INTERNATIONAL ECONOMIC REVIEW
Vol. 35, No. 3, August 1994

AUTOREGRESSIVE CONDITIONAL DENSITY ESTIMATION*

BY BRUCE E. HANSEN[1]

Engle's ARCH model is extended to permit parametric specifications for conditional dependence beyond the mean and variance. The suggestion is to model the conditional density with a small number of "parameters," and then model these parameters as functions of the conditioning information. This method is applied to two data sets. The first application is to the monthly excess holding yield on U.S. Treasury securities, where the conditional density used is a student's t distribution. The second application is to the U.S. Dollar/Swiss Franc exchange rate, using a new "skewed student t" conditional distribution.

1. INTRODUCTION

A typical econometric problem is to obtain an approximation to the distribution of a variable y_t, conditional on another (vector-valued) variable x_t. This includes the dynamic context where x_t contains lagged values of y_t.

Most applications include estimates of the conditional mean,

$$\mu_t = E(y_t|x_t). \tag{1}$$

The conditional mean may be thought of as the leading term in the conditional distribution. Many econometric applications are concerned with nothing further than the mean. The remaining error

$$e_t = y_t - \mu_t$$

in these contexts is implicitly modeled as independent of x_t.

Many applications include as well estimates of the conditional variance

$$\sigma_t^2 = \sigma^2(x_t) = E((y_t - \mu_t)^2|x_t) \tag{2}$$

which may be thought of as the second term in the conditional distribution.

The conditional variance can be used to define the normalized error

$$z_t = \frac{e_t}{\sigma_t} = \frac{y_t - \mu_t}{\sigma_t}. \tag{3}$$

The normalized error z_t is a random variable whose conditional distribution is derived from the conditional distribution of y_t by the transformations (1) and (2). In most regression models, however, the conditional distribution of z_t is simply

* Manuscript received June 1992; revised April 1993.

[1] My thanks go to Adrian Pagan for numerous discussions which stimulated my interest in this area, to Douglas Steigerwald and two referees for helpful comments, and to the National Science Foundation for financial support.

assumed to be independent of the conditioning variable x_t. This is typical, for example, in the "ARCH" literature which has sprung from the pioneering work of Engle (1982). While a useful simplifying assumption, there is absolutely no reason to expect the conditional distribution of the derived variable z_t to be independent of the conditioning information. Another way of saying this is that there is no reason to assume, in general, that the only features of the conditional distribution which depend upon the conditioning information are the mean and variance. Indeed, it seems quite reasonable that other features of the distribution (such as skewness and kurtosis) will depend on the conditioning information. Gallant, Hsieh, and Tauchen (1991) have made a clever argument of this form. They show that if the innovations e_t are generated by the mixture model $e_t = I_t^{1/2}\xi_t$ where ξ_t is iid and independent of I_t, then the variance of e_t, conditional on the past history of e_t alone, will not (in general) equal I_t, and thus the normalized error z_t will generally have a nonconstant conditional distribution. In fact, this is an implication of most stochastic volatility models. For various approaches to the latter see Andersen (1992), Shephard (1994), and Watanabe (1992).

The reason why most applications have ignored higher-order features of the conditional distribution may be because only the conditional mean and variance generate significant excitement. But this lack of excitement does not imply that higher-order features should be completely ignored. First, efficient estimation of the equations for the conditional mean and variance require a complete description of the conditional distribution. Second, the aim of conditional models is often prediction, and the accuracy of predictive distributions is critically dependent upon knowledge of the correct conditional distribution for the normalized error. This point has been recently made in Baillie and Bollerslev (1992). Third, empirical models of asset pricing are incomplete unless the full conditional model is specified. Full specification may be especially important in the context of options pricing, where the price is determined by not just the conditional mean and variance, but more complicated functions of the conditional distribution.

While it might be agreed that it is desirable to allow the conditional density of z_t to depend on x_t, it is probably not clear at all how to achieve this goal. One approach, offered by Gallant, Hsieh, and Tauchen (1991), is to model the joint density of y_t and x_t using a series expansion about the Gaussian density. This is an innovative approach, and has the potential to reveal a lot of information concerning the underlying distribution without having to impose a great deal of a priori information or structure. Their approach has several drawbacks, however. First, their parameterization is not parsimonious, and therefore requires very large data sets in order to achieve a reasonable degree of precision. Second, the methods are computationally expensive, and may lay outside the reach of many routine applications. Third, the techniques may be sensitive to choices of the number of expansion terms. Theorists have not yet solved many questions concerning implementation and the selection of the order of the expansion. As a result, these techniques will probably remain primarily in the hands of specialists.

This paper suggests an alternative parametric approach to modeling the conditional density of the normalized error. The approach may be regarded as a direct extension of Engle's idea to model the conditional variance as a function of lagged errors. My suggestion is to select a distribution which depends upon a low-dimensional parameter vector, and then let this "parameter vector" vary as a function of the conditional variables. In general, any distribution which has a closed-form density function may be used. In the applications presented in this paper, the student's t density and a generalization which allows for skewness are used.

It is useful to briefly review the relationship of this paper with the previous literature. Several density functions for the normalized error have been proposed beyond the Gaussian distribution originally used by Engle (1982). Many authors, including Engle and Bollerslev (1986) have used the conditional student's t distribution. Spanos (1991) suggested the multivariate t distribution, which yields conditional heteroskedasticity as a natural implication. Nelson (1989) used the generalized exponential distribution. Liu and Brorsen (1992) used a stable distribution. Engle and Gonzalez-Rivera (1991) used a nonparametric conditional distribution. While these authors have used different conditional distributions, they all have made the latter (equivalently, the shape parameters) constant over time. None have allowed the shape of distribution to change over time. In addition, few have allowed for skewness. The applications of this paper allow for both time-varying shapes in the conditional density, and for skewness in the density function.

This method is applied to two financial data set. The first is the excess holding yield on U.S. Treasury securities. The second is the Dollar/Franc exchange rate. In both applications strong evidence is found for variation in the conditional distribution beyond the mean and variance. A GAUSS386 program (requires the OPTMUM module) which replicates the empirical results is available on request from the author.

2. ARCD MODEL

2.1. *Probability Model.* The observed sample is $(y_t, x_t\colon t = 1, \ldots, n)$ which is assumed to be a realization of some jointly stationary process. We do not need to restrict the variables x_t to lie in a finite-dimensional space, so we can allow, for example, the variable x_t to include all of the (observed) past values of y_t.

We will restrict attention to distribution functions which have densities which can be written in the form

$$f(y|\alpha(x_t, \theta)) = \frac{d}{dy} P(y_t \leq y|x_t) \tag{4}$$

where θ is a finite-dimensional parameter vector and

$$\alpha_t = \alpha(x_t, \theta)$$

is a low-dimensional "time varying parameter" which fully describes the influence of x_t upon the conditional distribution. When the dimension of x_t is constant and finite, there is of course no loss in generality in writing the density function in this form, but when x_t is infinite dimensional or has a dimension which depends on t, then this class represents a meaningful restriction of the class of potential models.

For reasons which will become apparent, we will denote this class of models by the name "autoregressive conditional density models" (ARCD).

2.2. *Normalized Parameterizations.* It is particularly convenient for the reporting of applied research to rewrite the density function in terms of location and scale parameters. I will restrict attention in this exposition to cases where the location parameter is the conditional mean, and the scale parameter is the conditional variance, but the generalizations to cases where the mean or variance does not exist is straightforward and merely involves changes in notation. The idea is to parameterize the function $f(y|\alpha)$ so that we have the partition

$$\alpha_t = (\mu_t, \sigma_t^2, \eta_t)$$

where

$$\mu_t = \mu(\theta, x_t) = E(y_t|x_t) \tag{5}$$

is the conditional mean,

$$\sigma_t^2 = \sigma^2(\theta, x_t) = E((y_t - \mu_t)^2|x_t) \tag{6}$$

is the conditional variance, and

$$\eta_t = \eta(\theta, x_t)$$

contain the remaining parameters of the conditional distribution, which we will sometimes refer to as "shape" parameters.

The conditional mean and variance allow us to define the normalized variable

$$z_t(\theta) = \frac{y_t - \mu(\theta, x_t)}{\sigma(\theta, x_t)}. \tag{7}$$

We will denote the conditional density function for z_t by

$$g(z|\eta_t) = \frac{d}{dz} P(z_t < z|\eta_t) \tag{8}$$

say. Densities (4) and (8) are related by

$$f(y_t|\mu_t, \sigma_t^2, \eta_t) = \frac{1}{\sigma_t} g(z_t|\eta_t).$$

Most ARCH-type applications use probability models of the form (5) through (8), but with η_t assumed to be time invariant. The ARCD modeling strategy simply builds on this foundation by allowing the shape parameters of the density function to be time varying as well.

This formalization is convenient since there is a large literature which concerns the specification of the mean equation (5) and the variance equation (6). Parametric models include ARCH, GARCH, E-GARCH, N-ARCH, A-ARCH, plus ARCH-M versions of each (see Hentshcel 1991 for a recent summary). Nonparametric models for the mean and variance equations have also been suggested, as in Pagan and Hong (1991) and Gourerioux and Monfort (1992).

