Forthcoming: Journal of Econometrics

Estimation of Stochastic Volatility Models with Diagnostics *

A. Ronald Gallant
University of North Carolina

David Hsieh
Duke University

George Tauchen
Duke University

JEL Classification: C14.
Keywords: stochastic volatility, efficient method of moments, diagnostics.
First Draft: April 1994, Final Revision: December 1995.

Abstract

Efficient Method of Moments (EMM) is used to fit the standard stochastic volatility model and various extensions to several daily financial time series. EMM matches to the score of a model determined by data analysis called the score generator. Discrepancies reveal characteristics of data that stochastic volatility models cannot approximate. The two score generators employed here are "Semiparametric ARCH" and "Nonlinear Nonparametric". With the first, the standard model is rejected, although some extensions are accepted. With the second, all versions are rejected. The extensions required for an adequate fit are so elaborate that nonparametric specifications are probably more convenient.

*Corresponding author: George Tauchen, Duke University, Department of Economics, Social Science Building, Box 90097, Durham NC 27708-0097 USA, phone 1-919-660-1812, FAX 1-919-684-8974, e-mail get@tauchen.econ.duke.edu.

1 Introduction

The stochastic volatility model has been proposed as a description of data from financial markets by Clark (1973), Tauchen and Pitts (1983), Taylor (1986, 1994), and others. The appeal of the model is that it provides a simple specification for speculative price movements that accounts, in qualitative terms, for broad general features of data from financial markets such as leptokurtosis and persistent volatility. Also, it is related to diffusion processes used in derivatives pricing theory in finance; see Mathieu and Schotman (1994) and references therein. The standard form as set forth, for instance, in Harvey, Ruiz, and Shephard (1994), Jacquier, Polson, and Rossi (1994), and Danielsson (1994), takes the form of an autoregression whose innovations are scaled by an unobservable volatility process, usually distributed as a lognormal autoregression.

Estimation of the stochastic volatility model presents intriguing challenges, and a variety of procedures have been proposed for fitting the model. Extant methods include method of moments (Duffie and Singleton, 1993; Andersen and Sorensen, 1996), Bayesian methods (Jacquier, Polson, and Rossi, 1994; Geweke, 1994), simulated likelihood (Danielsson, 1994), and Kalman filtering methods (Harvey, Ruiz, and Shephard, 1994; Kim and Shephard, 1994). Two excellent recent surveys are Ghysels, Harvey, and Renault (1995) and Shephard (1995).

Here, we employ the Efficient Method of Moments (EMM) proposed by Bansal, Gallant, Hussey, and Tauchen (1993, 1995) and developed in Gallant and Tauchen (1996) to estimate and test the stochastic volatility model. EMM is a simulation-based moment matching procedure with certain advantages. The moments that get matched are the scores of an auxiliary model called the "score generator." If the score generator approximates the distribution of the data well, then estimates of the parameters of the stochastic volatility model are as efficient as if maximum likelihood had been employed (Tauchen, 1996; Gallant and Long, 1995). Failure to match these moments can be used as a statistical specification test and, more importantly, can be used to indicate features of data that the stochastic volatility model cannot accommodate (Tauchen, 1995).

The objective is to report and interpret the EMM objective function surface across a comprehensive set of specifications of the stochastic volatility model. We start with the standard,

and widely used setup, with Gaussian errors and short lag lengths, and we proceed to more complicated specifications with long lag lengths. The effort is aimed at generating a comprehensive accounting of how well the model and its extensions accommodate features of the data. An advantage of the EMM procedure is that it is computationally tractable enough to permit this exhaustive specification analysis. Our approach differs from typical practice in the stochastic volatility literature, which is to fit the standard setup and perhaps a single extension in one direction. Since various studies use different specifications, estimation methods, and data sets, it is difficult to reach firm conclusions on the plausibility of the stochastic volatility model. By using EMM, we can confront all of the various extensions, individually and jointly, to a judiciously chosen set of moments determined by a nonparametric specification search for the score generator. Other estimation methods are incapable of investigating the empirical plausibility of such an extended set of specifications for stochastic volatility on the large data sets used here.

We fit the univariate stochastic volatility model to a long time series comprised of 16,127 daily observations on adjusted movements of the Standard and Poor's Composite Price Index, 1928–87. We use such a long series because, among other things, we are interested in the long-term persistence properties of stock volatility.

For this estimation, we use two score generators based on the specification analysis of Gallant, Rossi, and Tauchen (1992). The first is an ARCH model with a homogeneous innovation distribution that is given a nonparametric representation. The specific specification is determined by a standard model selection procedure based on the BIC criterion and specification tests. This model is similar to the most widely used models in the ARCH family. Its score is termed the "Semiparametric ARCH Score". The second score generator is a fully nonparametric estimator of the distribution of a nonlinear process. It both nests the first and relaxes its homogeneity assumption. The specific specification is determined using the same model selection procedure as above. The corresponding score is termed the "Nonlinear Nonparametric Score". These two score generators, determined independently of the stochastic volatility model, are similar to models that are commonly fit to high-frequency financial data.

We undertake a similar exercise for a trivariate stochastic volatility model applied to

4,044 daily observations on adjusted movements of the Standard and Poor's Composite Price Index, adjusted movements of the \$/DM spot exchange rate, and the adjusted 90-day Euro-Dollar interest rate, 1977–92.

2 The Stochastic Volatility Model

2.1 Setup and Notation

Let y_t denote the first difference (either simple or logarithmic) over a short time interval, a day for instance, of the price of a financial asset traded on active speculative markets. The basic stochastic volatility for y_t is

$$\begin{aligned} y_t - \mu_y &= \sum_{j=1}^{L_y} c_j (y_{t-j} - \mu_y) + \exp(w_t) r_y z_t \\ w_t - \mu_w &= \sum_{j=1}^{L_w} a_j (w_{t-j} - \mu_w) + r_w \tilde{z}_t \end{aligned}$$

where μ_y, $\{c_j\}_{j=1}^{L_y}$, r_y, μ_w, $\{a_j\}_{j=1}^{L_w}$, and r_w are the parameters of the two equations, called the mean and volatility equations respectively. The processes $\{z_t\}$ and $\{\tilde{z}_t\}$ are mutually independent *iid* random variables with mean zero and unit variance. Whenever they exist, unconditional expectations are taken with respect to the joint distribution of the processes $\{z_t\}$ and $\{\tilde{z}_t\}$. The first two moments of the z_t and $\tilde{z}_t$ are not separately identified from the other parameters — hence the restriction to $\mathcal{E}(z_t) = \mathcal{E}(\tilde{z}_t) = 0$ and $\mathrm{Var}(z_t) = \mathrm{Var}(\tilde{z}_t) = 1$. Likewise, μ_w is not separately identified; we find numerically the best normalization is simply $\mu_w = 0$. A common assumption in the literature is that both z_t and $\tilde{z}_t$ are independent $N(0,1)$ random variables and that the lag lengths are short. Typically, $L_w = 1$ and $L_y = 1$, or $L_y = 0$. Below, we entertain other distributional assumptions and search over a broad set of lag lengths. The model implies restrictions on the serial covariance properties of $|y_t|^c$, $c > 0$, which are worked out in exhaustive detail in Ghysels, Harvey, and Renault (1995).

One interpretation of the process w_t, which has its origins in Clark (1973) and is refined in Tauchen and Pitts (1983), is that stochastic volatility reflects the random and uneven flow of new information to the financial market. Over the time period $t-1$ to t, a random number of individual pieces of information impinge the market. Each piece triggers an independent

price movement drawn from a time-homogeneous parent distribution. If $I_t = [\exp(w_t)]^2$ individual pieces impinge on the market then, conditional on I_t, the studentized innovation

$$[y_t - \mu_{t-1,t}]/\sqrt{I_t},$$

where

$$\mu_{t-1,t} = \mu_y + \sum_{j=1}^{L_y} c_j(y_{t-j} - \mu_y),$$

would follow a parent distribution, typically Gaussian. The process I_t is called the mixing process. It is unobservable and presumable serially correlated, which motivates the stochastic volatility specification given above.

2.2 Data Generator

The stochastic volatility model defines a strictly stationary and Markov process $\{s_t\}$, where $s_t = (y_t, w_t)'$. The process is Markovian of order $L_s = \max(L_y, L_w)$ with conditional density $p_s(s_t|s_{t-L_s}, \ldots, s_{t-1}, \rho)$ given by the stochastic volatility model, where

$$\rho = (\mu_y, c_1, \ldots, c_{L_y}, r_y, a_1, \ldots, a_{L_w}, r_w)'$$

is a vector that contains the free parameters of the stochastic volatility model.

The process $\{y_t\}$ is observed whereas $\{w_t\}$ is regarded as latent. Write $p_{y,J}(y_{t-J}, \ldots, y_t|\rho)$ for the implied joint density under the model of a stretch $y_{t-J}, \ldots, y_t$. Most integrals appearing in formulas in subsequent sections fail to admit closed form solutions. In practice, they must be approximated by quadrature or Monte Carlo integration, although likelihoods can sometimes be computed efficiently using the Kalman filter (Kim and Shephard, 1994). As will be seen, we need to compute expectations under the model of a variety of nonlinear functions. Monte Carlo integration is most convenient, and is effected by averaging over a long realization from the stochastic volatility model. For a general nonlinear function $g(y_{t-J}, y_{t-J+1}, \ldots, y_t)$, integrals of the form

$$\int \cdots \int g(y_{t-J}, y_{t-J+1}, \ldots, y_t) p_{y,J}(y_{t-J}, y_{t-J+1}, \ldots, y_t|\rho) \prod_{k=0}^{J} dy_{t-k}$$

are approximated by

$$\frac{1}{N} \sum_{\tau=J+1}^{N} g(\hat{y}_{\tau-J}, \hat{y}_{\tau-J+1}, \ldots, \hat{y}_\tau)$$

where $\{\hat{y}_\tau\}_{\tau=1}^N$ is a long simulated realization from the stochastic volatility model given a value ρ. This is accomplished by simulating $\{\hat{s}_\tau\}_{\tau=1}^N$, which is straightforward, and retaining the element $\hat{y}_\tau$ from $\hat{s}_\tau = (\hat{y}_\tau, \hat{w}_\tau)$.

