small in magnitude relative to the volatility. This implies that getting the drift right, while very difficult to do, would have little impact on our volatility estimates. Also, had we used the arbitrage-free drift, our results would not be directly comparable to CKLS, MR, and the other studies cited above that use this approximation.

As in the MR and CKLS papers, estimates of the LEVELS model show relatively high values for γ of 1.56 in TB13WK and 0.83 in TB30DY.[8] So, for TB13WK, the variance of unexpected interest rate changes is proportional to the cube of the level of interest rates. High sensitivity is also found in TB30DY. Therefore, the LEVELS model implies that as interest rates increase, volatility increases dramatically. The GARCH model, on the other hand, does not permit volatility to depend on interest rate levels, but instead allows volatility to change as news hits the market. So, in this model, there need be no apparent relationship between the estimated volatility and the level of the interest rate. In fact, the correlation between interest rate levels and GARCH volatility is only 0.608 (0.550) in TB13WK (TB30DY), while the correlation between interest rate levels and LEVELS volatility is 0.985 (0.948). Notice also that the GARCH model exhibits strong persistence in the volatility parameter for both datasets, as $a_1 + b$ is approximately one. Thus, shocks to the variance persist indefinitely.

Further useful insights can be gained from Figures 1 and 2, which graph the changes in T-bill yields along with the estimated conditional standard deviations from the LEVELS model. Figure 1 applies to TB13WK and Figure 2 to TB30DY. The striking observation from these graphs is that in several periods, especially 1983–1984, the LEVELS model misrepresents realized volatility. See Figure 3 for a close-up of this period for TB13WK. Interest rates were relatively high during these two years, but the Fed's interest rate targeting policy kept volatility low. In contrast, the GARCH model tracks realized volatility much better during this

[7]Mean reversion exists if $\beta < 0$, so a test for mean reversion is a test of whether $\beta = 0$ against the alternative that $\beta < 0$. However, under the null hypothesis of no mean reversion, r_t has a stochastic trend, implying that the usual t-test is inappropriate.

[8]This estimate of γ is sensitive to the time period under consideration. For example, the $\hat{\gamma}$ for TB30DY increases to 1.14 when we reduce the time period to June 1964–December 1989, the time period considered by CKLS.

TABLE 1

Statistical Models of the Short-Term Interest Rate: Weekly 13-Week Treasury Bill Yields, 02/09/73–07/06/90

	LEVELS	GARCH	TVP-LVLS	AsymTVP	GARCH-X	AsymG-X
Panel 1.						
α	0.0169	0.0204	0.0184	0.0104	0.0130	−0.0016
	(0.024)	(0.025)	(0.024)	(0.026)	(0.023)	(0.025)
β	−0.0010	−0.0016	−0.0011	0.0030	−0.0004	0.0061
	(0.004)	(0.004)	(0.003)	(0.004)	(0.003)	(0.004)
δ	−1.6499	−1.6490	−1.6506	−1.6711	−1.6502	−1.6804
	(0.006)	(0.006)	(0.006)	(0.008)	(0.006)	(0.009)
a_0	0.0002	0.0011	0.0004	0.0010	0.0016	0.0064
	(0.000)	(0.001)	(0.000)	(0.001)	(0.001)	(0.003)
a_1	·	0.1871	0.0236	0.0327	0.1702	0.1195
		(0.060)	(0.014)	(0.016)	(0.055)	(0.054)
a_2	·	·	·	0.0989	·	0.8782
				(0.046)		(0.217)
a_3	·	·	·	·	0.0045	0.0206
					(0.002)	(0.012)
b	·	0.8140	0.7969	0.4303	0.7791	0.2326
		(0.055)	(0.066)	(0.134)	(0.066)	(0.164)
γ	1.5594	·	0.4698	0.4362	3.7398	2.5516
	(0.119)		(0.118)	(0.110)	(0.660)	(0.480)
ν	3.0479	6.0188	6.3185	5.8171	6.1634	5.6928
	(0.354)	(1.147)	(1.250)	(1.213)	(1.177)	(1.614)
Panel 2.						
L	−14.92	61.15	69.27	128.49	69.10	132.41
SBC	−35.35	37.31	42.02	97.83	38.45	98.35
κ	8.48	4.30	4.44	5.41	4.52	6.22
ν_κ	4.71	5.40	5.35	5.11	5.33	4.97
$Q(\epsilon_t/\sigma_t)$	21.33	26.88	24.89	36.04	24.41	36.46
	(0.046)	(0.008)	(0.015)	(0.000)	(0.018)	(0.000)
$Q(\epsilon_t^2/\sigma_t^2)$	216.42	8.19	6.84	22.94	8.51	47.66
	(0.000)	(0.770)	(0.868)	(0.028)	(0.744)	(0.000)
Rate Level (λ_1)	0.000	9.026	0.338	1.452	2.483	0.233
	(0.994)	(0.003)	(0.561)	(0.228)	(0.115)	(0.630)
Asymmetry (λ_2)	9.718	18.038	16.006	3.331	13.178	7.457
	(0.002)	(0.000)	(0.000)	(0.068)	(0.000)	(0.006)
GARCH ($\lambda_3 \rightarrow \lambda_6$)	15.353	3.205	4.347	5.192	4.787	2.780
	(0.004)	(0.524)	(0.361)	(0.268)	(0.310)	(0.595)
Struct. Break (λ_7)	11.113	2.627	0.000	0.287	0.080	1.542
	(0.001)	(0.105)	(0.996)	(0.592)	(0.778)	(0.214)
R^2	0.209	0.286	0.295	0.523	0.303	0.616