2.3. *Flexible Density Functions.* The goal is desirable to select a density function $g(z|\eta)$ which generalizes the standard normal, is sufficiently flexible to generate the range of shapes which we think might be relevant in a particular application (such as heavy tails, skewness, or bi-modality), and yet is sufficiently parsimonious that η can be adequately modeled using time series techniques. To facilitate quasi-likelihood estimation, it is also important that the density $g(z|\eta)$ be available in closed form. Otherwise, estimation may be infeasible.

The statistics literature contains many flexible low-dimensional parametric distributions. But few have closed form density functions. For example, the well-known Pearson family can only be expressed as the solution to a differential equation. Other distributions, such as the Tukey-λ family, are only represented through their quantile function (the inverse of the distribution function). For a recent flexible generalization of the Tukey family see Ramberg, Tadikamalla, Dudewicz, and Mykytka (1979). Since the density function is unavailable, estimation of these models typically uses the method of moments or a close relative, rather than maximum likelihood.

In principle, a method of moments approach could be combined with the ARCD strategy proposed here. ARCH-styled models could be fit to a set of moments of the data, giving time-varying estimates of the conditional moments, and then could be inverted to find the matching parameters of the distribution. This would be a very different approach from that explored in this paper, but certainly is worth considering. I believe, however, that estimation based on the method of moments in potentially integrated GARCH models might involve severe inferential difficulties. The recent asymptotic distribution theory for GARCH models (see Lee and Hansen 1994) relies heavily on the fact that the likelihood scores are inversely proportional to the conditional variance, so that the scores will have bounded unconditional second moments, even though the data itself does not. Examining the nature of the proofs, it seems difficult to believe that a normal asymptotic distribution theory will hold for naive method of moments estimation in the presence of integrated variances. Until this distributional question is solved, it appears wise to avoid empirical strategies based on the method of moments.

2.4. *A Skewed Student's t Density.* The approach taken in this paper will be to rely on the popular student's t density and a simple skewed generalization. Recall, the student's t density (normalized to have unit variance) is

$$g(z|\eta) = \frac{\Gamma\left(\frac{\eta+1}{2}\right)}{\sqrt{\pi(\eta-2)}\,\Gamma\left(\frac{\eta}{2}\right)}\left(1+\frac{z^2}{(\eta-2)}\right)^{-(\eta+1)/2} \tag{9}$$

where $2 < \eta < \infty$. This density function is the basis for the empirical work reported in Section 3.

The student's t is a fairly restrictive parametric family, only allowing for variation in the location, scale, and tail thickness. To allow for a richer set of behaviors, we may need a more flexible family of probability densities. A minimal desirable extension is to allow for skewness. In order to keep in the ARCH tradition, it is also important to have density functions which can be easily parameterized so that the innovations are mean zero and unit variance. Otherwise, it will be difficult to separate the fluctuations in the mean and variance from the fluctuations in the shape of the conditional density.

Consider the following density function:

$$g(z|\eta, \lambda) = \begin{cases} bc\left(1 + \frac{1}{\eta - 2}\left(\frac{bz + a}{1 - \lambda}\right)^2\right)^{-(\eta+1)/2} & z < -a/b, \\ bc\left(1 + \frac{1}{\eta - 2}\left(\frac{bz + a}{1 + \lambda}\right)^2\right)^{-(\eta+1)/2} & z \geq -a/b, \end{cases} \tag{10}$$

where $2 < \eta < \infty$, and $-1 < \lambda < 1$. The constants a, b, and c are given by

$$a = 4\lambda c\left(\frac{\eta - 2}{\eta - 1}\right), \tag{11}$$

$$b^2 = 1 + 3\lambda^2 - a^2, \tag{12}$$

and

$$c = \frac{\Gamma\left(\frac{\eta + 1}{2}\right)}{\sqrt{\pi(\eta - 2)}\,\Gamma\left(\frac{\eta}{2}\right)}. \tag{13}$$

In the Appendix, we show that this is a proper density function with a mean of zero and a unit variance. This "skewed student's t distribution specializes to the student's t distribution (9) by setting $\lambda = 0$.

Inspection of the density function reveals that the density is continuous, and has a single mode at $-a/b$, which is of opposite sign with the parameter λ. Thus if $\lambda > 0$, the mode of the density is to the left of zero and the variable is skewed to the right, and vice-versa when $\lambda < 0$. Figure 1 displays plots of the density for a few parameterizations.

It should be emphasized that the general approach advocated here does not depend upon these particular choices for the density function. Other choices may have better properties in particular applications. We use density (10) for the empirical study reported in Section 4.

2.5. *Specification of Laws of Motion for Shape Parameters.* It is necessary to specify laws of motion for the "parameters" α_t. Many strategies are possible, but the one suggested here is to follow the lead of Engle (1982). Engle's ARCH model and its generalizations have all made σ_t^2 a function of the lagged errors

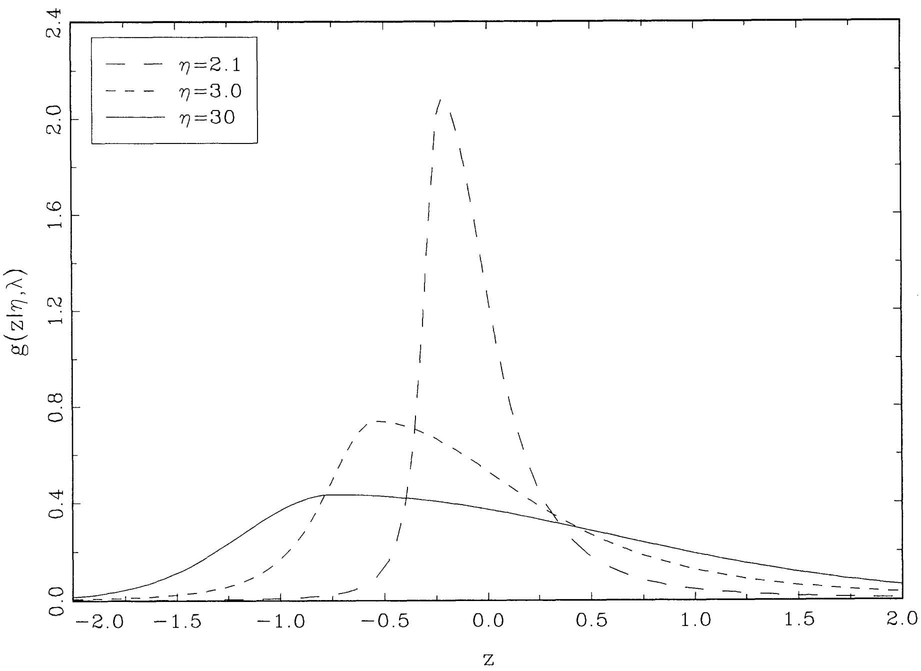

FIGURE 1
SKEWED STUDENT t DENSITY $\lambda = 0.5$

$$e_t = y_t - \mu_t.$$

Since this approach has been empirically successful for the conditional variance, then it seems reasonable to believe that this strategy could also work well for other time-varying parameters in η_t. That is, the proposed modeling strategy will be to specify laws of motion of the form

$$\eta_t = \eta(e_{t-1}, e_{t-2}, \dots, e_1).$$

As in the ARCH literature, we have to pay attention to boundary constraints. The conditional variance, for example, is constrained to be positive. Thus specifications of the form $\sigma_t^2 = a + be_{t-1}$ are avoided since they cannot guarantee positivity of the estimated variance sequence. One common solution (in this context) is to use specifications of the form $\sigma_t^2 = a + be_{t-1}^2$. Another solution is to use an appropriate transformation of the variance, such as $\ln \sigma_t^2 = a + be_{t-1} + ce_{t-1}^2$. Both methods have been used in the ARCH literature.

This constraint problem will certainly arise in the general ARCD context. Shape parameters arising from typical density functions often need to lie in restricted regions of the real line. Without the guidance of a priori theory, there is no uniformly correct approach, but a practical method which will "work" is to use a logistic transformation. Suppose that η_t is real valued and is related to a variable λ_t as

$$\eta_t = L + \frac{(U - L)}{1 + \exp(-\lambda_t)}.$$

Even if λ_t is allowed to vary over the entire real line, η_t will be constrained to lie in the region $[L, U]$. L and U should be chosen to reflect the region of interest for η_t. Combined with a law of motion for η_t such as

$$\lambda_t = a + be_{t-1} + ce_{t-1}^2$$

we obtain a relationship $\eta_t = \eta(e_{t-1})$ which is flexible yet constrained to the region $[L, U]$.

2.6. *Estimation and Inference.* We can write the conditional log-likelihood function as

$$\ln L(\theta | x_1, x_2, \dots, x_n) = \sum_{t=1}^{n} l_t(\theta) \tag{14}$$

where

$$l_t(\theta) = \ln g(z_t(\theta) | \eta_t(\theta)) - \ln \sigma(\theta, x_t).$$

The maximum likelihood estimate (MLE) of the model is the value $\hat{\theta}$ which maximizes the conditional log-likelihood (14). The optimum may be found using an appropriate optimization technique.