Here, computations are based on realizations of length 50,000 to 100,000, with the choice having no substantive effect on inferences. To let transients die off, first the volatility equation (which displays substantial persistence) runs for 10,000 periods; next, both the mean equation (which displays minor persistence) and the variance equations run together for another 100 periods, which are discarded; then both equations continue to run together to generate a realization of the desired length.

3 The EMM Estimator

In Sections 4 and 5 below we employ the Efficient Method of Moments (EMM) methodology as described in Gallant and Tauchen (1996) to estimate and test the stochastic volatility model. The title of the paper is suggestive — "Which Moments to Match?" — and the answer is simple and intuitive: Use the score vector of an auxiliary model that fits the data well to define a GMM criterion function. The EMM method has some computational advantages relative to indirect inference (Gourieroux, Monfort, and Renault, 1993) as it circumvents the need to refit the score generator to each simulated realization (compute the binding function) and it bypasses a Hessian computation. The ideas behind EMM are as follows.

We observe the data $\{\tilde{y}_t\}_{t=1}^n$, which is presumed to have been generated by the stochastic volatility model for some value $\rho^o \in R \subset \Re^{\ell_\rho}$, where ℓ_ρ is the length of ρ^o. The task is to estimate ρ^o and test the specification of the model.

Suppose that a probability model for the stochastic process $\{y_t\}_{t=-\infty}^{\infty}$ defined by the conditional density

$$f(y_t|y_{t-L}, y_{t-L+1}, \cdots, y_{t-1}, \theta) \qquad \theta \in \Theta \subset \Re^{\ell_\theta},$$

fits the data $\{\tilde{y}_t\}_{t=1}^n$ reasonably well. Fits well means that when its parameters are estimated

by quasi-maximum likelihood

$$\tilde{\theta}_n = \underset{\theta \in \Theta}{\operatorname{argmax}} \sum_{t=L+1}^{n} \log[f(\tilde{y}_t | \tilde{y}_{t-L}, \ldots, \tilde{y}_{t-1}, \theta)]$$

the model does reasonably well on statistical specification tests and the fit appears sensible from an economic perspective. The functional form of $f(y_t|x_{t-1}, \theta)$ need not have any direct connection to that of the true conditional distribution of y_t given $x_{t-1} = (y_{t-L}, y_{t-L+1}, \ldots, y_{t-1})$, which is

$$\frac{p_{y,L}(y_{t-L}, y_{t-L+1}, \ldots, y_t | \rho^o)}{p_{y,L-1}(y_{t-L}, y_{t-L+1}, \ldots, y_{t-1} | \rho^o)}.$$

It should provide a good approximation, though, for the EMM estimator to be nearly fully efficient (Tauchen, 1996; Gallant and Long, 1995).

The EMM estimator brings the information in $f(y|x, \tilde{\theta}_n)$ to bear on the task of estimating and testing the stochastic volatility model as follows. Define the criterion

$$m(\rho, \theta) = \int \cdots \int \frac{\partial}{\partial \theta} \log[f(y_t | y_{t-L}, \ldots, y_{t-1}, \theta)] p_{y,L}(y_{t-L}, \ldots, y_t | \rho) \prod_{k=0}^{L} dy_{t-k},$$

which is the expected score of the $f(y|x, \theta)$ model under the stochastic volatility model. Hence, $f(y|x, \theta)$ is called the "score generator". The induced parameter that $\tilde{\theta}_n$ estimates is that value θ^o for which $m(\rho^o, \theta) = 0$ (Gallant, 1987, Chapter 7, Theorem 8). This fact provides the motivation for the EMM estimator. One expects $m(\rho, \tilde{\theta}_n)$ to be near zero for values of ρ close to ρ^o.

The EMM estimator is

$$\hat{\rho}_n = \underset{\rho \in R}{\operatorname{argmin}} \; m'(\rho, \tilde{\theta}_n)(\tilde{\mathcal{I}}_n)^{-1} m(\rho, \tilde{\theta}_n)$$

where

$$\tilde{\mathcal{I}}_n = \frac{1}{n} \sum_{t=L+1}^{n} \Big[\frac{\partial}{\partial \theta} \log f(\tilde{y}_t | \tilde{x}_{t-1}, \tilde{\theta}_n)\Big] \Big[\frac{\partial}{\partial \theta} \log f(\tilde{y}_t | \tilde{x}_{t-t}, \tilde{\theta}_n)\Big]'.$$

and

$$\tilde{x}_{t-1} = (\tilde{y}_{t-L}, \tilde{y}_{t-L+1}, \ldots, \tilde{y}_{t-1})'.$$

In computing $\hat{\rho}_n$, we do not need to impose restrictions that the parameter space R contains only those ρ for which the model generates stationary data, as such restrictions are

automatically enforced on the computation (Tauchen, 1995). Also, as noted in Gallant and Tauchen (1996), one should, strictly speaking, use a weighted covariance estimator of

$$\mathcal{I}^o = \text{Var}\left[\frac{1}{\sqrt{n}}\sum_{t=L+1}^{n}\frac{\partial}{\partial\theta}\log f(y_t|y_{t-L}, y_{t-L+1}, \ldots, y_{t-1}, \theta^o)\right]$$

rather than $\tilde{\mathcal{I}}_n$, and formulas are given therein. However, it is unlikely that this generality will be necessary in practice because the use of a weighted covariance estimator means that one thinks that the score generator is a poor statistical approximation to the data generating process. A poor statistical approximation is unlikely because the score generator is, conceptually, a reduced form model, not a structural model, and is usually easy to modify by adding a few parameters so that it fits the data well.

Under regularity conditions stated in Gallant and Tauchen (1996), which are standard regularity conditions such that the maximum likelihood estimator of ρ in $p(y|x,\rho)$ is consistent and asymptotic normal and such that the quasi maximum likelihood estimator of θ in $f(y|x,\theta)$ is asymptotic normal, we have that $\hat{\rho}_n$ is consistent and

$$\sqrt{n}(\hat{\rho}_n - \rho^o) \xrightarrow{\mathcal{L}} N\left\{0, [(M^o)'(\mathcal{I}^o)^{-1}(M^o)]^{-1}\right\}$$

where $M^o = M(\rho^o, \theta^o)$ and $M(\rho,\theta) = (\partial/\partial\rho')m(\rho,\theta)$. M^o can be estimated consistently by $\hat{M}_n = M_n(\hat{\rho}_n, \tilde{\theta}_n)$. The order condition (necessary condition) for identification is $\ell_\rho \leq \ell_\theta$; sufficient conditions are discussed in Gallant and Tauchen (1996). The better the score generator approximates the conditional distribution of the data, then the closer is the asymptotic covariance matrix to that of maximum likelihood (Tauchen, 1996; Gallant and Long, 1995). If the score generator actually nests the true conditional distribution, then full efficiency obtains (Gallant and Tauchen, 1996).

$M_n(\rho,\theta)$ must be computed numerically in order to use the asymptotic distribution to get standard errors for setting confidence intervals on the elements of ρ^o. Alternatively, one can avoid computation of $\hat{M}_n$ by using the criterion difference statistic to set confidence intervals (Gallant, 1987, Chapter 7, Theorem 15). The latter approach is to be preferred in most time series applications because it will exclude values of ρ that imply an explosive process from the confidence interval (Tauchen, 1995).

For specification testing, which is the focus of this paper, we have that

$$n\, m'(\hat{\rho}_n, \tilde{\theta}_n)(\tilde{\mathcal{I}}_n)^{-1} m(\hat{\rho}_n, \tilde{\theta}_n) \xrightarrow{\mathcal{L}} \chi^2(df)$$

with $df = \ell_\theta - \ell_\rho$ under the null hypothesis that the maintained model $p_{y,L}(y_{t-L}, \ldots, y_t, \rho)$ is correct.

When a model fails a diagnostic test, one would like some suggestions as to what is wrong. Inspection of the quasi-t-ratios

$$\hat{T}_n = S_n^{-1} \sqrt{n}\, m(\hat{\rho}_n, \tilde{\theta}_n)$$

where $S_n = [\mathrm{diag}(\tilde{\mathcal{I}}_n)]^{1/2}$ can suggest reasons for model failure. As seen in Section 4, different elements of the score vector correspond to different features of the fit. Large quasi-t-ratios reveal the features of the data that the maintained model cannot approximate.

The elements of $\hat{T}_n$ are biased downward in absolute value because the standard errors S_n are too large due to the fact that

$$\sqrt{n}\, m(\hat{\rho}_n, \tilde{\theta}_n) \xrightarrow{\mathcal{L}} N\Big\{0, \mathcal{I}^o - (M^o)[(M^o)'(\mathcal{I}^o)^{-1}(M^o)]^{-1}(M^o)'\Big\}.$$

The downward bias can be corrected by computing $\hat{M}_n$ numerically and putting $S_n = \Big(\mathrm{diag}\{\tilde{\mathcal{I}}_n - (\hat{M}_n)[(\hat{M}_n)'(\tilde{\mathcal{I}}_n)^{-1}(\hat{M}_n)]^{-1}(\hat{M}_n)'\}\Big)^{1/2}$ in the formula for $\hat{T}_n$.