Columns 1, 2, 3, and 4 of Panel 1 report the maximum likelihood estimates from the model,

$$r_t - r_{t-1} = \alpha + \beta r_{t-1} + \delta D_{\text{crash}} + \epsilon_t, \tag{1}$$

$$\epsilon_t/\sigma_t \mid \Im_{t-1} \sim t_\nu, \tag{2}$$

$$E_{t-1}\epsilon_t^2 \equiv \sigma_t^2 = \psi_t^2 r_{t-1}^{2\gamma}; \quad \psi_t^2 = a_0 + a_1\epsilon_{t-1}^2 + a_2\eta_{t-1}^2 + b\psi_{t-1}^2, \tag{3}$$

where $\eta_t = \min(\epsilon_t, 0)$. Standard errors are in parentheses. Columns 5 and 6 replace equations (3) above with

$$E_{t-1}\epsilon_t^2 \equiv \sigma_t^2 = a_0 + a_1\epsilon_{t-1}^2 + +a_2\eta_{t-1}^2 + a_3(r_{t-1}/10)^{2\gamma} + b\sigma_{t-1}^2. \tag{4}$$

Panel 2 reports: log-likelihood values (L); Schwarz-Bayes information criterion (SBC); excess kurtosis of standardized residuals (κ); degrees of freedom implied by the excess kurtosis (ν_κ); Ljung-Box tests for up to twelfth order serial correlation in the standardized residuals and squared standardized residuals ($Q(\epsilon_t/\sigma_t)$ and $Q(\epsilon_t^2/\sigma_t^2)$, respectively); and a set of Wooldridge's (1990) robust conditional moment tests as discussed in the text. *P*-values are in parentheses.

TABLE 2

Statistical Models of the Short-Term Interest Rate: Monthly Total Returns on 30-Day Treasury Bills, 01/60–12/93

	LEVELS	GARCH	TVP-LVLS	AsymTVP	GARCH-X	AsymG-X
Panel 1.						
α	0.2449	0.1688	0.1853	0.1638	0.1600	0.2666
	(0.082)	(0.071)	(0.071)	(0.073)	(0.068)	(0.085)
β	−0.0415	−0.0316	−0.0299	−0.0070	−0.0290	−0.0165
	(0.017)	(0.016)	(0.015)	(0.016)	(0.015)	(0.016)
a_0	0.0334	0.0303	0.0106	0.0127	0.0224	0.0076
	(0.011)	(0.014)	(0.004)	(0.006)	(0.015)	(0.034)
a_1	·	0.2238	0.0193	0.0000	0.1299	0.0000
		(0.062)	(0.012)	(·)	(0.055)	(·)
a_2	·	·	·	0.0676	·	0.6464
				(0.032)		(0.094)
a_3	·	·	·	·	0.3167	0.6243
					(0.132)	(0.238)
b	·	0.7412	0.7217	0.5824	0.6762	0.2405
		(0.053)	(0.077)	(0.075)	(0.082)	(0.138)
γ	0.8291	·	0.5432	0.5040	1.4371	1.3008
	(0.092)		(0.094)	(0.142)	(0.369)	(0.386)
ν	7.9631	9.4818	43.5508	43.5322	22.8144	19.0472
	(2.417)	(3.532)	(73.650)	(75.410)	(18.148)	(14.151)
Panel 2.						
L	−449.67	−443.27	−429.81	−401.75	−431.47	−395.63
SBC	−464.69	−461.30	−450.84	−425.78	−455.51	−422.67
κ	3.98	3.79	3.14	3.60	3.28	3.96
ν_κ	5.51	5.58	5.91	5.67	5.83	5.51
$Q(\epsilon_t/\sigma_t)$	86.81	54.00	121.48	127.20	115.97	103.28
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
$Q(\epsilon_t^2/\sigma_t^2)$	97.33	5.36	12.61	27.72	10.72	23.95
	(0.000)	(0.719)	(0.398)	(0.006)	(0.553)	(0.021)
Rate Level (λ_1)	5.367	10.400	1.505	2.197	3.895	1.412
	(0.021)	(0.001)	(0.220)	(0.138)	(0.048)	(0.235)
Asymmetry (λ_2)	11.096	8.409	13.267	0.673	11.075	1.003
	(0.001)	(0.004)	(0.000)	(0.412)	(0.001)	(0.317)
GARCH ($\lambda_3 \rightarrow \lambda_6$)	12.781	3.411	6.684	8.869	6.699	3.104
	(0.012)	(0.492)	(0.154)	(0.064)	(0.153)	(0.541)
Struct. Break (λ_7)	8.990	3.248	1.162	1.930	3.397	4.327
	(0.003)	(0.072)	(0.281)	(0.165)	(0.065)	(0.038)
R^2	0.187	0.220	0.265	0.536	0.259	0.558

Columns 1, 2, 3, and 4 of Panel 1 report the maximum likelihood estimates from the model,

(1) $$r_t - r_{t-1} = \alpha + \beta r_{t-1} + \epsilon_t,$$

(2) $$\epsilon_t/\sigma_t \mid \Im_{t-1} \sim t_\nu,$$

(3) $$E_{t-1}\epsilon_t^2 \equiv \sigma_t^2 = \psi_t^2 r_{t-1}^{2\gamma}; \quad \psi_t^2 = a_0 + a_1\epsilon_{t-1}^2 + a_2\eta_{t-1}^2 + b\psi_{t-1}^2,$$

where $\eta_t = \min(\epsilon_t, 0)$. Standard errors are in parentheses. Columns 5 and 6 replace equations (3) above with