Under the assumption of correct specification, the likelihood scores

$$\frac{\partial}{\partial \theta} l_t(\theta) = \frac{\partial}{\partial \theta} \ln g(z_t(\theta) | \eta_t(\theta)) - \frac{\partial}{\partial \theta} \ln \sigma(\theta, x_t)$$

are martingale differences and have variance

$$V = V(\theta_0), \qquad V(\theta) = E \frac{\partial}{\partial \theta} l_t(\theta) \frac{\partial}{\partial \theta} l_t(\theta)' = -E \frac{\partial}{\partial \theta \partial \theta'} l_t(\theta),$$

where θ_0 denotes the true parameter value. If $El_t(\theta) < \infty$ and $E\,(\partial/\partial\theta)\, l_t(\theta) < \infty$ uniformly in θ then the MLE will be consistent. If $V < \infty$ as well and the likelihood is sufficiently well behaved in the neighborhood of θ_0 then the MLE will be asymptotically normal as well. While these are not unreasonable expectations, it is my expectation that a rigorous proof will be quite difficult to accomplish in this general setting. Lumsdaine (1991) established consistency and asymptotic normality for the Gaussian GARCH(1,1) quasi-MLE under the assumption that z_t is iid with 32 finite moments. Lee and Hansen (1994) achieved a similar result under the weaker condition that z_t is a stationary martingale difference with a bounded conditional fourth moment. Lee (1993) extended these results to incorporate the Gaussian GARCH-M model. All of these papers have confined attention to the case in which the conditional density used for estimation is the standard normal. Extension of this theory to cover the general context considered here would be desirable, but beyond the scope of the present study. We will simply assume that such theorems hold, and proceed conventionally.

Since any particular probability model is unlikely to be the "correct" model, but should more accurately be viewed as an approximation to the underlying probability structure, it is reasonable to report "robust" standard errors, as suggested by White (1982), in addition to the more conventional standard errors. These give asymptotically valid confidence intervals for the "pseudo-true" parameter values which minimize the information distance between the true probability measure and the quasi-likelihood. The robust standard errors are the square roots of the diagonal elements of the matrix

$$\hat{\Omega} = \hat{M}^{-1}\hat{V}\hat{M}^{-1}$$

where

$$\hat{M} = -\sum_{t=1}^{n} \frac{\partial}{\partial\theta\partial\theta'} l_t(\hat{\theta})$$

and

$$\hat{V} = \sum_{t=1}^{n} \frac{\partial}{\partial\theta} l_t(\hat{\theta}) \frac{\partial}{\partial\theta} l_t(\hat{\theta})'.$$

2.7. *Parameter Constancy.* A parameter constancy test has been introduced by Lee and Hansen (1992) which is particularly easy to implement. The test statistic is a member of the family of tests introduced by Nyblom (1989) and modified by Hansen (1990). The statistic is an approximate LM test of the null that the parameters θ are constant against the alternative that the parameters θ follow a martingale process. The statistic is based on the cumulative moments

$$S_t = \sum_{i=1}^{t} \frac{\partial}{\partial\theta} l_i(\hat{\theta})$$

and takes the form

$$L = \frac{1}{n} \sum_{t=1}^{n} S_t'\hat{V}^{-1}S_t.$$

Under the same regularity conditions which guarantee asymptotic normality of the pseudo-MLE, the statistic L has an asymptotic distribution which depends only on the number of parameters in θ. This distribution is tabulated in Nyblom (1989) and Hansen (1990). The statistic L tests the null that the entire vector θ is stable against the alternative that the entire vector may be unstable. A statistic which tests the stability of an individual parameter is given by

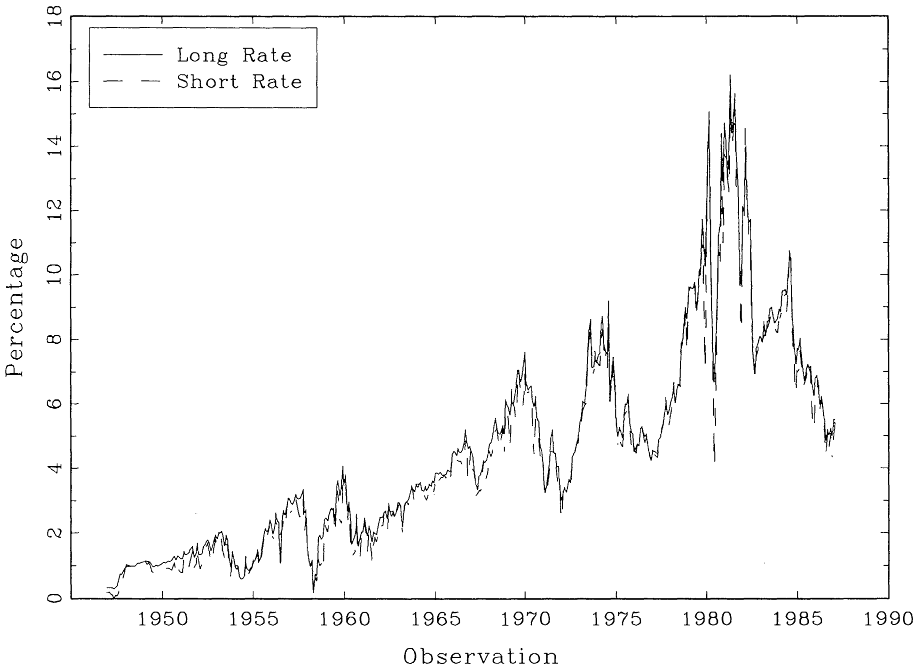

FIGURE 2
INTEREST RATES

$$L_k = \frac{1}{n} \sum_{t=1}^{n} S_{kt}^2 / \hat{V}_{kk}$$

where S_{kt} is the kth element of S_t and $\hat{V}_{kk}$ is the kth diagonal element of $\hat{V}$. The asymptotic 1 percent critical value for the individual statistics is 0.75, and the asymptotic 5 percent critical value is 0.47.

3. A CONDITIONAL STUDENT MODEL FOR THE TERM STRUCTURE

3.1. *Basic Structure.* This section describes a study concerning the short-run term structure of interest rates. The data, monthly observations on returns to U.S. Treasury securities for the period December 1946 to February 1987, come from Table 13.A.1 of McCulloch (1990). His returns series were calculated from the prices of whole securities, and were adjusted for changes in tax legislation. Figure 2 plots the one-month yield rate R_t and the instantaneous yield rate r_t.

From his tables, the excess holding yield, y_t was calculated as

$$y_t = \frac{(1 + R_t)^2}{1 + r_{t+1}} - (1 + r_t),$$

and the interest differential, i_t, was calculated as

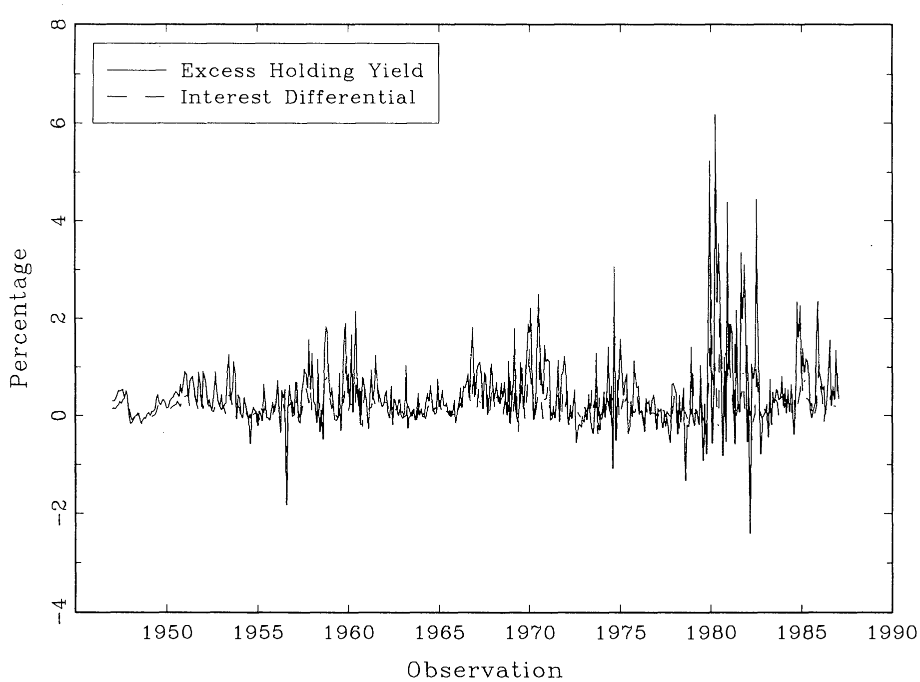

FIGURE 3
TRANSFORMED VARIABLES

$$i_t = R_t - r_t.$$

These two series are displayed in Figure 3.

In our earlier notation, $x_t = (y_{t-1}, y_{t-2}, \dots ; i_t, i_{t-1}, \dots)$, since we are interested in obtained the distribution of the excess holding yield, conditional on the current interest differential and lagged values of these two series. As discussed in Engle, Lilien and Robins (1987) and Pagan and Hong (1991), the interest differential plays an important role in empirical models of the excess holding yield, even though the expectations hypothesis implies otherwise.

3.2. *Specification of the Conditional Mean.* The main thrust of this exercise is not on the conditional mean or variance, but is to demonstrate that allowing for higher-order dependence yields significant gains. Yet the specification of the mean and variance equations cannot be taken lightly, for it is clear that errors in their specification may result in spurious higher-order findings. At the same time, it is important (from both computational and precision viewpoints) not to heavily overparameterize the model. The approach adapted in this application is to model the equations sequentially, using the vehicle of the Gaussian likelihood to select the equations for the mean and variance. This will enable us to feasibly estimate and compare a large number of models. As discussed above, it is known that the Gaussian quasi-MLE is consistent (and asymptotically normal) for the parameters of the correctly specified conditional mean and variance functions.