We have not corrected the bias in this paper because we believe the correction to be unnecessary for two reasons. First, $\mathcal{I}^o - (M^o)[(M^o)'(\mathcal{I}^o)^{-1}(M^o)]^{-1}(M^o)'$ is the familiar formula for the variance of GLS residuals and experience with GLS regressions suggests that the difference between $\mathcal{I}^o - (M^o)[(M^o)'(\mathcal{I}^o)^{-1}(M^o)]^{-1}(M^o)'$ and $\mathcal{I}^o$ is negligible in most applications. Secondly, we do not rely on the quasi-t-ratios for inference, we only rely on them for suggestions as to how the stochastic volatility model might be enhanced. When we act upon a suggestion, we check it with the χ^2 statistic. This methodological approach is similar to the well established F-protected t-test methodology as employed in the statistical Analysis of Variance.

4 Univariate Empirical Results

4.1 Data

The data to which we fit the univariate stochastic volatility model is a long time series comprised of 16,127 daily observations, $\{\tilde{y}_t\}_{t=1}^{16,127}$, on adjusted movements of the Standard and Poor's Composite Price Index, 1928–87. This series is the univariate stock series used in Gallant, Rossi, and Tauchen (1992, 1993). The raw series is the Standard and Poor's Composite Price Index (SP), daily, 1928–87. We use a long time series, because, among other things, we want to investigate the long-term properties of stock market volatility. As described in Gallant, Rossi, and Tauchen (1992), the raw series is converted to a price movements series, $100[\log(SP_t) - \log(SP_{t-1})]$, and then adjusted for systematic calendar effects in location and scale. Financial data are known to exhibit calendar effects, that is, systematic shifts in location and scale due to different trading patterns across days of the week, holidays, and year-end tax trading. Calendar effects comprise a very small portion of the total variation in the series, although they should still be accounted for in order not to adversely affect subsequent analysis. The raw and adjusted data are plotted in Figure 1. Though long time series sometimes exhibit structural regime switches, there is no such shift apparent in the figure.

4.2 Score Generators

To implement the EMM estimator we require a score generator $f(y|x,\theta)$ that fits these data well. As documented in in Gallant, Rossi, and Tauchen (1992, 1993) the seminonparametric (SNP) density proposed by Gallant and Tauchen (1989) does so. Moreover, when refitted to subperiods, estimates are stable.

The SNP density is a member of a class of parameterized conditional densities

$$\mathcal{H}_K = \left\{ f_K(y|x,\theta) : \theta = (\theta_1, \theta_2, \cdots, \theta_{\ell_K}) \right\}$$

which expands $\mathcal{H}_1 \subset \mathcal{H}_2 \subset \cdots$ as K increases. It has two desirable properties from the perspective of EMM estimation: (1) The union $\mathcal{H} = \bigcup_{K=1}^{\infty} \mathcal{H}_K$ is quite rich and it is reasonable to assume that the true density $p(y|x)$ of stationary data from a financial market is

contained in $\mathcal{H}$. (2) If θ is estimated by quasi-maximum likelihood, viz.

$$\tilde{\theta}_n = \arg\max_{\theta \in \Re^{\ell_K}} \frac{1}{n} \sum_{t=L+1}^{n} \log[f_K(\tilde{y}_t|\tilde{y}_{t-L}, \ldots, \tilde{y}_{t-1}, \theta)],$$

and if K grows with sample size n [either adaptively as a random variable $\tilde{K}_n$ or deterministically as a function $K(n)$], then

$$\tilde{p}_n(y|x) = f_K(y|x, \tilde{\theta}_n)$$

is a consistent (Gallant and Nychka, 1987) and efficient (Fenton and Gallant, 1996a; Gallant and Long, 1995) nonparametric estimator of $p(y|x)$ with desirable qualitative features (Fenton and Gallant, 1996b).

A standard method of describing a conditional density $f(y|x, \theta)$ is to set forth a location function μ_x and a scale function R_x that reduces the process $\{y_t\}_{t=-\infty}^{\infty}$ to an innovation process $\{z_t\}_{t=-\infty}^{\infty}$ via the transformation

$$z_t = r_{x_{t-1}}^{-1}(y_t - \mu_{x_{t-1}}).$$

The description is completed by setting forth a conditional density $h(z|x)$ for the innovation process. We follow this recipe in describing $f_K(y|x, \theta) \in \mathcal{H}_K$.

The location function μ_x is affine in x

$$\mu_{x_{t-1}} = b_0 + b' x_{t-1}.$$

It is presumed to depend on $L_\mu \leq L$ lags which is accomplished by putting leading elements of b to zero as required. Note that were one to put r_x to a constant and eliminate the dependence of the innovation density on x by writing $h(z)$ instead of $h(z|x)$ then $\{y_t\}_{t=-\infty}^{\infty}$ would be a vector autoregression (VAR).

The scale function r_x is affine in the absolute values of x

$$r_{x_{t-1}} = \rho_0 + \rho' |x_{t-1}|$$

It is presumed to depend on $L_R \leq L$ lags which is accomplished by putting leading elements of ρ to zero as required. Note that were one to eliminate the dependence of the innovation density on x by writing $h(z)$ instead of $h(z|x)$ then $\{y_t\}_{t=-\infty}^{\infty}$ would be an ARCH-type process akin to that proposed by Nelson (1991).

For a vector $\zeta = (\zeta_1, \ldots, \zeta_\ell)$ with real elements and a vector $\lambda = (\lambda_1, \ldots, \lambda_\ell)$ with integer elements, let ζ^λ denote the monomial $\prod_{i=1}^{\ell} \zeta_i^{\lambda_i}$ of degree $|\lambda| = \sum_{i=1}^{\ell} |\lambda_i|$ and consider

$$h_K(z|x) = \frac{[P_K(z,x)]^2 \phi(z)}{\int [P_K(u,x)]^2 \phi(u)\, du}$$

formed from the polynomial

$$P_K(z,x) = \sum_{\alpha=0}^{K_z} \Big(\sum_{|\beta|=0}^{K_x} a_{\alpha\beta} x^\beta \Big) z^\alpha$$

where $\phi(z) = (2\pi)^{-1/2} e^{-z'z/2}$. $P_K(z,x)$ is a polynomial of degree K_z in z whose coefficients are, in turn, polynomials of degree K_x in x. The product $[P_K(z,x)]^2 \phi(z)$ is a Hermite polynomial in z with positivity enforced whose coefficients depend on x. The shape of the innovation density $h_K(z_t|x_{t-1})$ varies with x_{t-1} which permits $h_K(z_t|x_{t-1})$ to exhibit general, conditional shape heterogeneity. By putting selected elements of the matrix $A = [a_{\alpha\beta}]$ to zero, $P_K(z,x)$ can be made to depend on only $L_p \leq L$ lags from x. One may note that if K_z is put to zero, then the innovation density $h_K(z|x)$ is Gaussian. If $K_z > 0$ and $K_x = 0$, then the density can assume arbitrary shape but innovations are homogeneous.

The change of variables $y_t = r_{x_{t-1}} z_t + \mu_{x_{t-1}}$ to obtain the density

$$f_K(y_t|x_{t-1}, \theta) = \frac{\{P_K[r_{x_{t-1}}^{-1}(y_t - \mu_{x_{t-1}}), x_{t-1}]\}^2 \phi[r_{x_{t-1}}^{-1}(y_t - \mu_{x_{t-1}})]}{|r_{x_{t-1}}|^{1/2} \int [P_K(u, x_{t-1})]^2 \phi(u)\, du}.$$

completes the description of the SNP density. The vector θ contains the coefficients $A = [a_{\alpha\beta}]$ of the Hermite polynomial, the coefficients $[b_0, b]$ of the location function, and the coefficients $[\rho_0, \rho]$ of the scale function. To achieve identification, the coefficient $a_{0,0}$ is set to 1. The tuning parameters are L_u, L_r, L_p, K_z, and K_x, which determine the dimension ℓ_K $(= \ell_\theta)$ of θ.

When data is heavy tailed, as is typical for data from financial markets, numerical stability can be enhanced without affecting theoretical results by forming the vector of lags x_{t-1} from a series $\{y_t^*\}$ consisting of $\{y_t\}$ that have been centered by subtracting the sample mean, scaled by dividing by the sample standard error, and transformed by the logistic map that takes the interval $(-\infty, \infty)$ into the interval (-4, 4). That has been done both here and in the results reported for this series by Gallant, Rossi, and Tauchen (1992, 1993). Note that it is only the lagged dependent variables x_{t-1} that are logistic transformed; the contemporaneous y_t is not.

We selected the tuning parameters L_u, L_r, L_p, K_z, and K_x following the protocol that is described in detail in Bansal, Gallant, Hussey, and Tauchen (1995). Briefly, the model is expanded sequentially according to the BIC (Schwarz, 1978) model selection criterion. It is then expanded further if a battery of statistical specification tests indicate that the BIC specification is inadequate. Following this protocol, we selected the model $L_u = 2$, $L_r = 18$, and $K_z = 4$ with 26 free parameters, when innovations are constrained to be homogenous (that is, $K_x = 0$, and $L_p = 1$ imposed). This is a semiparametric density with a parametric part comprised of an AR(2)-ARCH(18) model with unconstrained lag coefficients and a nonparametric error density, which is analogous to the model proposed by Engle and Gonzales-Rivera (1991). We term the score from this fit the "Semiparametric ARCH Score" in legends for figures and tables. When the homogeneity constraint is dropped, and we follow the same protocol, we select the model $L_u = 2$, $L_r = 18$, $L_p = 2$, $K_z = 4$, and $K_x = 1$ with 36 free parameters; this specification does better under BIC than the model with homogeneous errors. This fitted model differs in only minor respects from the preferred SNP specification reported in Gallant, Rossi, and Tauchen (1992). (The differences are due to minor enhancements to the computer program.) We term the score from this fit the "Nonlinear Nonparametric Score".