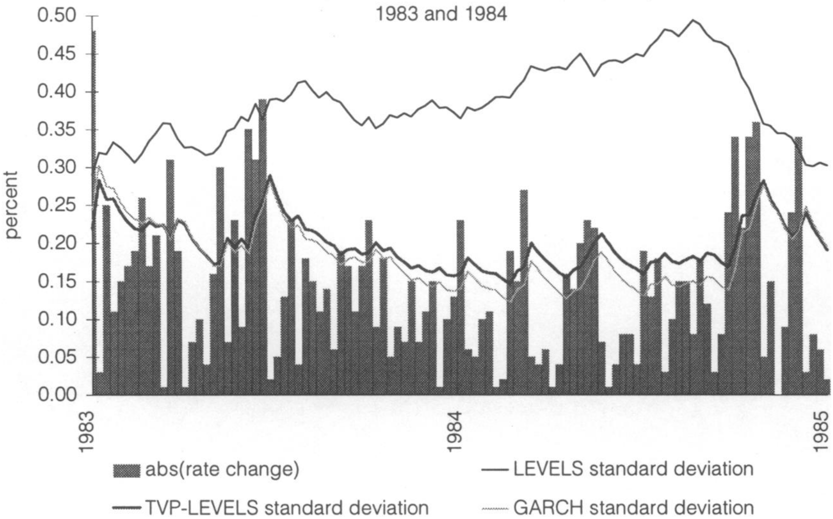
(4) $$E_{t-1}\epsilon_t^2 \equiv \sigma_t^2 = a_0 + a_1\epsilon_{t-1}^2 + +a_2\eta_{t-1}^2 + a_3(r_{t-1}/10)^{2\gamma} + b\sigma_{t-1}^2.$$

Panel 2 reports: log-likelihood values (L); Schwarz-Bayes information criterion (SBC); excess kurtosis of standardized residuals (κ); degrees of freedom implied by the excess kurtosis (ν_κ); Ljung-Box tests for up to twelfth order serial correlation in the standardized residuals and squared standardized residuals ($Q(\epsilon_t/\sigma_t)$ and $Q(\epsilon_t^2/\sigma_t^2)$, respectively); and a set of Wooldridge's (1990) robust conditional moment tests as discussed in the text. *P*-values are in parentheses.

period. However, while we do not show the GARCH volatility in Figures 1 and 2 (to keep the graphs readable), the GARCH model under predicts volatility during the 1979–1982 monetary targeting experiment by the Fed, when interest rates were relatively high.

FIGURE 3

Weekly 13-Week Treasury Bill Volatility
1983 and 1984



Many of these observations and criticisms of the LEVELS and GARCH models are verified in the test statistics reported in the bottom panels of the tables. First, the likelihood functions for the GARCH models are higher than for the LEVELS models (61.15 vs. −14.92 for TB13WK, and −443.27 vs. −449.67 for TB30DY). While this is not a formal test because the models are not nested, it does suggest that the GARCH model tracks volatility better than the LEVELS model. A similar conclusion is obtained from the Schwarz-Bayes information criterion (SBC). Also, both the Ljung-Box and the CM tests[9] for remaining GARCH effects find that the LEVELS model fails to model adequately the serial correlation in the volatility process. This result is not necessarily expected, because the LEVELS variances are strongly serially correlated. In fact, the first order serial correlation coefficient for the LEVELS variances is 0.981 in TB13WK and 0.932 in TB30DY. In addition, the rate levels test (λ_1) indicates that the LEVELS model captures the dependence of volatility on levels, though the p-value in TB30DY is only 0.021. In contrast, the tests on the GARCH model reveal that it captures serial correlation in volatility (the Ljung-Box and CM GARCH tests are both insignificant), but fails to pick up the dependence of volatility on rate levels (the CM rate level test statistics are strongly significant).

[9]When implementing the CM tests, we first divided the generalized residual and the gradient by the conditional variance. This transformation does not affect the asymptotic distribution of the tests, but has the advantage of making our tests asymptotically equivalent to the more well-known Lagrange Multiplier tests in classical circumstances.

Our structural break test (λ_7) reveals that the LEVELS model is misspecified during the Fed's monetary targeting experiment between October 1979 and October 1982. This result differs from CKLS, who found no structural break in their model. Our results differ because our test is designed to detect a different kind of misspecification than the CKLS test. We examine whether the estimated volatility was correct, on average, during the Fed's experiment, while their test examines whether the parameters of the model changed. Our conclusions are supported by Ball and Torous (1994), who estimate a stochastic switching model and find evidence of structural breaks in the LEVELS volatility process. The structural break test also shows that the GARCH model is misspecified during this period in the TB30DY data, but not in the TB13WK data. Also, as evidenced by the asymmetry tests (λ_2), both these models misspecify the way that volatility responds differently to positive shocks than to negative shocks. This suggests that the asymmetric models that we introduced above might dominate the symmetric ones.

Finally, the GARCH model with a conditional t-distribution permits excess kurtosis in the data. For example, in TB13WK, the standardized residuals from the GARCH model exhibit excess kurtosis of 4.30. The relationship between excess kurtosis and degrees of freedom in a t-distribution is

$$\nu_\kappa = \frac{6}{\kappa} + 4,$$

where κ is the excess kurtosis and ν_κ is the degrees of freedom. So if $\kappa = 4.30$, then $\nu_\kappa = 5.40$. This is only about one half of a standard deviation away from our estimated degrees of freedom of 6.01. A similar conclusion can be drawn from TB30DY. We take this as informal evidence that the GARCH model with conditional t errors adequately captures the leptokurtosis in short-term interest rate data.