TABLE 1
EXCESS HOLDING YIELD: UNRESTRICTED GAUSSIAN MODEL

Variables	Estimate	St. Error	Robust SE	Nyblom L_k
Mean Equation				
Intercept	0.02	0.03	0.03	0.05
σ_t	0.09	0.11	0.11	0.07
i_t	1.17	0.15	0.17	0.17
i_t^2	−0.78	0.23	0.23	0.04
i_{t-1}	0.25	0.14	0.16	0.03
i_{t-1}^2	0.14	0.19	0.18	0.03
i_{t-2}	0.24	0.14	0.16	0.10
i_{t-2}^2	−0.12	0.15	0.12	0.36
i_{t-3}	0.12	0.14	0.16	0.05
i_{t-3}^2	0.26	0.11	0.12	0.03
y_{t-1}	0.03	0.09	0.10	0.12
y_{t-1}^2	0.10	0.05	0.04	0.08
$i_t y_{t-1}$	0.25	0.16	0.17	0.04
$i_{t-1} y_{t-1}$	−0.35	0.18	0.16	0.04
$i_{t-2} y_{t-1}$	0.02	0.17	0.12	0.07
$i_{t-3} y_{t-1}$	−0.19	0.17	0.15	0.05
Variance Equation				
Intercept	−0.00001	0.0008	0.0010	0.09
$e_{t-1}^2 - \sigma_{t-1}^2$	0.21	0.06	0.10	0.12
i_t^2	0.07	0.03	0.03	0.26
σ_{t-1}^2	1.01	0.02	0.04	0.04
Log L	323.7			
Nyblom L	4.12			

Table 1 reports the Gaussian maximum likelihood estimates of a fairly general specification of the conditional mean, with a fairly simple specification of the conditional variance. In all of the tables, the maximum likelihood estimates, the conventional standard errors, and the White robust standard errors are reported. The Nyblom L_k statistics for each parameter are reported. In the variance equation, the variance is reported as a linear function of σ_{t-1}^2 and $e_{t-1}^2 - \sigma_{t-1}^2$. This was done so that the coefficient on the former can be interpreted as a measure of persistence in the variance. The point estimate is 1.01, which indicates an integrated (persistent) conditional variance. This is consistent with a large volume of GARCH studies, as documented in Engle and Bollerslev (1986).

In Table 1, a large number of the individual coefficients appear insignificantly different from zero. A more parsimonious model was selected by successfully eliminating the variable with the smallest t-statistic, until the model reported in Table 2 was obtained. The only exceptions to the smallest t-statistic rule were that the intercept was always maintained, and the conditional standard deviation was retained until the final step. The latter was done since the possibility of a significant "GARCH-M" effect has long been believed to be important for the excess holding yield on Treasury securities. The model of Table 2 has eight fewer parameters than the model of Table 1, with an increase in the log-likelihood of only 3.04, which is far from a statistically significant difference.

It is interesting to compare these results with an alternative, simpler specification reported in Table 3. The major difference is that only the current value of the interest differential is included in the conditional mean equation. In this specifica-

TABLE 2
EXCESS HOLDING YIELD: RESTRICTED GAUSSIAN MODEL

Variables	Estimate	St. Error	Robust SE	Nyblom L_k
Mean Equation				
Intercept	0.04	0.23	0.25	0.04
i_t	1.24	0.12	0.13	0.13
i_t^2	−0.51	0.14	0.13	0.10
i_{t-1}	0.35	0.12	0.12	0.08
i_{t-2}	0.19	0.09	0.10	0.11
i_{t-3}^2	0.30	0.09	0.09	0.03
y_{t-1}^2	0.12	0.09	0.03	0.12
$i_{t-1}y_{t-1}$	−0.29	0.03	0.10	0.07
Variance Equation				
Intercept	−0.0001	0.0007	0.0009	0.10
$e_{t-1}^2 - \sigma_{t-1}^2$	0.22	0.06	0.10	0.14
i_t^2	0.07	0.03	0.03	0.30
σ_{t-1}^2	1.01	0.03	0.04	0.04
Log L	326.7			
Nybolm L	2.46			

tion, the conditional standard deviation appears to be statistically significant in the mean equation, as is commonly found in this literature. Note that the likelihood ratio statistic for this restricted model is 38.4, which is statistically significant at the 1 percent level. This restricted model also fails the Nyblom-Hansen parameter stability test. The L statistic of 3.6 exceeds the 1 percent null critical value of 2.6. The individual stability tests suggest that the coefficient on i_t is not stable. Note that these problems do not arise for the general models of Tables 1 and 2, where extra lags of the interest differential are included. An important lesson here is that the stability tests are useful diagnostics. If the model of Table 3 were estimated first, the large stability test statistics would alert a careful researcher that further study of the dynamic specification is needed.

Another interesting contrast between the models of Table 3 and Tables 1 and 2 is the difference between the conventional standard errors and the robust standard errors. In Tables 1 and 2 the two estimates are nearly the same, but in Table 3 the estimates are quite different. This is also informal evidence against the specification

TABLE 3
EXCESS HOLDING YIELD: NAIVE GAUSSIAN MODEL

Variables	Estimate	St. Error	Robust SE	Nyblom L_k
Mean Equation				
Intercept	0.02	0.04	0.06	0.10
σ_t	0.32	0.11	0.16	0.06
i_t	0.99	0.13	0.17	0.70
y_{t-1}	0.10	0.06	0.07	0.42
Variance Equation				
Intercept	0.0004	0.0013	0.0017	0.20
$e_{t-1}^2 - \sigma_{t-1}^2$	0.21	0.04	0.07	0.09
i_t^2	0.16	0.07	0.10	0.34
σ_{t-1}^2	0.97	0.03	0.05	0.08
Log L	342.9			
Nyblom L	3.57			

TABLE 4
EXCESS HOLDING YIELD: STUDENT t MODEL

Variables	Estimate	St. Error	Robust SE	Nyblom L_k
Mean Equation				
Intercept	0.04	0.02	0.02	0.05
t_t	1.17	0.13	0.16	0.32
i_t^2	−0.45	0.15	0.16	0.03
i_{t-1}	0.35	0.10	0.10	0.05
i_{t-2}	0.19	0.09	0.10	0.24
i_{t-3}^2	0.28	0.09	0.11	0.02
y_{t-1}^2	0.11	0.04	0.03	0.05
$i_{t-1}y_{t-1}$	−0.29	0.12	0.11	0.04
Variance Equation				
Intercept	−0.00008	0.00090	0.0010	0.13
$e_{t-1}^2 - \sigma_{t-1}^2$	0.20	0.07	0.10	0.10
i_t^2	0.11	0.05	0.07	0.31
σ_{t-1}^2	0.99	0.03	0.04	0.06
Degrees of Freedom	5.7	1.56	1.60	.16
Log L	315.6			
Nyblom L	3.16			

(this informal comparison could be made rigorous using a White information matrix test).

For the rest of the analysis, we will use the specification for the conditional mean and variance as given in Table 2. The specification of the conditional variance was also examined. Additional lags of the e_{t-1}^2 and i_t^2 were also included, but were not statistically significant and so the model was not augmented. It appears that the model reflected in Table 2 provides a good specification for the conditional mean and variance. We now turn to modeling other features of the conditional distribution.

3.3. *Student t Likelihood.* We start with a conventional student's t model with a constant degrees of freedom parameter. The MLE for this model are given in Table 4. The parameter estimates and standard errors for the conditional mean and variance are not dramatically different than those from the Gaussian MLE. The degrees of freedom parameter is estimated to be 5.7, which implies a fairly fat tail. The fit of the model is a dramatic improvement over the Gaussian, with the log-likelihood changing by 11.1.

3.4. *Conditional Student Likelihood.* We next allowed the degrees of freedom parameter to be time-varying. A logistic function was used to bound the time-varying conditional degrees of freedom parameter to lie between a lower bound of 2.1 and an upper bound of 30. The upper bound was selected simply because the student's t distribution is virtually indistinguishable from the standard normal for any value of η above 30. The lower bound is perhaps more critical. The normalized student's t density is not defined for $\eta = 2$, so needs to be bounded away from 2. Some visual experimentation suggested that setting $L = 2.1$ wasn't too extreme a choice, and the numerical operations didn't appear to find this choice offensive. The function was completed by making the logistically transformed η_t a quadratic function of the information set. The complete specification is

TABLE 5
EXCESS HOLDING YIELD: CONDITIONAL STUDENT t MODEL

Variables	Estimate	St. Error	Robust SE	Nyblom L_k
Mean Equation				
Intercept	0.02	0.02	0.02	0.05
i_t	1.14	0.11	0.14	0.30
i_t^2	−0.34	0.08	0.07	0.03
i_{t-1}	0.42	0.10	0.09	0.07
i_{t-2}	0.16	0.09	0.10	0.27
i_{t-3}^2	0.28	0.09	0.11	0.02
y_{t-1}^2	0.12	0.03	0.02	0.06
$i_{t-1}y_{t-1}$	−0.35	0.10	0.10	0.03
Variance Equation				
Intercept	0.00003	0.00112	0.001142	0.14
$e_{t-1}^2 - \sigma_{t-1}^2$	0.23	0.08	0.12	0.13
i_t^2	0.09	0.05	0.05	0.25
σ_{t-1}^2	1.03	0.04	0.05	0.05
Degrees of Freedom				
Intercept	−2.44	0.55	0.60	0.08
e_{t-1}	−0.23	0.66	0.48	0.23
e_{t-1}^2	−0.05	0.37	0.23	0.07
i_t	3.33	1.97	1.94	0.14
i_t^2	3.27	2.59	2.64	0.04
$e_{t-1}i_t$	−4.14	2.44	2.39	0.03
Log L	309.2			
Nyblom L	3.81			

$$\frac{\eta_t - 2.1}{27.9} = \frac{1}{1 + \exp(-\lambda_t)}$$

$$\lambda_t = \lambda_0 + \lambda_1 e_{t-1} + \lambda_2 e_{t-1}^2 + \lambda_3 i_t + \lambda_4 i_t^2 + \lambda_5 e_{t-1} i_t. \tag{15}$$

This function is quite flexible and will allow for a wide range of relationships.