We emerge from this exercise with two sets of scores with which to confront the stochastic volatility model. The first, the Semiparametric ARCH Score, is defined by a score generator that is very similar to models widely employed in the ARCH literature, though a bit more flexibly parameterized. The second, the Nonlinear Nonparametric Score, is defined by a score generator determined via a complete specification search that accounts for the full complexity of the data.

4.3 Fit to the Semiparametric ARCH Score

Table 1 shows the optimized values of the EMM objective function scaled to follow a chi square, as described in Section 3. Table 2 shows the parameter estimates for the various specifications reported in Table 1. From the top panel of Table 1, labeled Gaussian, it is seen that the standard stochastic volatility model fails to approximate the distribution of the data adequately; it is overwhelmingly rejected. However, as seen from the objective function

surface laid out across the various panels of the table, certain extensions of the standard stochastic volatility model fit the data better.

We describe these extensions and seek to determine which features of the data they seem to approximate well and which features poorly. Guided by the objective function, we inspect the EMM quasi-t-ratios $\hat{T}_n$. The elements of $\hat{T}_n$ provide suggestive diagnostics, as pointed out in Section 3.

Figure 2 shows these EMM quasi-t-ratios as a bar chart for the case $L_y = 2$, $L_w = 2$, and Gaussian z's. This is the standard stochastic volatility specification

$$\begin{aligned} y_t - \mu_y &= c_1(y_{t-1} - \mu_y) + c_2(y_{t-2} - \mu_y) + \exp(w_t) r_y z_t \\ w_t - \mu_w &= a_1(w_{t-1} - \mu_w) + a_2(w_{t-2} - \mu_w) + r_w \tilde{z}_t. \end{aligned}$$

The source of the rejection of this model is failure to match the features defined by the polynomial part of the SNP score. Either $\exp(w_t)$ is not the correct transformation of the latent variance process or z_t is not Gaussian.

Modified Exponential

To explore the first possibility, consider the model

$$\begin{aligned} y_t - \mu_y &= c_1(y_{t-1} - \mu_y) + c_2(y_{t-2} - \mu_y) + T_e(w_t) r_y z_t \\ T_e(w_t) &= \exp(b_{e0} + b_{e1} w_t) + b_{e2} w_t^2 + b_{e3} I_+(w_t) w_t^2 \\ w_t - \mu_w &= a_1(w_{t-1} - \mu_w) + a_2(w_{t-2} - \mu_w) + r_w \tilde{z}_t \end{aligned}$$

where $I_+(w)$ is 1 if w is positive and is 0 otherwise. The idea is to modify the Taylor expansion of $\exp(\cdot)$ by replacing the quadratic term with a differentiable quadratic spline that has one knot at zero. Inspection of the bar chart (not shown) indicates failure. The fit is improved by better matching the VAR and ARCH scores at the expense of further mismatch to the polynomial part of the SNP score. The exponential transformation appears not to be a problem, so we consider non-Gaussian densities for z_t.

t-Errors

A natural way to relax the Gaussian assumption is to use t-errors. Consider the model

$$\begin{aligned} y_t - \mu_y &= c_1(y_{t-1} - \mu_y) + c_2(y_{t-2} - \mu_y) + \exp(w_t) r_y \tau_{\nu t} \\ w_t - \mu_w &= a_1(w_{t-1} - \mu_w) + a_2(w_{t-2} - \mu_w) + r_w \tilde{z}_t \end{aligned}$$

where $\{\tau_{\nu t}\}$ is *iid* Student-t with ν degrees of freedom. The objective function is so flat for values of the degrees of freedom parameter $\nu \in (10, 20)$ that the optimizer gets stuck and makes no progress when it sees ν as free parameter along with the rest. Thus, in the second panel of Table 1 we report the value of the objective function for $\nu = 10, 15, 20, 25$. The specification with t errors helps, but still the model does not fit the data. Figure 3 shows the bar chart for the case $\nu = 15$; the stochastic volatility model fails to fit the score of the SNP polynomial for the cubic term, suggesting a failure to generate skewness.

Spline Error Transformation

More flexibility than with the t is available from a spline transformation to the Gaussian innovation. Consider

$$\begin{aligned} y_t - \mu_y &= c_1(y_{t-1} - \mu_y) + c_2(y_{t-2} - \mu_y) + \exp(w_t) r_y T_z(z_t) \\ T_z(z_t) &= b_{z0} + b_{z1} z_t + b_{z2} z_t^2 + b_{z3} I_+(z_t) z_t^2 \\ w_t - \mu_w &= a_1(w_{t-1} - \mu_w) + a_2(w_{t-2} - \mu_w) + r_w \tilde{z}_t \end{aligned}$$

The idea is to allow a deviation from the Gaussian specification by transforming z_t through a differentiable quadratic spline that has one knot at zero. To achieve identification, the constraints $(2\pi)^{-1/2} \int T_z(v) \exp(-v^2/2) dv = 0$ and $(2\pi)^{-1/2} \int T_z^2(v) \exp(-v^2/2) dv = 1$ are imposed on the b_{zj}. From Table 1 it is seen that the added flexibility of the spline transform sharply reduces the objective function value. The EMM quasi-t-ratios for this "spline-transform" fit are shown in Figure 4. The transform works; the moments of the polynomial part of the Semiparametric ARCH score are adequately matched.

The effects of the spline are to fatten the tails and introduce an asymmetry as seen in Figure 5. The solid line in the upper left panel is a plot of the spline T_z. This plot can also

be interpreted as a plot of the quantiles of the distribution of the random variable $T_z(z_t)$ on the vertical axis against the quantiles of the standard normal distribution on the horizontal axis. If a distribution is Gaussian, then its quantile-quantile plot is a 45-degree line. A comparison with the 45-degree line in the upper left panel of Figure 5 indicates heavy tails, because the solid line plots below the 45-degree line on the left and above on the right, and an asymmetry, because the solid line deviates more from the 45-degree line on the left than on the right. The asymmetry is also apparent from a comparison with the solid line in the upper right panel of Figure 5 which shows a quantile-quantile plot of the six degrees of freedom Student t-distribution. The asymmetry and heavy tails are features of the data that have been captured by the Semiparametric ARCH Score as can be seen in the lower left panel of Figure 5. The EMM moment matching procedure has transferred these characteristics to the spline-transform stochastic volatility model. The asymmetry and heavy tails are real features of the data, not artifacts of the SNP fit, as can be seen from the solid line in the lower right panel of Figure 5 which is a quantile-quantile plot of a kernel density estimate from ARCH residuals.

Chaotic Volatility

Interestingly, one can do as well with a deterministic variance process. EMM quasi-t-ratios (not shown) that result when the variance equation of the model

$$\begin{aligned} y_t - \mu_y &= c_1(y_{t-1} - \mu_y) + c_2(y_{t-2} - \mu_y) + T_w(w_t) r_y T_z(z_t) \\ T_w(w_t) &= b_{w0} + b_{w1} w_t + b_{w2} w_t^2 + b_{w3} I_+(w_t) w_t^2 \\ T_z(z_t) &= b_{z0} + b_{z1} z_t + b_{z2} z_t^2 + b_{z3} I_+(z_t) z_t^2 \end{aligned}$$

is a moving average in 40 lags

$$w_t = \sum_{j=0}^{40} \frac{40 - j}{40} v_{t-j}$$

from a chaotic Mackey-Glass sequence

$$v_t = v_{t-1} + 10.5 \left(\frac{0.2 v_{t-5}}{1 + v_{t-5}^{10}} - 0.1 v_{t-1} \right)$$

are similar to those shown in Figure 4. This Mackey-Glass variant on the spline-transform stochastic volatility model does slightly better on the SNP scores and slightly worse on the

ARCH scores.

Long Memory

Figure 4 suggests the standard stochastic volatility model has some trouble matching the scores of the flexibly parameterized ARCH model, and somewhat more so at the longer ARCH lags. Bollerslev and Mikkelsen (1996), Ding, Granger, and Engle (1993), and Breidt, Crato, and Lima (1994) present evidence that long-memory models like those of Granger and Joyeux (1980) might be needed to account for the high degree of persistence in financial volatility. Harvey (1993) contains an extensive discussion of the properties of long memory in stochastic volatility models. We thus explore if inclusion of both short- and long-memory helps in fitting the stochastic volatility model.