Our results suggest that models explicitly incorporating both serial correlation in volatility and dependence of volatility on levels should be superior to either of the baseline models. Results for the TVP-LEVELS and AsymTVP models are presented in the middle two columns of Tables 1 and 2. Consider first the TVP-LEVELS model. The estimated volatilities are plotted in Figures 1, 2, and 3. These estimated volatilities seem to track realized volatility better than the levels model, especially during the 1983–1984 period.

In the TVP-LEVELS model, we find strong evidence of a variance process that differs from both the LEVELS process and the GARCH process. a_1 and b are jointly significantly different from zero, implying that the volatility parameter is time varying. Similarly, γ is significantly different from zero, implying that the variance is an increasing function of levels. In fact, the correlation between the predicted variance and the interest rate level is 0.729 in TB13WK and 0.758 in TB30DY. Also, this model seems to correct many of the misspecifications in both the LEVELS and GARCH variance processes. The TVP-LEVELS model passes most of our volatility-related specification tests. It captures both the serial correlation in volatility and the dependence on levels. In addition, the combined use of rate levels and shocks gives a functional form that is flexible enough to capture the Fed's structural shift. The estimated degrees of freedom are not statistically significantly different from those implied by our excess kurtoses measures, suggesting that we have modeled adequately the leptokurtosis in the data sets. Finally, SBC comparisons indicate that the TVP-LEVELS model is statistically preferable to both the LEVELS and the GARCH models. However, the asymmetry test (λ_2) is strongly significant, suggesting that the TVP-LEVELS model does not correctly capture the way that volatility responds differently to positive shocks than to negative shocks.

Notice that a_1 has dropped substantially compared to the GARCH model, causing $a_1 + b$ to drop from about 1 in the GARCH model to about 0.83 in ours. We conjecture that, in contrast to the GARCH model, persistence in the volatility parameter is finite.[10] This suggests that the common finding of integrated GARCH in the literature could be due, at least in part, to misspecification of the volatility equation by ignoring the dependence on levels. Also, our nonzero estimate for γ means that, in contrast to the GARCH model, interest rates in the TVP-LEVELS model can never be less than zero (if the data frequency is sufficiently high). Finally, and most importantly, the estimated γ is 0.459 in TB13WK and 0.543 in TB30DY, which are not significantly different from 0.50. This suggests that the "square root" models, where $\gamma = 0.50$, characterize short-term interest rate data better than other one-factor models.[11] Therefore, square root-type volatility processes with time-varying volatility parameters are probably better than high γ processes when pricing interest rate derivatives that depend on the volatility process.

As noted above, the TVP-LEVELS model fails the asymmetry test. We therefore estimated the AsymTVP model, and give the results in the fourth column of Tables 1 and 2. While the AsymTVP model statistically dominates the TVP-LEVELS model (the likelihood ratio tests are 118.44 for TB13WK and 56.12 for TB30DY), most of the results discussed above for the TVP-LEVELS are unaffected. The only notable differences are that the model now passes the asymmetry test for both datasets, but fails the $Q(\epsilon^2/\sigma^2)$ test. Perhaps more important, the estimates of γ are similar to the TVP-LEVELS estimates. In the TB13WK dataset, γ drops slightly to 0.436 and, in the TB30DY dataset, it drops slightly to 0.504. Neither of these values are statistically different from 0.50, again pointing to the potential practical usefulness of the square root-type volatility models.

Consider the GARCH-X and AsymG-X models.[12] Estimates for these models are given in the final two columns of Tables 1 and 2. Recall that, in contrast to the TVP models, the GARCH-X models do not permit sensitivity of volatility to levels to depend on information flows. Thus, economically, the GARCH-X models are fundamentally different from the TVP models. However, statistically, for the data sets considered here, these models prove to be very similar. For the symmetric versions, the TVP models have higher likelihood functions and Schwarz-Bayes Criteria while, for the asymmetric versions, the reverse is true. Again, these models

[10]We were unable to prove this statement, because volatility persistence in the TVP-LEVELS model can no longer be measured by $a_1 + b$. Volatility is now a function of both the volatility parameter, ψ_t^2, and interest rate levels. Therefore, volatility persistence is a function of both persistence in the volatility parameter (as measured by $a_1 + b$) and persistence in interest rate levels.

[11]For lack of a better term, we use the phrase "square root model" to refer to the $\gamma = 0.50$ models. The more common usage of the phrase "square root model" would require $a_1 = b = 0$.

[12]We divided the r_{t-1} term in equations (6) and (7) by 10 when estimating the model. Failing to do this gave $\hat{a}_3 \approx 0.00000000015$, which caused estimation problems.

are not nested, so these likelihood-based comparisons should be interpreted with caution. Also, as for the TVP models, the symmetric GARCH-X model fails the asymmetry test (λ_2), while the AsymG-X model fails the $Q(\epsilon^2/\sigma^2)$ test. However, the symmetric GARCH-X model still fails the rate level test (λ_1) for TB13WK, and the AsymG-X model still fails the asymmetry test (λ_2) for TB13WK and the structural break test (λ_7) for TB30DY. This suggests that there is still room for modeling improvements in the GARCH-X group of models.