To optimize the global likelihood, I found that it was easiest to first use the normalized residuals from the previously estimated model, and fit equation (15) alone. This provided a good set of starting values for the complete likelihood.

The estimates are reported in Table 5. Most of the coefficient estimates of the mean and variance equations are quite similar to those of Table 4, and most of the standard errors are smaller. In particular, note that the estimate of "persistence in variance" (the coefficient on σ_{t-1}^2 in the variance equation) is 1.03, indicating that the allowance for time-variation in the student's t parameter does not change the finding of integration in the variance. The likelihood ratio statistic against the conventional student's t model has an asymptotic p-value of 2.5 percent. Noting some concern with relying on the validity of the asymptotic approximation, we interpret this as evidence against the assumption that the conditional distribution of the normalized errors is independent of the conditioning information. This particular model (the conditional student t) may not be the "truth," but it does appear to give a statistically significant increase in fit, and therefore a better description of the time series process for excess holding yields.

Parameter estimates from tables often do not give a good feel between condi-

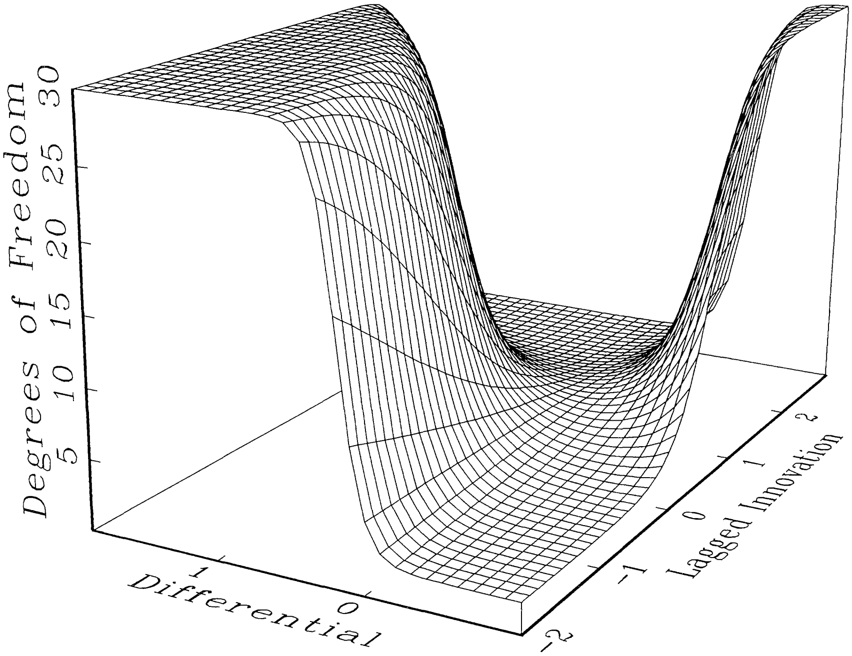

FIGURE 4
ESTIMATED RESPONSE FUNCTION

tioning variables and the objects of interest, and this is certainly true concerning the estimated relationship for the degrees of freedom, so I have displayed the nonlinear relationship in a 3-D graph in Figure 4. The vertical axis gives the estimated degrees of freedom, and the other axes the interest differential and lagged residual. It is easy to see a strong quadratic effect in the interest differential (so the degrees of freedom is small for i_t near zero), and a more mild quadratic effect in e_{t-1}.

Figure 5 displays the estimated degrees of freedom parameter over the sample period. Note that most of the estimates are close to 5, with some visits down to the lower boundary of 2.1 (implying a very fat tailed distribution) and some up towards, and even hitting, the upper boundary of 30 (implying a near-Gaussian distribution). Unfortunately, the "degrees of freedom" parameterization disguises some information, since the shape of the density is much more sensitive to changes in η when η is small than when it is large. The plot of Figure 5 emphasizes the large movements between 10 and 30, which are probably less significant than the movements between 2 and 3. To alleviate this deficiency, we plot in Figure 6 the inverse of the degrees of freedom, $1/\eta_t$. In this picture, the lower boundary, 0, represents a Gaussian density, and the upper boundary, 1/2, represents the limit of the fattailed densities. Another method to assess the behavior of the estimated process for the degrees of freedom parameter is to estimate its unconditional density. This is shown in Figure 7. This shows clearly that η_t is typically close to the modal value, 5.

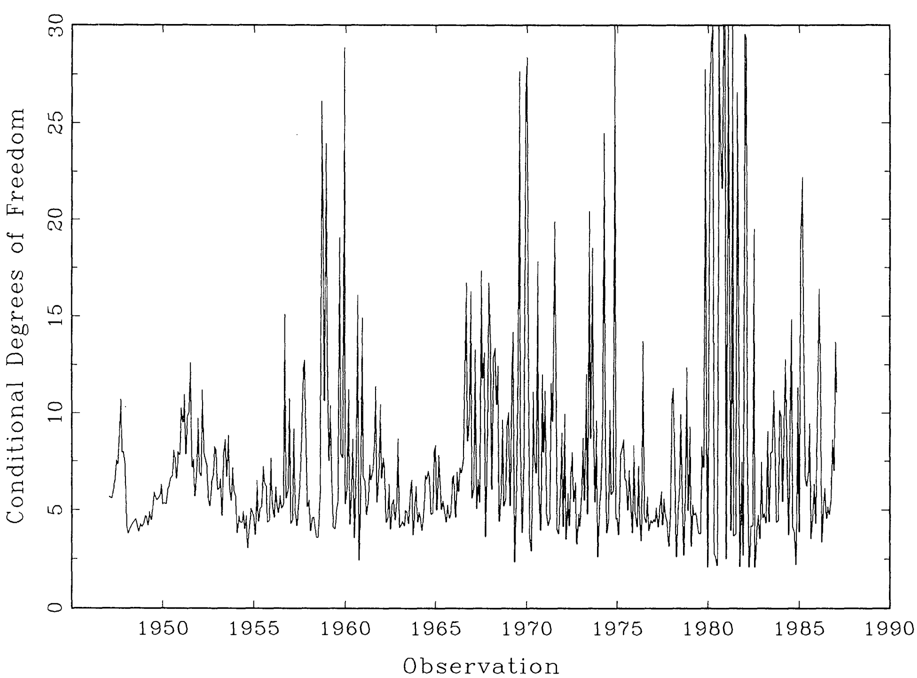

FIGURE 5

INTEREST RATES: CONDITIONAL DEGREE OF FREEDOM

4. A SKEWED STUDENT'S t MODEL FOR THE EXCHANGE RATE

One commonly analyzed series in the ARCH literature is the weekly Dollar/Swiss Franc exchange rate for July 1973 through August 1985. Engle and Bollerslev (1986) studied this series, and suggested a GARCH(1,1) specification with a student's t density. Maximum likelihood estimates for this specification are given in Table 6. While this model survives a number of standard specification tests (such as tests for omitted variables) the degrees of freedom parameter decisively fails the Nyblom constancy test. The test statistic 1.79 is over twice the 1 percent critical value. This indicates that the model specification is not adequate.