The long-memory stochastic volatility model is

$$\begin{aligned} y_t - \mu_y &= c_1(y_{t-1} - \mu_y) + c_2(y_{t-2} - \mu_y) + \exp(w_t^*) r_y z_t \\ w_t^* - \mu_w &= (1 - \mathcal{L})^{-d} z_{wt} \\ z_{wt} &= \sum_{j=1}^{L_w} a_j z_{w,t-j} + r_w \tilde{z}_t \end{aligned}$$

where $\{z_t\}$ and $\{\tilde{z}_t\}$ are *iid* Gaussian, $(1 - \mathcal{L})^{-d} = \sum_{k=0}^{\infty} \psi_k(d) \mathcal{L}^k$, and the coefficients $\psi_k(d)$ are obtained from the series expansion of $f(x) = (1 - x)^{-d}$, valid for $|d| < 1$, as described in Sowell (1990). Motivating this specification is the fact that for $|d| < 1/2$, $(1 - \mathcal{L})^d v_t = \epsilon_t$, $\{\epsilon_t\}$ *iid* with finite variance, defines a strictly stationary process whose moving average representation is $v_t = (1 - \mathcal{L})^{-d} \epsilon_t = \sum_{k=1}^{\infty} \psi_k(d) \epsilon_{t-k}$; the autocovariance function of v_t decays arithmetically to zero, instead of exponentially to zero as in the case of an autoregression of finite lag length. For $1/2 \leq d < 1$, $(1 - \mathcal{L})^d v_t = \epsilon_t$, defines a nonstationary process. $\{w_t^*\}$ is thus obtained by passing the autoregressive process $\{z_{wt}\}$ through the long-memory moving average filter. For $d = 0$, this generates exactly the same autoregressive volatility process as earlier, while for $0 < |d| < 1/2$, it defines a strictly stationary volatility process with both short- and long-memory components.

Since we need very long realizations for Monte Carlo integration, it is impractical to simulate exactly from this model by, say, computing the Cholesky factorization of the covariance matrix of w_t and proceeding in the usual manner. Instead, we follow Bollerslev

and Mikkelsen (1996) and use a method that truncates the moving average filter and lets the process run for a long while to attenuate the effects of transients. Their calculations suggest that truncation at 1,000 suffices, so we use the moving average filter $\sum_{k=0}^{1000} \psi_k(d)\mathcal{L}^k$. (Because of the truncation, this method technically generates a stationary process for all $|d| < 1$.) They trim off the first 7,000 realizations; we trim off the first 10,000. Some would argue that this method does not actually generate realizations from a long-memory volatility process. The point is well taken but, nonetheless, the Bollerslev-Mikkelsen approach still defines a volatility process $\{w_t^*\}$ with extremely high persistence.

The bottom part of Table 1 shows the optimized objective function when the long-memory parameter, d, is estimated jointly with the other parameters of the model subject to a normalization on μ_w for identification. We only estimate the long-memory version for $L_w = 1$ and $L_w = 2$, since the job of the long-memory specification is to take care of longer lags. For the block labeled "Gaussian & Long Memory" the mean equation is

$$y_t - \mu_y = c_1(y_{t-1} - \mu_y) + c_2(y_{t-2} - \mu_y) + \exp(w_t^*) r_y z_t$$

while for the block labeled "Spline & Long Memory," the mean equation is

$$y_t - \mu_y = c_1(y_{t-1} - \mu_y) + c_2(y_{t-2} - \mu_y) + \exp(w_t^*) r_y T_z(z_t)$$

where the two-parameter quadratic spline $T_z(\cdot)$ is as defined above.

As seen from Table 1, long memory helps, but the Gaussian stochastic volatility model cannot accommodate all of the structure implicit in the semiparametric ARCH model. With the spline transform, it can. Figure 6 shows the bar chart for the case $L_w = 2$. The impact on the objective function value of long memory is similar to that of introducing two or three extra freely parameterized lags into the volatility equation. Overall, long-memory helps about as much as introducing six free lags into the volatility specification.

4.4 Fit to the Nonlinear Semiparametric Score

Table 3 displays the objective function surface for versions of the stochastic volatility model against the Nonlinear Nonparametric Score; Table 4 shows the estimated parameter values. From Table 3, the standard model is overwhelmingly rejected. The various extensions provide

much improvement over the standard Gaussian model, but nothing comes as close as the spline variants against the Semiparametric ARCH Score. We now examine the performance of the extensions in more detail.

The bar chart for the $L_y = 2$, $L_w = 2$, Gaussian stochastic volatility specification is shown as Figure 7. The ARCH part of the score is fit poorly, as is the SNP part. The quasi-t-ratios are not orthogonal, so that failure to fit the SNP scores could manifest itself as large ARCH quasi-t-ratios and conversely. The spline-transform variant (not shown) does just about as poorly.

The full Nonlinear Nonparametric Score embodies various conditional nonlinearities, such as the asymmetric "leverage effect" of Nelson (1991) that are discussed in Gallant, Rossi, and Tauchen (1992, 1993). We explore the effects of introducing asymmetry into the stochastic volatility model. A common approach in the stochastic volatility literature (Harvey and Shephard, 1996) is to generate asymmetry by introducing correlations across innovations in the mean and variance equations:

$$\begin{aligned} y_t - \mu_y &= c_1(y_{t-1} - \mu_y) + c_2(y_{t-2} - \mu_y) + \exp(w_t) r_y z_t \\ w_t - \mu_w &= a_1(w_{t-1} - \mu_w) + a_2(w_{t-2} - \mu_w) + r_w(\tilde{z}_t + g z_{t-1}) \end{aligned}$$

where g is a free parameter to be estimated. This variant does better but still does poorly on the chi-square statistics shown in Table 3. The bar chart (not shown) shows large SNP quasi-t-ratios, which suggests that the spline-transform be applied to the asymmetric variant. The model that results is

$$\begin{aligned} y_t - \mu_y &= c_1(y_{t-1} - \mu_y) + c_2(y_{t-2} - \mu_y) + \exp(w_t) r_y T_z(z_t) \\ T_z(z_t) &= b_{z0} + b_{z1} z_t + b_{z2}(z_t)^2 + b_{z3} I_+(z_t)(z_t)^2 \\ w_t - \mu_w &= a_1(w_{t-1} - \mu_w) + a_2(w_{t-2} - \mu_w) + r_w(\tilde{z}_t + g z_{t-1}) \end{aligned}$$

The fit improves but is still inadequate, as indicated by the chi-square statistics shown in Table 3.

Finally, we consider long-memory in the variance equation. We estimate with the spline transformation:

$$y_t - \mu_y = c_1(y_{t-1} - \mu_y) + c_2(y_{t-2} - \mu_y) + \exp(w_t^*) r_y T_z(z_t)$$

$$
\begin{aligned}
T_z(z_t) &= b_{z0} + b_{z1} z_t + b_{z2}(z_t)^2 + b_{z3} I_+(z_t)(z_t)^2 \\
w_t^* - \mu_w &= (1-\mathcal{L})^{-d} z_{wt} \\
z_{wt} &= \sum_{j=1}^{L_w} a_j z_{w,t-j} + r_w \tilde{z}_t.
\end{aligned}
$$

We also estimate a model with the spline transformation and cross-correlation in innovations:

$$
\begin{aligned}
y_t - \mu_y &= c_1(y_{t-1} - \mu_y) + c_2(y_{t-2} - \mu_y) + \exp(w_t^*) r_y T_z(z_t) \\
T_z(z_t) &= b_{z0} + b_{z1} z_t + b_{z2}(z_t)^2 + b_{z3} I_+(z_t)(z_t)^2 \\
w_t^* - \mu_w &= (1-\mathcal{L})^{-d} z_{wt} \\
z_{wt} &= \sum_{j=1}^{L_w} a_j z_{w,t-j} + r_w(z_t + g\tilde{z}_{t-1})
\end{aligned}
$$

As seen from the lower two panels of Table 3 long memory helps, but, as in fitting to the Semiparametric ARCH Score, long memory has about the same impact on the objective function as does introducing a few more free lags into the volatility specification. Figure 8 shows the bar chart with long memory for the case $L_w = 2$ and correlated errors. Comparing this figure to Figure 7 shows that the combined effects of the spline transformation, the asymmetry, and the long memory improves the fit substantially, but despite all of these added complications the model fails to fit both the ARCH and SNP scores.

This, we think, is about as far as one can go and stay within the spirit of the stochastic volatility model. A specification that probably would capture the full complexity of the data is to let the coefficients of the transformation

$$
T_z(z_t) = b_{z0} + b_{z1} z_t + b_{z2} z_t^2 + b_{z3} I_+(z_t) z_t^2,
$$

depend upon lagged $z's$ and perhaps add a few more unconstrained lag coefficients. However, this degree of complexity is so close to a nonparametric specification that we see little point to it. Why not just fit the series nonparametrically and have done with it?

5 Trivariate Estimation

Modern asset pricing theory holds that there is a pricing kernel (or marginal rate of substitution) that discounts gross returns to unity. Using methods similar to ours, Andersen

(1996) obtains a good fit of a continuous time stochastic volatility model to high frequency Treasury returns. As Treasury returns reflect pure nominal pricing kernel movements, Andersen's findings taken together with asset pricing theory suggest that a stochastic volatility model should be able to account for the co-movements of several assets. As one of the distinguishing features of the EMM method is its ability to accommodate multivariate data, we investigate this possibility using several assets over a shorter, and therefore potentially more homogeneous, time horizon than in the previous section.