One important difference between the GARCH-X models and the TVP models is the estimated sensitivity of volatility to levels. In the TB30DY dataset, $\hat{\gamma} \approx 1.4$ for the symmetric GARCH-X model and 1.3 for the asymmetric GARCH-X model, while in the TB13WK dataset, $\hat{\gamma} \approx 3.7$ for the symmetric GARCH-X model and 2.6 for the AsymG-X model. At first glance, this seems like an unreasonably high sensitivity. But recall that the interest rate in the variance equation was divided by 10 in these models. Taking this into account, the levels term in the GARCH-X model for TB13WK, for example, can be written as

$$0.0045\left(\frac{r_{t-1}}{10}\right)^{7.4795} = \frac{0.0045}{10^{7.4795}} r_{t-1}^{7.4795} = 10^{-10}\left(1.492 r_{t-1}^{7.4795}\right).$$

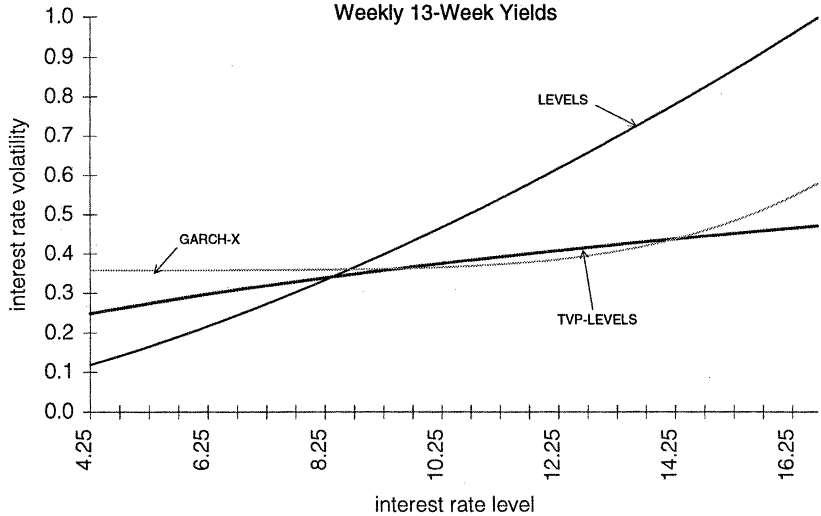
So, the low constant term dampens much of the apparent sensitivity. To examine the effect of this low constant term, and to directly compare the sensitivities implied by the LEVELS, TVP-LEVELS, and GARCH-X models in the TB13WK dataset, consider Figure 4. Figure 4 plots volatility as a function of interest rate levels for these three models.[13] Sensitivity of volatility to levels is measured by the slope of these lines. The GARCH-X model is insensitive to rate levels for low interest rates, but becomes very sensitive as interest rates increase. In contrast, the LEVELS model is very sensitive for all interest rate levels, and the TVP-LEVELS model is moderately sensitive for all interest rate levels. Interestingly, the TVP-LEVELS and GARCH-X sensitivities are similar for the ranges of interest rates experienced over the last 20 years (7 percent to 14 percent). It is only at the extremes that these two models have different implications about the sensitivity of volatility to levels.

Longstaff and Schwartz (1992) proposes GARCH-X models in which γ is restricted to be one. Although we do not present the results of this restriction in the tables, imposing $\gamma = 1$ gives models that are statistically only marginally better than the GARCH models and clearly inferior to the models considered here. Also, these models fail the rate level test (λ_1), indicating that $\gamma = 1$ in the GARCH-X model does not permit adequate sensitivity of volatility to levels.

Another model evaluation criterion we use to compare the three classes of interest rate models is forecasting power. To evaluate forecasting performance, we use the model parameters in the top panel of Tables 1 and 2 to construct a one-

[13]For each of the lines in Figure 4, we fixed all the terms in the respective volatility equations (except the interest rate level) at their sample means, and allowed the interest rate to vary from the sample minimum of 4.25 percent to the sample maximum of 16.68 percent. So, for example, the equation of the GARCH-X line in Figure 4 is $\sigma_t = 0.128 + 0.0045 * (r_{t-1}/10)^{7.479}$. The equation for the TVP-LEVELS line is $\sigma_t = 0.0158 * r_{t-1}^{0.9396}$.

FIGURE 4
Sensitivity of Volatility to Interest Rates
Weekly 13-Week Yields



step-ahead forecast of volatility, $\hat{\sigma}_t$, for each observation in our sample.[14] Using these, we calculate mean squared forecast errors,

$$\text{MSFE} \quad = \quad \frac{1}{T}\sum_{t=1}^{T}\left(|\hat{\epsilon}_t| - \zeta\hat{\sigma}_t\right)^2,$$

where

$$\zeta \quad = \quad 2\frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)}\sqrt{\frac{\nu-2}{(\nu-1)^2\pi}}.$$

The scale factor ζ is necessary because, under the assumed distribution for ϵ_t, $\mathrm{E}|\epsilon_t| = \zeta\sigma$. Instead of presenting MSFEs, which have very little economic meaning by themselves, we present the proportion of the variance of absolute residuals that can be explained by the models' conditional volatility estimates, denoted R^2. This is computed from

$$R^2 \quad = \quad 1 - \frac{\text{MSFE}}{\frac{1}{T}\sum_{t=1}^{T}\left(|\hat{\epsilon}_t| - \overline{|\hat{\epsilon}_t|}\right)^2},$$

and is reported in the final row of Tables 1 and 2. In terms of forecasting power, the LEVELS model performs the worst for both datasets, while the new class of models we propose performs the best. Within our class, the asymmetric versions forecast much better than the symmetric versions.[15]

[14]Note that because the parameters were estimated using the entire dataset, we are not evaluating out-of-sample forecasting ability.