As a first pass, we try a conditional student's t model, making the logistically

TABLE 6

EXCHANGE RATE: STUDENT t MODEL

Variables	Estimate	St. Error	Robust SE	Nyblom L_k
Mean Equation				
Intercept	0.033	0.0025	0.030	0.34
$e_{t-1}^2 - \sigma_{t-1}^2$	0.15	0.04	0.05	0.23
σ_{t-1}^2	1.01	0.02	0.02	0.40
Degrees of Freedom	8.2	2.8	2.5	1.79
Log L	1142.6			
Nyblom L	2.3			

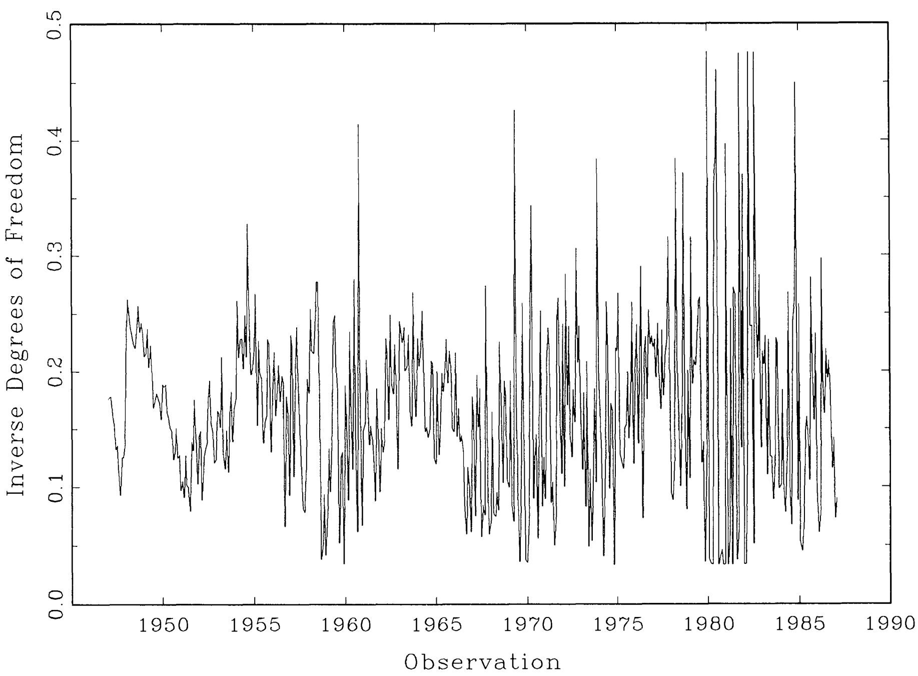

FIGURE 6
INTEREST RATES: INVERSE DEGREE OF FREEDOM

transformed student's t parameter (bounded between 2.1 and 30) a linear function of e_{t-1} and e_{t-1}^2. These results are given in Table 7. The p-value for the increase in the likelihood is 10 percent, which cannot be taken as strong evidence for the augmented model, and the Nyblom stability test statistic still rejects the specification.

As a next pass, Table 8 reports estimates of the skewed student's t model of equation (10), constraining the shape parameters (η and λ) to be constant over time. The estimates for the variance equation and the degrees of freedom are essentially the same as before. The skewness parameter is negative, implying a skew to the left.

TABLE 7
EXCHANGE RATE: CONDITIONAL STUDENT t MODEL

Variables	Estimate	St. Error	Robust SE	Nyblom L_k
Variance Equation				
Intercept	0.031	0.025	0.031	0.29
$e_{t-1}^2 - \sigma_{t-1}^2$	0.17	0.05	0.06	0.20
σ_{t-1}^2	1.01	0.02	0.02	0.35
Degrees of Freedom				
Intercept	−1.07	0.73	0.79	1.59
e_{t-1}	−0.38	0.24	0.19	0.22
e_{t-1}^2	−0.08	0.07	0.06	0.33
Log L	1140.36			
Nyblom L	2.44			

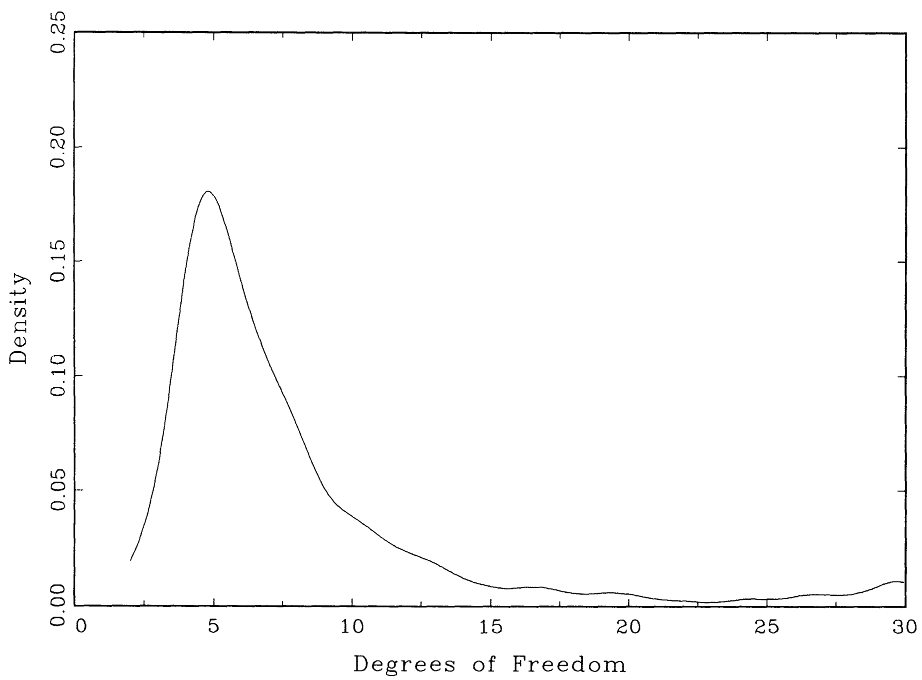

FIGURE 7
INTEREST RATES: DENSITY OF DEGREE OF FREEDOM

Next, Table 9 reports estimates of the general model, allowing for time variation in both λ and η. As before, η_t is bounded between 2.1 and 30. λ_t is bounded between $-.9$ and .9, using the logistic function. Both logistically transformed variables are specified as quadratic functions of e_{t-1}. The results indicate that both the degree of freedom and skewness parameters are negatively related to lagged error e_t and its square. Interestingly, the estimates for the variance equation are slightly different than in the previous specifications, with the estimate for the coefficient on $e_{t-1}^2 - \sigma_{t-1}^2$ rising from 0.15 to 0.20, and that for σ_{t-1}^2 rising from 1.01 to 1.03.

To assess the statistical significance of the general model, it is interesting to

TABLE 8
EXCHANGE RATE: SKEWED STUDENT t MODEL

Variables	Estimate	St. Error	Robust SE	Nyblom L_k
Variance Equation				
Intercept	0.032	0.0024	0.029	0.40
$e_{t-1}^2 - \sigma_{t-1}^2$	0.15	0.04	0.05	0.25
σ_{t-1}^2	1.00	0.02	0.02	0.43
Degrees of Freedom	8.1	2.7	2.5	1.60
Skew Parameter	−0.09	0.05	0.05	1.42
Log L	1141.2			
Nyblom L	3.1			

compare the four likelihoods of Tables 6 through 9. Simply allowing for η_t to be time-varying (Table 7) or the density to be skewed (Table 8) only produces a marginally significant change in the likelihood. But allowing for both effects simultaneously (Table 9) produces a LR test statistic (against the student's t model of Table 6) of 13.5 which has an asymptotic p-value of 2 percent. This again provides strong evidence that parametrically-specified time-varying conditional densities are statistically important as descriptions of the time series properties of financial data.

Figure 8 displays the time series $\hat{\eta}_t$, and Figure 9 displays $\hat{\eta}_t^{-1}$. From the latter it is clear that η_t is primarily hovering around 10, with occasional excursions into the more fat-tailed region. Figure 10 displays an estimate of the density for η_t. Figure 11 displays the estimates λ_t. The sequence is typically near zero, with the density becoming conditionally skewed after large squared innovations. Figure 12 displays a nonparametric estimate of the density of the process λ_t.

Unfortunately, the Nyblom stability test statistics for both the conditional degrees of freedom and skewness equations indicate misspecification. Attempts to rectify this problem by adding extra lags of e_{t-1} to the equations had no effect (the parameter estimates were very small and insignificant). It is also possible that these test statistics are revealing a nonstationary feature of the conditional distribution, which cannot be easily incorporated in an ARCH-type framework. This calls for further research.

5. CONCLUSION

This paper has generalized Engle's ARCH model to let shape parameters beyond the variance depend upon conditioning information. This is achieved simply by using a low-dimensional parametric family for the conditional density, and letting each parameter be a parametric function of the data. Two particular examples of this approach, using a conditional student t distribution and a new conditional skewed student t distribution, are developed and used to model the one-month

TABLE 9
EXCHANGE RATE: CONDITIONAL SKEWED STUDENT t MODEL

Variables	Estimate	St. Error	Robust SE	Nyblom L_k
Variance Equation				
Intercept	0.037	0.029	0.037	0.33
$e_{t-1}^2 - \sigma_{t-1}^2$	0.20	0.06	0.10	0.18
σ_{t-1}^2	1.03	0.03	0.04	0.41
Degrees of Freedom				
Intercept	−1.10	0.73	0.91	1.06
e_{t-1}	−0.54	0.21	0.20	0.23
e_{t-1}^2	−0.08	0.05	0.05	0.34
Skew Parameter				
Intercept	−0.06	0.14	0.14	0.95
e_{t-1}	−0.13	0.09	0.09	0.21
e_{t-1}^2	−0.10	0.05	0.07	0.07
Log L	1135.9			
Nyblom L	3.22			