Let y_t denote an $M \times 1$ vector containing the first differences (either simple or logarithmic) over a short time interval, a day for instance, of the prices of a financial asset traded on active speculative markets. A multivariate stochastic volatility model for y_t is

$$\begin{aligned} y_t - \mu_y &= \sum_{j=1}^{L_y} C_j(y_{t-j} - \mu_y) + \text{diag}[\exp(w_t)]R_y z_t \\ w_t - \mu_w &= \sum_{j=1}^{L_w} A_j(w_{t-j} - \mu_w) + R_w \tilde{z}_t \end{aligned}$$

where μ_y is an $M \times 1$ vector, the C_j are $M \times M$ matrices for $j = 1, 2, \ldots, L_y$, and R_y is an $M \times M$ upper triangular matrix. Similarly, μ_w is an $M \times 1$ vector, the A_j are $M \times M$ matrices for $j = 1, 2, \ldots, L_w$, and R_w is an $M \times M$ upper triangular matrix. The processes $\{z_t\}$ and $\{\tilde{z}_t\}$ are mutually independent *iid* random variables with mean zero and variance I_M. Throughout, $\exp(\cdot)$ denotes the elementwise exponential of a vector argument, $\text{diag}(v)$ with a vector argument denotes the diagonal matrix with the elements $v_1, \ldots, v_M$ down the diagonal, and $\text{diag}(B)$ with a matrix argument denotes the vector $(b_{11}, \ldots, b_{MM})'$ with the diagonal elements of B as its elements. Thus,

$$\text{diag}[\exp(w_t)] = \begin{bmatrix} e^{w_{1t}} & 0 & \cdots & 0 \\ 0 & e^{w_{2t}} & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & e^{w_{Mt}} \end{bmatrix}$$

The data to which we fit this stochastic volatility model ($M = 3$) consists of 4,044 daily observation on three variables: adjusted movements of the Standard and Poor's Composite Price Index, adjusted movements of the \$/DM spot exchange rate, and the adjusted 90-day Euro-Dollar interest rate, 1977–92. In this case $M = 3$, $y_t = (y_{1t}, y_{2t}, y_{3t})'$, and the data set

is $\{\tilde{y}_t\}_{t=1}^{4044}$. The raw series consists of the Standard and Poor's Composite Index (SP), the \$/DM exchange rate (DM), and the three-month Euro-dollar interest rate (ED). The three series were collected daily, January 4, 1977–December 31, 1992. The stock index and the exchange rate are converted to raw price movements series, $100[\log(SP_t) - \log(SP_{t-1})]$, and $100[\log(DM_t) - \log(DM_{t-1})]$. The two raw price movement series and the raw ED series are then each adjusted for systematic calendar effects. The adjustment procedure is the same as Gallant, Rossi, and Tauchen (1992) except for the use of a robust regression method instead of ordinary least squares.

The estimation treats the three series as strictly stationary. This seems reasonable for stock returns and exchange rate movements, but requires discussion for the interest rate. As is well known, short-term interest rates collected at high frequencies display extreme persistence characteristic of (near) unit-root processes. However, recent empirical results of Aït-Sahalia (1996), and confirmed in Tauchen (1996), indicate that, although interest rates display little mean revision in the central part of the data, they display substantial mean reversion at very low and very high values. Hence, interest rates appear nonstationary, or nearly so, when considered with linear methodology, when in fact they are stationary when considered with nonlinear methods.

As in Section 4, to implement the EMM estimator we require a score generator that approximates these data well. We use the multivariate SNP model as described in Gallant, Rossi, and Tauchen (1992). It is derived along the same lines as set forth in Section 4.2 and has the following functional form

$$f(y|x,\theta) = \frac{\left\{P[R_x^{-1}(y-\mu_x), x^*]\right\}^2 \phi[R_x^{-1}(y-\mu_x)]}{|\det(R_x)| \int [P(z,x^*)]^2 \phi(z)\, dz}$$

where

$$\mu_x = b_o + Bx^*$$

$$\text{vech}(R_x) = \rho_o + P|x^*|.$$

vech(R) denotes the elements of the upper triangle of R stored as a column vector, $|x|$ denotes element-wise absolute value, x^* is a vector of lagged values of y_t, and $\phi(z) = (2\pi)^{-M/2} e^{-z'z/2}$. The asterisk indicates that prior to forming lags, the y_t have been centered by subtracting

the sample mean, scaled by dividing elementwise by sample standard errors, and then transformed elementwise by the logistic map that takes the interval $(-\infty, \infty)$ into the interval (-4, 4). $P(z, x^*)$ is a polynomial of degree K_z in z whose coefficients are, in turn, polynomials of degree K_x in x^*. μ_x is a function of the first L_u lags in x^* which is accommodated by inserting zeros in B at the appropriate locations; similarly R_x is a function of the first L_r lags in x^* and $P(z, x^*)$ a function of the first L_p lags in x^*. The multivariate model has two additional tuning parameters I_z and I_x that indicate that high order interaction in the polynomial $P(z, x^*)$ have been put to zero: $I_z = 0$ means that no interactions are suppressed, $I_z = 1$ means that the highest order interactions are suppressed, namely those of degree K_z and so on; similarly for K_x. We only allow $P|x^*|$ to contribute to the diagonal of R_x by inserting zeroes in the appropriate elements of P.

As in Section 4, if $K_z = 0$, $K_x = 0$, $L_u > 0$, and $L_r > 0$ then the SNP density is a form of ARCH model with Gaussian innovations. If $K_z > 0$, $K_x = 0$, $L_u > 0$, and $L_r > 0$ then the SNP density is a form of ARCH model with conditionally homogeneous, non-Gaussian innovations. The SNP model with $K_z > 0$ and $K_x = 0$ can accurately approximate any conditionally homogeneous innovation process by taking K_z large enough. If $K_z > 0$, $K_x > 0$, $L_u > 0$, $L_r > 0$, and $L_p > 0$ then the SNP model can accurately approximate any Markovian, stationary process by taking K_z and K_x large enough, including those that exhibit nonlinearities such as conditional skewness and kurtosis (Gallant, Hsieh, and Tauchen, 1991).

We fit the SNP model by quasi maximum likelihood following the protocol that is described in Bansal, Gallant, Hussey, and Tauchen (1995) and is summarized in Section 4. Following this protocol, we select the model $L_u = 4$, $L_r = 16$, $K_z = 8$, and $I_z = 7$ when innovations are constrained to be homogenous ($K_x = 0$, $L_p = 1$). The score from this fit we term the "Semiparametric ARCH Score". We also report results for the model $L_u = 4$, $L_r = 16$, $L_p = 1$, $K_z = 8$, $I_z = 7$, $K_x = 2$, and $I_x = 1$, where the homogeneity constraint is dropped, and term the score from this fit the "Nonlinear Nonparametric Score". We encountered difficulty fitting the stochastic volatility model to the even larger specification, $L_u = 4$, $L_r = 16$, $L_p = 1$, $K_z = 8$, $I_z = 7$, $K_x = 3$, and $I_x = 2$, dictated by following the protocol and do not report EMM results for that score. In all cases, the linear VAR model at the core of the SNP hierarchy is constrained to be zero after lag 2, except for lags of the interest

rate which go out to lag 4, which reflects our prior knowledge that interest rates display much more complicated patterns of linear persistence than do stock returns or exchange rate movements.

Following the EMM procedure described in Section 3 we obtain the chi-square statistics shown in Table 5. As seen from the table, the stochastic volatility model fails to approximate the distribution of these data adequately; it is overwhelmingly rejected.

6 Conclusion

The standard stochastic volatility model, which has received substantial attention in the literature, is an empirically implausible model for stock returns. Our exhaustive search across many specifications indicates that the model must be extended to include (i) an asymmetric thick-tailed distribution for innovations in the mean equation, (ii) long-term dependence in the volatility equation, and (iii) cross correlation between innovations in the mean and volatility equations. When introduced individually, each of these extensions improves the fit somewhat. When introduced together, they produce a stochastic volatility model that is quite elaborate and can accommodate features of the data best described as "Semiparametric ARCH". However, the model still cannot accommodate features that could be described as "Nonlinear Nonparametric." Although not as exhaustive, our investigation for the trivariate data series on stock returns, interest rates, and exchange rates leads to a similar result.

These findings thus cast doubt on the statistical reliability of estimated stochastic volatility models that do not include all three of the extensions. At a minimum, estimates of stochastic volatility models should be accompanied by diagnostic tests in the directions found empirically important here. An even stronger conclusion, which emerges from the failure to fit the Nonlinear Nonparametric features, is that the stochastic volatility model cannot be made to fit financial market data without losing scientific content. The reason is that the conditional heterogeneity in higher moments exhibited by the stochastic volatility model is imparted solely by the volatility equation and therefore cannot be decoupled from the volatility equation. Without the decoupling, the model is not rich enough to approximate data from financial markets. With a decoupling, the stochastic volatility model becomes akin to a nonparametric specification and there are far more computationally convenient nonpara-

metric estimators. Our findings stand in contrast to results of Kim and Shephard (1994), Geweke (1994), and others who find evidence in favor of fairly standard stochastic volatility models. The reason is that we step outside the narrow confines of stochastic volatility and entertain the possibility of very general and flexible auxiliary models. These models provide the diagnostics discrediting stochastic volatility.

Acknowledgement

This material is based upon work supported by the National Science Foundation. We thank Rob Engle and Jorgen Wolters for helpful remarks at various stages of this research, and we thank the two referees and the editor, Helmut Lutkepohl, for thoughtful remarks.

7 References

Aït-Sahalia, Yacine, 1996, Testing continuous-time models of the spot interest rate. Review of Financial Studies, forthcoming.

Andersen, Torben G. and Jesper Lund, 1996, Estimating continuous-time stochastic volatility models, Journal of Econometrics, forthcoming.

Andersen, Torben G. and Bent Sorensen, 1996, GMM estimation of a stochastic volatility model: A monte carlo study, Journal of Business and Economic Statistics, forthcoming. %TeX—

Bansal, Ravi, A. Ronald Gallant, Robert Hussey, and George Tauchen, 1993, Computational aspects of nonparametric simulation estimation, in: David A. Belsley, ed., Computational techniques for econometrics and economic analysis (Kluwer Academic Publishers, Boston, MA) 3–22.