[15]We could account for the number of parameters in each model by using the "adjusted R^2", $\overline{R}^2$ =

B. Tradeoffs between the LEVELS and Unrestricted Models

When we consider two models where one is a generalization of the other, a natural question arises about the benefits of moving to the more general model. One self-evident benefit is that estimating a flexible functional form is more likely to yield the correct interest rate process. This is particularly important when pricing long-dated, path-dependent interest rate derivatives, such as index amortizing rate (IAR) swaps, CMO swaps, swaptions, mortgages, adjustable rate preferred securities, and CMOs. The values of these securities are sensitive to small changes in short-term interest rate volatility. To illustrate, consider the pricing of IAR swaps. These swaps have their notional principal reduced over time according to an amortization schedule based on the level of a reference interest rate on certain fixed dates in the future (typically every three or six months). The value of this swap depends on the probability distribution of the reference rate on each reset date. In addition, as the amount of principal remaining on any reset date is a function of past interest rate levels, IAR swaps are "path-dependent." This path depends on the volatility process, further amplifying the importance of accurately modeling the temporal behavior of interest rate volatility.

Clearly, our class of volatility processes will yield different volatility paths than, say, a square root process with a constant volatility parameter or the $\gamma = 1.5$ process of CKLS. To demonstrate that they also generate different probability distributions of future interest rate levels, we simulated both the LEVELS model and the TVP-LEVELS model 5000 times, using the TB13WK estimation results, with July 6, 1990 (the final date in TB13WK), as the starting date. On this day, the 13-week T-bill rate was 7.9 percent. To focus on the volatility process, we used the mean equation $r_t - r_{t-1} = 0.0022$ for all the simulations[16] and allowed only the volatility process to differ. Figures 5 and 6 graph the 5th, 25th, 50th, 75th, and 95th percentiles of the 5000 simulation paths for each horizon up to 180 weeks, for the LEVELS and TVP-LEVELS models, respectively. Two striking differences are: i) the narrower confidence intervals from the TVP-LEVELS model; and ii) the high skewness from the LEVELS model. The 25th and 75th percentiles are not very different, suggesting that the most significant differences between the two distributions are in their tails, especially the upper tail.

For some of the securities mentioned above (swaps, mortgages, and CMOs), average predicted amortization will be less with the LEVELS model than with the TVP-LEVELS model due to the larger upper tail. This results in longer predicted lives with the LEVELS model. Thus, the LEVELS model would lead to higher predicted cash flows, implying it would overprice these securities relative to the TVP models. Similarly, the LEVELS model would overprice adjustable rate preferred securities (due to higher average payout) and long-dated call options on interest rates (e.g., interest rate caps), while under pricing interest rate floors.

A potentially important remaining question is which of our new models is better for pricing derivatives like those mentioned above. To investigate this question, we performed the above simulation using all four of our new models: TVP-

$1-((T-1)/(T-k))(1-R^2)$, where k is the number of parameters in the volatility process. However, for our models, $\overline{R}^2 \approx R^2$ because the ratio $(T-1)/(T-k)$ is close to one.

[16]The average weekly change in the interest rate over our sample period was 0.0022.

FIGURE 5

Confidence Intervals for Simulated Interest Rate from LEVELS Model

Weekly 13-Week Yields

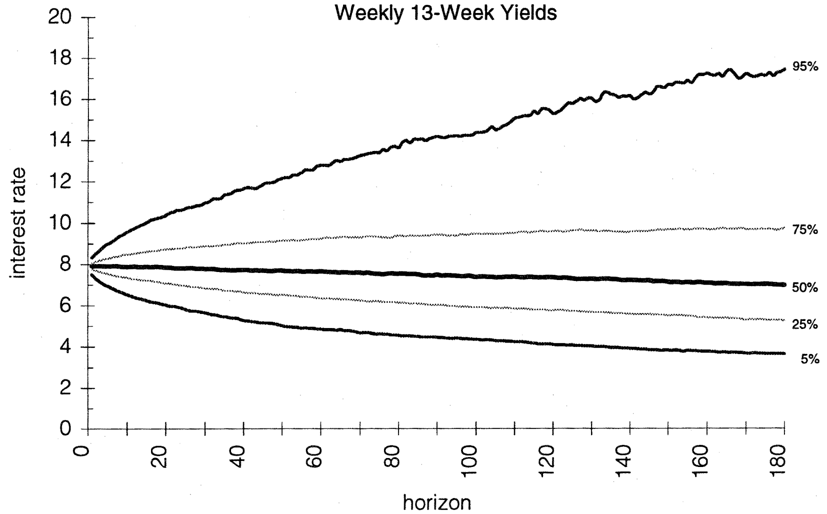
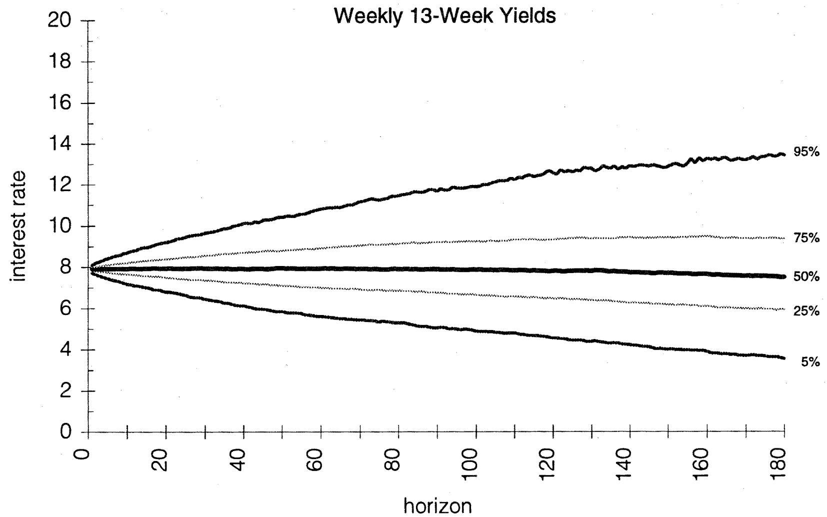