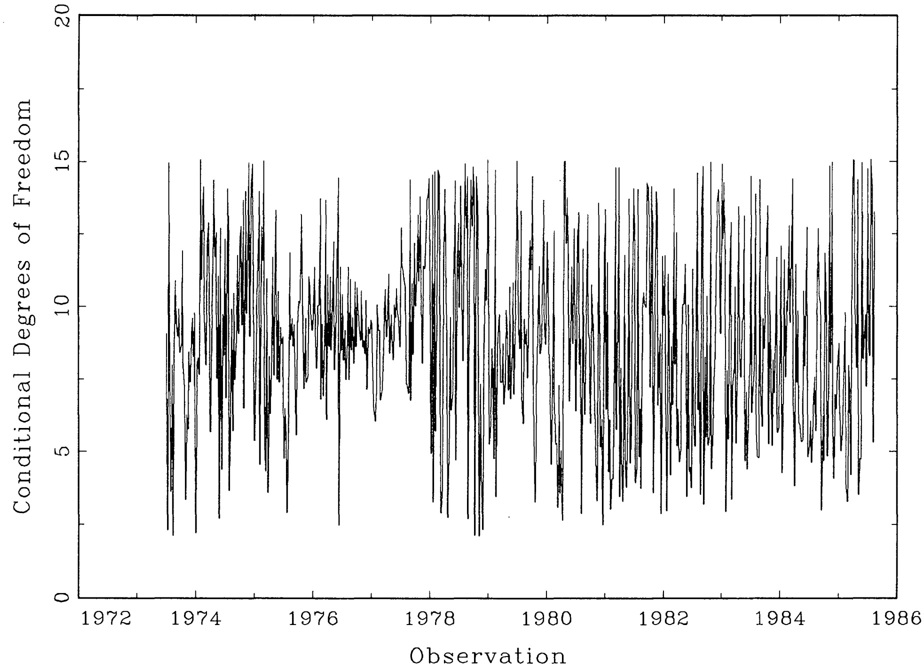

FIGURE 8
EXCHANGE RATES: CONDITIONAL DEGREE OF FREEDOM

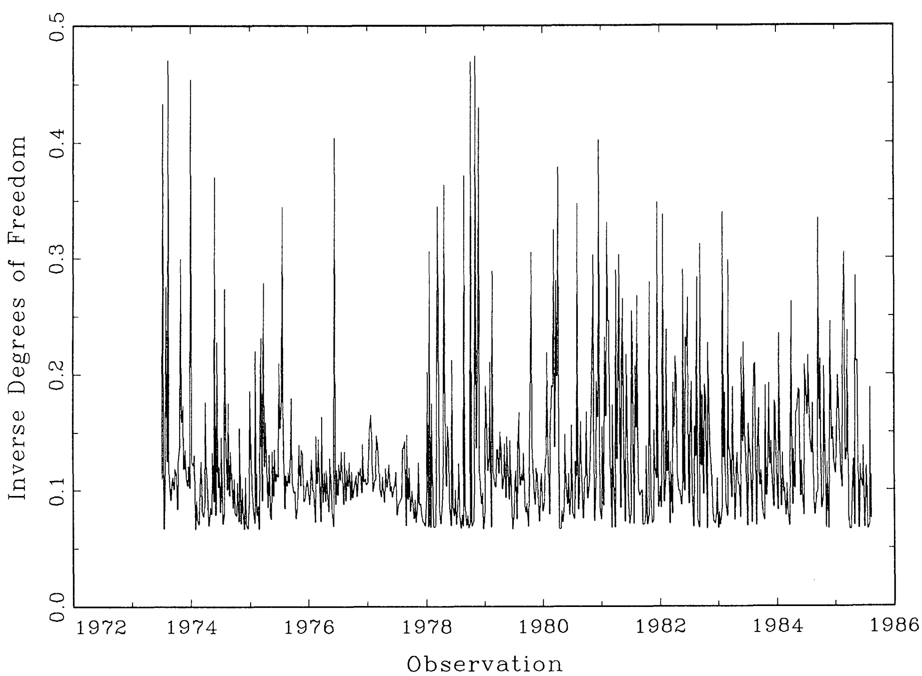

FIGURE 9
EXCHANGE RATES: INVERSE DEGREE OF FREEDOM

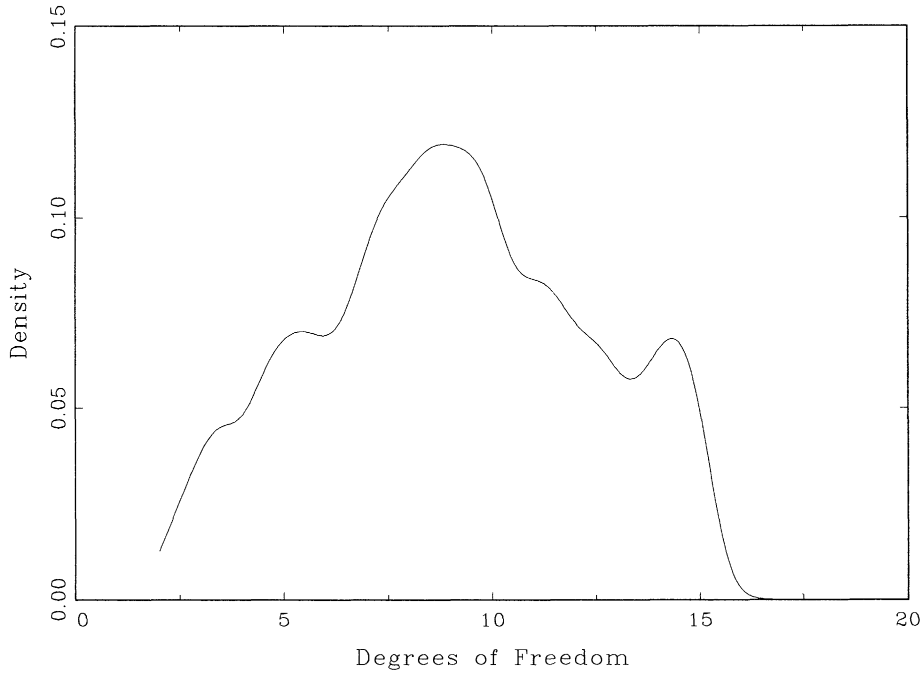

FIGURE 10
EXCHANGE RATES: DENSITY OF DEGREE OF FREEDOM

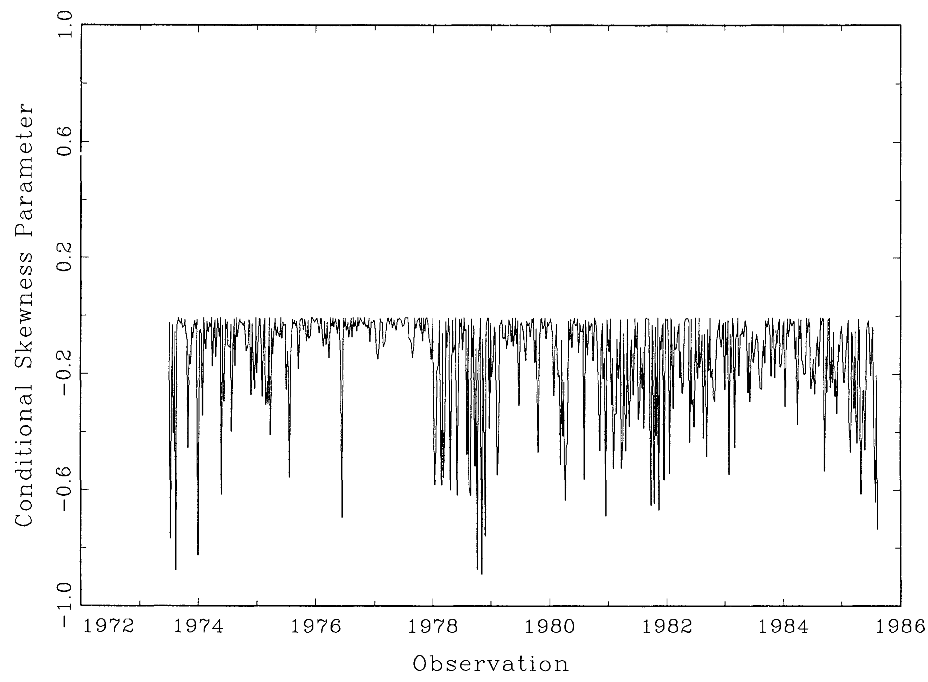

FIGURE 11
EXCHANGE RATES: CONDITIONAL SKEWNESS PARAMETER

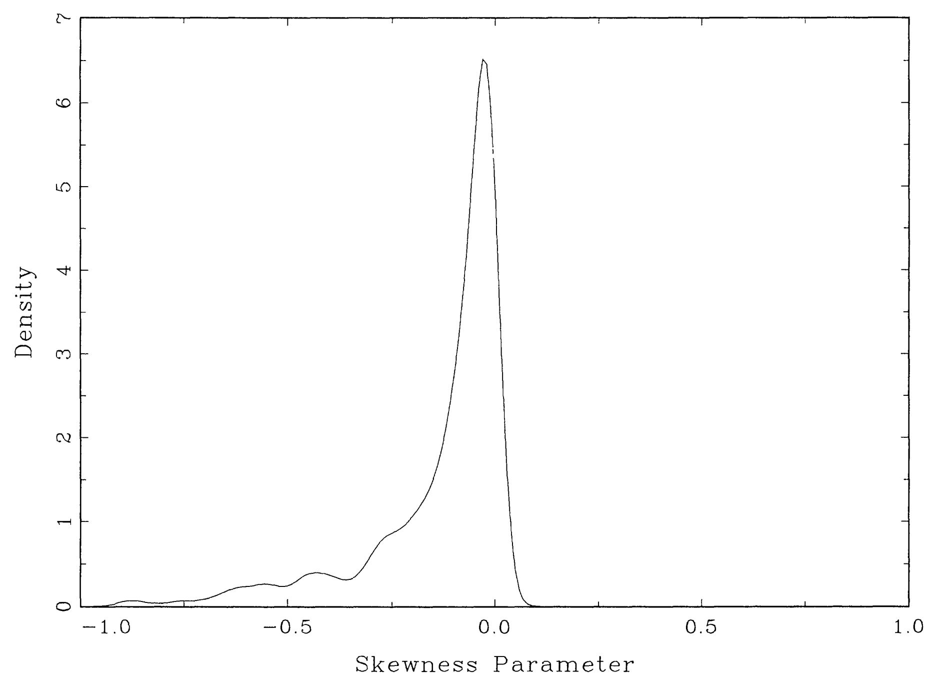

FIGURE 12
EXCHANGE RATES: DENSITY OF SKEWNESS PARAMETER

excess holding yield on U.S. Treasury securities, and monthly Dollar/Franc exchange rate, respectively. The shape parameters of the conditional densities are found to be statistically significant at the 5 percent level.