Bansal, Ravi, A. Ronald Gallant, Robert Hussey, and George Tauchen, 1995, Nonparametric estimation of structural models for high-frequency currency market data, Journal of Econometrics 66, 251–287.

Bollerslev, Tim and Hans Ole Æ. Mikkelsen, 1996, Modeling and pricing long-memory in stock market volatility, Journal of Econometrics, forthcoming.

Breidt, F. Jay, Nuno Crato, and Pedro de Lima, 1994, Modeling long-memory stochastic volatility, Manuscript (Iowa State University, Ames, IA).

Clark, Peter K., 1973, A subordinated stochastic process model with finite variance for speculative prices, Econometrica 41, 135–156.

Danielsson, Jon, 1994, Stochastic volatility in asset prices: Estimation with simulated maximum likelihood, Journal of Econometrics 61, 375-400.

Ding, Zhuanxin, Clive W.J. Granger, and Robert F. Engle, 1993, A long memory property of stock market returns and a new model, Journal of Empirical Finance 1, 83–108.

Duffie, Darrell and Kenneth J. Singleton, 1993, Simulated moments estimation of Markov models of asset prices, Econometrica 61, 929–952.

Engle, Robert F. and Gloria Gonzales-Rivera, 1991, Semiparametric ARCH models, Journal of Business and Economic Statistics 9, 345–360.

Fenton, Victor M. and A. Ronald Gallant, 1996a, Convergence rates of SNP density estimators, Econometrica, forthcoming.

Fenton, Victor M. and A. Ronald Gallant, 1996b, Qualitative and asymptotic performance of SNP density estimators, Journal of Econometrics, forthcoming.

Gallant, A. Ronald, 1987, Nonlinear statistical models, (Wiley, New York, NY).

Gallant, A. Ronald, David A. Hsieh, and George Tauchen, 1991, On fitting a recalcitrant series: The pound/dollar exchange rate, 1974–83, in: William A. Barnett, James Powell, and George Tauchen, eds., Nonparametric and semiparametric methods in econometrics and statistics, Proceedings of the fifth international symposium in economic theory and econometrics, (Cambridge University Press, Cambridge) 199–240.

Gallant, A. Ronald and Jonathan R. Long, 1995, Estimating stochastic differential equations efficiently by minimum chi-square, Manuscript (University of North Carolina, Chapel Hill, NC).

Gallant, A. Ronald and Douglas W. Nychka, 1987, Seminonparametric maximum likelihood estimation, Econometrica 55, 363–390.

Gallant, A. Ronald, Peter E. Rossi, and George Tauchen, 1992, Stock prices and volume, Review of Financial Studies 5, 199–242.

Gallant, A. Ronald, Peter E. Rossi, and George Tauchen, 1993, Nonlinear dynamic structures, Econometrica 61, 871–907.

Gallant, A. Ronald and George Tauchen, 1992, A nonparametric approach to nonlinear time series analysis: Estimation and simulation, in: E. Parzen, D. Brillinger, M. Rosenblatt,

M. Taqqu, J. Geweke, and P. Caines, eds., New dimensions in time series analysis, Part II, (Springer-Verlag, New York, NY) 71–92.

Gallant, A. Ronald and George Tauchen, 1989, Seminonparametric estimation of conditionally constrained heterogeneous processes: Asset pricing applications, Econometrica 57, 1091–1120.

Gallant, A. Ronald and George Tauchen, 1996, Which moments to match, Econometric Theory, forthcoming.

Geweke, John, 1994, Bayesian comparison of econometric models, Manuscript (Federal Reserve Bank, Minneapolis, MI).

Ghysels, Eric, Andrew Harvey, and Eric Renault, 1995, Stochastic volatility, in: G. S. Maddala, ed., Handbook of statistics, Vol. 14, Statistical methods in finance, (North Holland, Amsterdam) forthcoming.

Gourieroux, C., A. Monfort, and E. Renault, 1993, Indirect inference, Journal of Applied Econometrics 8, S85–S118.

Granger, C. W. and R. Joyeux, 1980, An introduction to long-memory time series models and fractional differencing, Journal of Time Series Analysis 1, 15-29.

Harvey, Andrew C., 1993, Long memory and stochastic volatility, Manuscript (London School of Economics, London).

Harvey, Andrew C., Ester Ruiz, and Neil Shephard, 1994, Multivariate stochastic variance models, Review of Economic Studies 61, 129–158.

Harvey, Andrew C. and Neil Shephard, 1996, Estimation of an asymmetric stochastic volatility model for asset returns, Journal of Business and Economic Statistics, forthcoming.

Jacquier, Eric, Nicholas G. Polson, and Peter E. Rossi, 1994, Bayesian analysis of stochastic volatility models, Journal of Business and Economic Statistics 12, 371–417.

Kim, Sanjoon and Neil Shephard, 1994, Stochastic volatility: Optimal likelihood inference and comparison with ARCH models, Manuscript (Nuffield College, Oxford).

Mathieu, Ronald and Peter Schotman, 1994, Stochastic volatility and the distribution of exchange rate news, Manuscript (University of Limburg, Maastricht, The Netherlands).

Nelson, D., 1991, Conditional heteroskedasticity in asset returns: A new approach, Econometrica 59, 347–370.

Schwarz, Gideon, 1978, Estimating the dimension of a model, Annals of Statistics 6, 461–464.

Shephard, Neil, 1995, Statistical aspects of ARCH and stochastic volatility, Manuscript (Nuffield College, Oxford).

Sowell, Fallaw, 1990, The fractional unit root distribution, Econometrica 58, 495–508.

Tauchen, George, 1995, The objective function of simulation estimators near the boundary of the stability set, Manuscript (Duke University, Durham, NC).

Tauchen, George, 1996, New minimum chi-square methods in empirical finance, in: D. Kreps and K. Wallis, eds., Advances in econometrics, Seventh world congress, (Cambridge University Press, Cambridge) forthcoming.

Tauchen, George and Mark Pitts, 1983, The price variability-volume relationship on speculative markets, Econometrica 51, 485–505.

Taylor, Steven J., 1986, Modeling financial time series, (Wiley, New York, NY).

Taylor, Steven J., 1994, Modeling stochastic volatility, Mathematical Finance 4, 183–204.

Table 1. Univariate Price Change Series: Optimized Value of the Criterion for the Semiparametric ARCH Score Generator.

Score Generator (SNP)						SV Model			Objective Function		
L_u	L_r	L_p	K_z	K_x	ℓ_θ	L_y	L_w	ℓ_ρ	χ^2	df	p-val
Gaussian											
2	18	1	4	0	26	2	1	6	86.432	20	< 0.0001
2	18	1	4	0	26	2	2	7	79.001	19	< 0.0001
2	18	1	4	0	26	2	3	8	72.672	18	< 0.0001
2	18	1	4	0	26	2	4	9	69.188	17	< 0.0001
2	18	1	4	0	26	2	5	10	67.823	16	< 0.0001
2	18	1	4	0	26	2	6	11	61.093	15	< 0.0001
$t(\nu)$, $\nu = 10, 15, 20, 25$											
2	18	1	4	0	26	2	2	8	78.186	18	< 0.0001
2	18	1	4	0	26	2	2	8	68.931	18	< 0.0001
2	18	1	4	0	26	2	2	8	69.111	18	< 0.0001
2	18	1	4	0	26	2	2	8	69.898	18	< 0.0001
Spline											
2	18	1	4	0	26	2	1	8	41.920	18	0.0011
2	18	1	4	0	26	2	2	9	41.351	17	0.0008
2	18	1	4	0	26	2	3	10	37.700	16	0.0016
2	18	1	4	0	26	2	4	11	36.107	15	0.0017
2	18	1	4	0	26	2	4	12	33.768	14	0.0022
2	18	1	4	0	26	2	6	13	18.638	13	0.1348
Gaussian & Long-Memory											
2	18	1	4	0	26	2	0	6	67.691	20	< 0.0001
2	18	1	4	0	26	2	1	7	67.061	19	< 0.0001
2	18	1	4	0	26	2	2	8	65.463	18	< 0.0001
Spline & Long-Memory											
2	18	1	4	0	26	2	0	8	34.923	18	0.0097
2	18	1	4	0	26	2	1	9	26.718	17	0.0623
2	18	1	4	0	26	2	2	10	21.781	16	0.1504

L_u is the number of lags in the linear part of the SNP model; L_r is the number of lags in the ARCH part; L_p the number of lags in the polynomial part, $P(z, x)$. The polynomial $P(z, x)$ is of degree K_z in z and K_x in x; by convention, $L_p = 1$ if $K_x = 0$. ℓ_θ is the number of free parameters associated with the SNP model. L_y is the number of lags in the linear conditional mean specification of the stochastic volatility model, and L_w is the number of lags in the volatility specification. ℓ_ρ is the number of free parameters of the stochastic volatility model. χ^2 is the EMM objective function scaled to be distributed $\chi^2(df)$ under the maintained assumption of correct specification of the stochastic volatility model. Some relevant quantiles are $\chi^2_{0.99}(20) = 37.566$, $\chi^2_{0.99}(15) = 30.578$.

Table 2. Univariate Price Change Series: Fitted Parameter Values for the Semiparametric ARCH Score Generator.