FIGURE 6

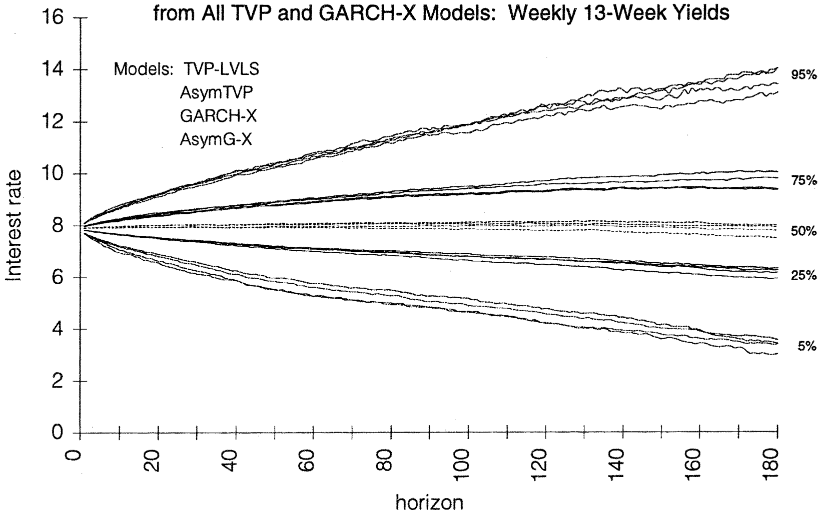
Confidence Intervals for Simulated Interest Rate from TVP-LEVELS Model

Weekly 13-Week Yields



LEVELS, AsymTVP, GARCH-X, and AsymG-X. The resulting distributions are overlaid on each other in Figure 7. These distributions are very similar, suggesting any one of the four proposed models will yield similar derivatives prices. This result is somewhat surprising, because it suggests, for example, that correctly modeling asymmetries will not greatly affect interest rate derivatives prices. It also suggests that the GARCH-X and TVP models will yield similar derivatives prices, even though these volatility processes are quite different. However, in light of Figure 4, we see that this conclusion holds only when the starting interest rate in the simulations is between about 7 percent and 14 percent. Similar simulations (not reported in the Figures) using starting values outside of this range resulted in substantially different future distributions, with the GARCH-X models having wider confidence intervals. Therefore, the choice of model can be crucial during times when interest rates are near either historical highs or lows.

FIGURE 7
Confidence Intervals for Simulated Interest Rates from All TVP and GARCH-X Models: Weekly 13-Week Yields



Though the benefits of our generalized models are clear, there is a potentially significant cost in that no simple analytical solution exists to the pricing of derivatives from these flexible interest rate processes. However, our empirical volatility processes set up naturally for Monte Carlo evaluation. Therefore, their generality over the existing models comes at little cost for the valuation of securities that already require Monte Carlo evaluation, such as the long-dated, path-dependent interest rate derivatives discussed above.

In summary, our results indicate that allowing volatility to be a function of both interest rate levels and shocks to the interest rate market could have a large impact on the pricing of long-dated interest rate based derivative securities. However, the added flexibility of our models may not be needed for short-lived securities. Our results suggest that volatility persists (for short horizons). Therefore, we conjecture that one could assume the volatility parameter is constant and calculate an implied volatility estimate. According to our results, the appropriate model used to extract implied volatilities should assume a square root-type volatility process. This implied volatility estimate would be one way to proxy for the average forecasted time-varying volatility parameter over the life of the option.

VI. Conclusions

In this paper, we look at two commonly used, empirical one-factor interest rate models. In the first, volatility is a function only of interest rate levels. We call these LEVELS models. In the second, volatility is a function of information shocks to interest rates. These are GARCH models. We present evidence that shows that GARCH models rely too heavily on serial correlation in variances and fail to capture the relationship between interest rate levels and volatility. On the other hand, the LEVELS specification over emphasizes the dependence of volatility on interest rate levels and fails to capture the serial correlation in conditional variances. Furthermore, the LEVELS model is not robust to the Fed-induced structural break or the inclusion of the 1987 crash. We propose an alternative class of models that captures both the serial correlation in variances and the dependence of variances on levels. We use traditional specification tests, robust conditional moment tests and tests of forecasting power to demonstrate that our class of models characterizes the volatility process better than either the LEVELS or GARCH models. We then show that the pricing of fixed income derivatives is sensitive to the volatility model used.

Two important conclusions emerge from this paper. First, the sensitivity of interest rate volatility to levels is exaggerated in the literature. We disagree with the conclusion of Chan, Karolyi, Longstaff, and Sanders (1992) that, "the relation between interest rate volatility and the level of r is the most-important feature of any dynamic model of the short-term riskless rate" (p. 1217). A comparison of the models presented in this paper indicates that, while this relationship *is* important, adequately modeling the volatility parameter as a function of unexpected "news" is equally important. A second conclusion is that existing theoretical models of interest rates are misspecified in the way they model volatility. Fortunately, our results suggest a potentially fruitful path by which to improve these models. A new generation of theoretical models might seek to account for both the relationship between interest rate levels and volatility *and* the relationship between interest rate shocks and volatility.