A GAUSS386 program (requires the OPTMUM module) which replicates the empirical results is available on request from the author.

University of Rochester, U.S.A.

APPENDIX

In this appendix we show that density (10) is a proper density with a mean of zero and unit variance. It will be convenient, however, to first analyze a random variable Z with density

$$h(y|\eta, \lambda) = \begin{cases} bc\left(1 + \dfrac{1}{\eta - 2}\left(\dfrac{y}{1 - \lambda}\right)^2\right)^{-(\eta + 1)/2} & y < 0, \\ bc\left(1 + \dfrac{1}{\eta - 2}\left(\dfrac{y}{1 + \lambda}\right)^2\right)^{-(\eta + 1)/2} & y \geq 0, \end{cases} \tag{16}$$

where the constants b and c are given in (12) and (13). Let $g(x|\eta)$ denote the student's t density normalized to have a unit variance, as in (9), which equals $h(x|\eta, 0)$. By the transformation $x = y/(1-\lambda)$ we see

$$\int_{-\infty}^{0} h(y|\eta, \lambda)\, dy = (1-\lambda)\int_{-\infty}^{0} g(x|\eta)\, dx = \frac{1-\lambda}{2},$$

and by the transformation $x = y/(1+\lambda)$ we find

$$\int_{0}^{\infty} h(y|\eta, \lambda)\, dy = (1+\lambda)\int_{0}^{\infty} g(x|\eta)\, dx = \frac{1+\lambda}{2}.$$

Thus

$$\int_{-\infty}^{\infty} h(y|\eta, \lambda)\, dy = \frac{1-\lambda}{2} + \frac{1+\lambda}{2} = 1$$

and $h(\cdot|\eta, \lambda)$ is a proper density.

Using the same set of transformations we find

$$\int_{-\infty}^{0} yh(y|\eta, \lambda)\, dy = (1-\lambda)^2\int_{-\infty}^{0} c\left(1 + \frac{x^2}{\eta - 2}\right) dx = -c(1-\lambda)^2\left(\frac{\eta - 2}{\eta - 1}\right)$$

and

$$\int_{0}^{\infty} yh(y|\eta, \lambda)\, dy = (1+\lambda)^2\int_{0}^{\infty} c\left(1 + \frac{x^2}{\eta - 2}\right) dx = c(1+\lambda)^2\left(\frac{\eta - 2}{\eta - 1}\right).$$

Thus

$$EY = \int_{-\infty}^{\infty} yh(y|\eta, \lambda)\, dy = c\left(\frac{\eta - 2}{\eta - 1}\right)[(1+\lambda)^2 - (1-\lambda)^2] = 4\lambda c\left(\frac{\eta - 2}{\eta - 1}\right) = a$$

(a is defined in equation (11)).

We also find that

$$\int_{-\infty}^{0} y^2h(y|\eta, \lambda)\, dy = (1-\lambda)^3\int_{-\infty}^{0} x^2g(x|\eta)\, dx = \frac{(1-\lambda)^3}{2}$$

where the final inequality uses the fact that the density $g(x|\eta)$ is symmetric and has a variance of unity. Similarly,

$$\int_{0}^{\infty} y^2h(y|\eta, \lambda)\, dy = \frac{(1+\lambda)^3}{2}.$$

Thus

$$EY^2 = \frac{(1-\lambda)^3}{2} + \frac{(1+\lambda)^3}{2} = 1 + 3\lambda^2 = b^2 + a^2$$

by definitions (11) and (12).

Now consider the random variable given by the transformation

$$Z = \frac{Y-a}{b}.$$

Its density is given by (10), which shows that this is a proper density. We can easily see that

$$EZ = \frac{EY-a}{b} = \frac{a-a}{b} = 0$$

and

$$EZ^2 = \frac{EY^2 - 2aEY + a^2}{b^2} = \frac{a^2 + b^2 - 2a^2 + a^2}{b^2} = 1,$$

which establishes that the density (10) has a mean of zero and unit variance, as desired.

REFERENCES

ANDERSEN, T., "Volatility," Working Paper No. 144, Kellogg Graduate School, Northwestern University, 1992.

BAILLIE, R. T. AND T. BOLLERSLEV, "Prediction in Dynamic Models with Time-Dependent Conditional Variances," *Journal of Econometrics* 52 (1992), 91–114.

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

——— AND T. BOLLERSLEV, "Modeling the Persistence of Conditional Variances," *Econometric Reviews* 5 (1986), 1–50.

——— AND G. GONZALEZ-RIVERA, "Semiparametric ARCH Models," *Journal of Business and Economic Statistics* 9 (1991), 345–360.

———, D. M. LILIEN, AND R. P. ROBINS, "Estimating Time Varying Risk Premia in the Term Structure: The ARCH-M Model," *Econometrica* 55 (1987), 391–407.

GALLANT, R. A., D. A. HSIEH, AND G. E. TAUCHEN, "On Fitting a Recalcitrant Series: The Pound/Dollar Exchange Rate, 1974–1983," in W. Barnett, J. Powell, and G. Tauchen, eds., *Semiparametric and Nonparametric Methods in Econometrics and Statistics* (Cambridge: Cambridge University Press, 1991), 199–240.

GOURIEROUX, C. AND A. MONFORT, "Qualitative Threshold ARCH Models," *Journal of Econometrics* 52 (1992), 159–200.

HANSEN, B. E., "Lagrange Multiplier Tests for Parameter Instability in Non-Linear Models," mimeo, Department of Economics, University of Rochester, 1990.

HENTSCHEL, L., "The Absolute Value GARCH Model and the Volatility of U.S. Stock Returns," mimeo, Simon Graduate School, University of Rochester, 1991.

LEE, S. W., "Asymptotic Properties of the Quasi-Maximum Likelihood Estimator of the GARCH-M Model," mimeo, Department of Economics, Indiana University, 1993.

——— AND B. E. HANSEN, "Asymptotic Properties of the Maximum Likelihood Estimator and Test of the Stability of Parameters of the GARCH and IGARCH Models," mimeo, Department of Economics, University of Rochester, 1992.

——— AND ———, "Asymptotic Theory for the GARCH(1,1) Quasi-Maximum Likelihood Estimator," *Econometric Theory* 10 (forthcoming 1994).

LIU, S. M. AND B. W. BRORSEN, "Maximum Likelihood Estimation of the Stable Distribution with a Time-Varying Scale Parameter," mimeo, Department of Economics, Oklahoma State University, 1992.

LUMSDAINE, R. L., "Asymptotic Properties of the Maximum Likelihood Estimator in GARCH(1,1) and IGARCH(1,1) Models," mimeo, Department of Economics, Princeton University, 1991.

MCCULLOCH, J. H., "Appendix B: U.S. Term Structure Data, 1946–87," in B. M. Friedman and F. H. Hahn, eds., *Handbook of Monetary Economics*, Vol. 1 (New York: North-Holland, 1990), 672–715.

NELSON, D. B., "Conditional Heteroskedasticity in Asset Returns: A New Approach," *Econometrica* 59 (1990), 347–370.

NYBLOM, J., "Testing the Constancy of Parameters over Time," *Journal of the American Statistical Association* 84 (1989), 223–230.

PAGAN, A. R. AND Y. S. HONG, "Non-Parametric Estimation and the Risk Premium," in W. Barnett, J. Powell, and G. Tauchen, eds., *Semiparametric and Nonparametric Methods in Econometrics and Statistics* (Cambridge: Cambridge University Press, 1991), 51–75.

RAMBERG, J. S., P. R. TADIKAMALLA, E. J. DUDEWICZ, AND E. F. MYKYTKA, "A Probability Distribution and Its Uses in Fitting Data," *Technometrics* 21 (1979), 201–209.

SHEPHARD, N. G., "Local Scale Models: State Space Alternatives of Integrated GARCH Processes," *Journal of Econometrics* 60 (1994), 181–202.

SPANOS, A., "Modeling Dynamic Heteroskedasticity Using Thick-Tailed Distributions: The Student's T and Related Linear Models," mimeo, Department of Economics, Virginia Polytechnic Institute, 1991.

WATANABE, T., "Alternative Approach to Conditional Heteroskedasticity in Stock Returns: Approximate Non-Gaussian Filtering," mimeo, Department of Economics, Yale University, 1992.

WHITE, H., "Maximum Likelihood Estimation of Misspecified Models," *Econometrica* 50 (1982), 1–25.