References

Ball, C. A., and W. N. Torous. "A Stochastic Volatility Model for Short Term Interest Rates and the Detection of Structural Shifts." Unpubl. Manuscript, Anderson School of Management, Univ. of California at Los Angeles (1994).

Bliss, R. R., and D. C. Smith. "Stability of Interest Rate Processes." Unpubl. Manuscript, Indiana Univ. (1994).

Bollerslev, T. "Generalized Autoregressive Conditional Heteroskedasticity." *Journal of Econometrics,* 31 (1986), 307–327.

Bollerslev, T.; R. T. Chou; and K. F. Kroner. "ARCH Modeling in Finance: A Review of the Theory and Empirical Evidence." *Journal of Econometrics,* 52 (1992), 5–59.

Brennan, M., and E. Schwartz. "Analyzing Convertible Bonds." *Journal of Financial and Quantitative Analysis,* 15 (1980), 907–929.

Broze, L.; O. Scaillet; and J.-M. Zakoian. "Testing for Continuous-Time Models of the Short Term Interest Rate." Unpubl. Manuscript, CORE, Belgium (1993).

Cai, J. "A Markov Model of Unconditional Variance in ARCH." *Journal of Business and Economic Statistics* (1994).

Cecchetti, S.; R. Cumby; and S. Figlewski. "Estimation of the Optimal Futures Hedge." *Review of Economics and Statistics,* 70 (1988), 623–630.

Chan, K. C.; G. A. Karolyi; F. A. Longstaff; and A. B. Sanders. "An Empirical Comparison of Alternative Models of the Short-Term Interest Rate." *Journal of Finance,* 47 (1992), 1209–1227.

Cox, J. "Notes on Option Pricing I: Constant Elasticity of Variance Diffusions." Unpubl. Manuscript, Stanford Univ. (1975).

Cox, J.; J. Ingersoll; and S. Ross. "An Analysis of Variable Rate Loan Contracts." *Journal of Finance,* 35 (1980), 389–403.

———. "A Theory of the Term Structure of Interest Rates." *Econometrica,* 53 (1985), 145–166.

Dietrich-Campbell, B., and E. Schwartz. "Valuing Debt Options: Empirical Evidence." *Journal of Financial Economics,* 16 (1986), 321–343.

Dothan, U. "On the Term Structure of Interest Rates." *Journal of Financial Economics,* 6 (1978), 59–69.

Engle, R. F. "Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of U.K. Inflation." *Econometrica,* 50 (1982), 987–1008.

Engle, R. F.; D. Lilien; and R. Robins. "Estimating Time Varying Risk Premia in the Term Structure:The ARCH-M Model." *Econometrica,* 55 (1987), 391–407.

Engle, R. F.; V. Ng; and M. Rothschild. "Asset Pricing with a Factor ARCH Covariance Structure: Empirical Estimates for Treasury Bills." *Journal of Econometrics,* 45 (1992), 213–238.

Evans, M. "Interpreting the Term Structure Using the Intertemporal Capital Asset Pricing Model: An Application of the Non-linear ARCH-M Model." Unpubl. Manuscript, New York Univ. (1989).

Flannery, M.; A. Hameed; and R. Harjes. "Asset Pricing, Time-Varying Risk Premia and Interest Rate Risk." Unpubl. Manuscript, Univ. of Florida (1992).

Gagnon, L.; I. G. Morgan; and E. H. Neave. "Pricing Eurodollar Time Deposit Futures." Unpubl. Manuscript, School of Business, Queen's Univ. (1993).

Gibbons, M. R., and K. Ramaswamy. "A Test of the Cox, Ingersoll and Ross Model of the Term Structure." *Review of Financial Studies,* 6 (1993), 619–658.

Hamilton, J. D., and R. Susmel. "Autoregressive Conditional Heteroskedasticity and Changes in Regime." Unpubl. Manuscript, Department of Economics, Univ. of California, San Diego (1992).

Heath, D.; R. A. Jarrow; and A. Morton. "Bond Pricing and the Term Structure of Interest Rates: A New Methodology for Contingent Claims Evaluation." *Econometrica,* 60 (1992), 77–105.

Kloeden, P. E., and E. Platen. *Numerical Solution of Stochastic Differential Equations.* New York, NY: Springer-Verlag (1992).

Longstaff, F., and E. Schwartz. "Interest Rate Volatility and the Term Structure: A Two-Factor General Equilibrium Model." *Journal of Finance,* 47 (1992), 1259–1282.

Marsh, T., and E. Rosenfeld. "Stochastic Processes for Interest Rates and Equilibrium Bond Prices." *Journal of Finance,* 38 (1983), 635–646.

Merton, R. "Theory of Rational Option Pricing." *Bell Journal of Economics and Management Science,* 4 (1973), 141–183.

Newey, W. K., and D. G. Steigerwald. "Consistency of Quasi-Maximum Likelihood Estimators for Models with Conditional Heteroskedasticity." Unpubl. Manuscript, Department of Economics, Univ. of California, Santa Barbara (1994).

Oldfield, G. S., and R. J. Rogalski. "The Stochastic Properties of Term Structure Movements." *Journal of Monetary Economics,* 19 (1987), 229–245.

Sanders, A., and H. Unal. "On the Intertemporal Behavior of the Short-Term Rate of Interest." *Journal of Financial and Quantitative Analysis,* 23 (1988), 417–423.

Vasicek, O. "An Equilibrium Characterization of the Term Structure." *Journal of Financial Economics,* 5 (1977), 177–188.

Wooldridge, J. M. "A Unified Approach to Robust, Regression-Based Specification Tests." *Econometric Theory*, 6 (1990), 17–43.