μ_y	r_y	c_1	c_2	b_{z0}	b_{z1}	b_{z2}	b_{z3}	r_w	a_1	a_2	a_3	a_4	a_5	a_6	d
Gaussian															
0.038	0.927	0.105	0.066		1.000			0.095	0.976						
0.037	0.918	0.105	0.066		1.000			0.155	0.961	-0.662					
0.036	0.909	0.103	0.068		1.000			0.149	0.964	0.010	-0.641				
0.036	0.906	0.103	0.068		1.000			0.172	0.959	-0.209	-0.267	-0.547			
0.037	0.916	0.104	0.068		1.000			0.116	0.972	0.156	-0.185	-0.616	0.304		
0.036	0.909	0.102	0.069		1.000			0.144	0.965	-0.067	-0.592	-0.002	-0.485	0.397	
t															
0.034	0.935	0.103	0.066		$\nu = 10$			0.031	0.993	0.561					
0.036	0.937	0.104	0.066		$\nu = 15$			0.049	0.989	0.378					
0.035	0.928	0.104	0.066		$\nu = 20$			0.124	0.971	-0.521					
0.036	0.926	0.104	0.066		$\nu = 25$			0.132	0.969	-0.580					
Spline															
0.023	0.942	0.110	0.063	0.030	0.921	-0.079	0.097	0.083	0.980						
0.022	0.939	0.109	0.063	0.030	0.933	-0.072	0.083	0.129	0.969	-0.537					
0.022	0.931	0.109	0.064	0.032	0.960	-0.056	0.049	0.138	0.966	0.036	-0.623				
0.021	0.923	0.108	0.064	0.032	0.988	-0.039	0.014	0.212	0.947	-0.558	-0.479	-0.356			
0.022	0.935	0.109	0.063	0.029	0.928	-0.073	0.089	0.029	0.993	1.634	-1.853	1.078	-0.247		
0.023	0.947	0.109	0.064	0.029	0.942	-0.064	0.071	0.012	0.998	2.005	-1.970	0.939	0.220	-0.378	
Gaussian & Long Memory															
0.036	0.908	0.104	0.067		1.000			0.224							0.540
0.035	0.904	0.103	0.067		1.000			0.242	-0.159						0.550
0.034	0.908	0.104	0.066		1.000			0.221	0.039	-0.192					0.539
Spline & Long Memory															
0.022	0.922	0.111	0.062	0.033	1.046	-0.003	-0.059	0.259							0.515
0.021	0.887	0.110	0.060	0.038	1.208	0.095	-0.267	0.369	-0.268						0.493
0.020	0.863	0.109	0.061	0.044	1.355	0.187	-0.463	0.429	-0.419	0.128					0.486

The rows of Table 2 correspond to the rows of Table 1. Due to identification restrictions across parameters, the number of parameters in a row do not necessarily correspond to the number of free parameters shown in Table 1.

Table 3. Univariate Price Change Series: Optimized Value of the Criterion for the Nonlinear Nonparametric Score Generator.

Score Generator (SNP)						SV Model			Objective Function		
L_u	L_r	L_p	K_z	K_x	ℓ_θ	L_y	L_w	ℓ_ρ	χ^2	df	p-val
Gaussian											
2	18	2	4	1	36	2	1	6	173.361	30	< 0.0001
2	18	2	4	1	36	2	2	7	164.337	29	< 0.0001
2	18	2	4	1	36	2	3	8	155.449	28	< 0.0001
2	18	2	4	1	36	2	4	9	151.243	27	< 0.0001
2	18	2	4	1	36	2	5	10	149.350	26	< 0.0001
2	18	2	4	1	36	2	6	11	147.984	25	< 0.0001
Spline											
2	18	2	4	1	36	2	1	8	151.290	28	< 0.0001
2	18	2	4	1	36	2	2	9	150.765	27	< 0.0001
2	18	2	4	1	36	2	3	10	144.411	26	< 0.0001
2	18	2	4	1	36	2	4	11	143.310	25	< 0.0001
2	18	2	4	1	36	2	5	12	143.310	24	< 0.0001
2	18	2	4	1	36	2	6	13	142.461	23	< 0.0001
Gaussian-Asymmetric											
2	18	2	4	1	36	2	1	7	111.497	29	< 0.0001
2	18	2	4	1	36	2	2	8	111.487	28	< 0.0001
2	18	2	4	1	36	2	3	9	97.536	27	< 0.0001
2	18	2	4	1	36	2	4	10	93.969	26	< 0.0001
2	18	2	4	1	36	2	5	11	91.075	25	< 0.0001
2	18	2	4	1	36	2	6	12	85.711	24	< 0.0001
Spline-Asymmetric											
2	18	2	4	1	36	2	1	9	78.972	27	< 0.0001
2	18	2	4	1	36	2	2	10	78.197	26	< 0.0001
2	18	2	4	1	36	2	3	11	75.483	25	< 0.0001
2	18	2	4	1	36	2	4	12	70.109	24	< 0.0001
2	18	2	4	1	36	2	5	13	69.881	23	< 0.0001
2	18	2	4	1	36	2	6	14	69.645	22	< 0.0001
Spline & Long Memory											
2	18	2	4	1	36	2	0	8	152.654	28	< 0.0001
2	18	2	4	1	36	2	1	9	146.479	27	< 0.0001
2	18	2	4	1	36	2	2	10	143.477	26	< 0.0001
Spline-Asymmetric & Long Memory											
2	18	2	4	1	36	2	0	9	94.678	27	< 0.0001
2	18	2	4	1	36	2	1	10	72.049	26	< 0.0001
2	18	2	4	1	36	2	2	11	71.609	25	< 0.0001

L_u is the number of lags in the linear part of the SNP model; L_r is the number of lags in the ARCH part; L_p the number of lags in the polynomial part, $P(z, x)$. The polynomial $P(z, x)$ is of degree K_z in z and K_x in x; by convention, $L_p = 1$ if $K_x = 0$. ℓ_θ is the number of free parameters associated with the SNP model. L_y is the number of lags in the linear conditional mean specification of the stochastic volatility model, and L_w is the number of lags in the volatility specification. ℓ_ρ is the number of free parameters of the stochastic volatility model. χ^2 is the EMM objective function scaled to be distributed $\chi^2(df)$ under the maintained assumption of correct specification of the stochastic volatility model. Some relevant quantiles are $\chi^2_{0.99}(30) = 50.892$, $\chi^2_{0.99}(25) = 44.314$, $\chi^2_{0.99}(20) = 37.566$.

Table 4. Univariate Price Change Series: Fitted Parameter Values for the Nonlinear Non-parametric Score Generator.

μ_y	r_y	c_1	c_2	g	b_{z0}	b_{z1}	b_{z2}	b_{z3}	r_w	a_1	a_2	a_3	a_4	a_5	a_6	d
Gaussian																
0.052	0.767	0.114	0.052			1.000			0.092	0.966						
0.051	0.762	0.114	0.051			1.000			0.152	0.944	-0.712					
0.050	0.756	0.111	0.052			1.000			0.195	0.930	-0.826	-0.510				
0.050	0.748	0.109	0.054			1.000			0.223	0.920	-0.746	-0.625	-0.404			
0.050	0.746	0.108	0.054			1.000			0.229	0.919	-0.642	-0.733	-0.318	-0.284		
0.050	0.745	0.109	0.054			1.000			0.229	0.919	-0.642	-0.732	-0.318	-0.284	-0.009	
Spline																
0.046	0.769	0.114	0.052		0.010	0.905	-0.068	0.117	0.072	0.974						
0.047	0.770	0.114	0.052		0.010	0.901	-0.071	0.123	0.038	0.986	0.475					
0.047	0.770	0.115	0.052		0.009	0.914	-0.062	0.106	0.010	0.997	1.651	-0.802				
0.047	0.771	0.115	0.051		0.009	0.915	-0.061	0.105	0.019	0.994	0.670	0.834	-0.799			
0.047	0.771	0.115	0.051		0.009	0.915	-0.061	0.105	0.019	0.994	0.670	0.834	-0.799			
0.047	0.772	0.115	0.051		0.009	0.914	-0.063	0.107	0.046	0.986	-0.146	0.549	0.575	0.000	-0.695	
Gaussian-Asymmetric																
0.047	0.839	0.115	0.055	-0.791		1.000			0.070	0.976						
0.047	0.839	0.115	0.055	-0.801		1.000			0.068	0.976	0.018					
0.045	0.834	0.114	0.057	-0.494		1.000			0.122	0.964	0.076	-0.671				
0.045	0.836	0.114	0.058	-0.510		1.000			0.106	0.970	0.286	-0.552	-0.267			
0.045	0.839	0.114	0.057	-0.630		1.000			0.085	0.974	0.220	-0.128	-0.690	0.361		
0.045	0.837	0.112	0.058	-0.547		1.000			0.101	0.970	0.212	-0.806	0.307	-0.610	0.479	
Spline-Asymmetric																
0.033	0.849	0.119	0.054	-1.454	0.024	0.919	-0.074	0.100	0.045	0.979						
0.033	0.848	0.119	0.054	-1.720	0.024	0.917	-0.075	0.102	0.035	0.982	0.134					
0.032	0.863	0.117	0.053	-1.038	0.026	0.924	-0.073	0.094	0.054	0.980	0.427	-0.489				
0.031	0.856	0.117	0.056	-0.938	0.026	0.939	-0.063	0.075	0.065	0.977	0.361	-0.808	0.329			
0.031	0.857	0.117	0.056	-0.859	0.026	0.942	-0.061	0.071	0.069	0.977	0.430	-0.882	0.390	-0.083		
0.032	0.856	0.117	0.055	-1.063	0.025	0.939	-0.062	0.075	0.055	0.979	0.279	-0.224	-0.191	0.622	-0.491	
Spline & Long Memory																
0.046	0.760	0.114	0.052		0.010	0.965	-0.032	0.044	0.200							0.507
0.046	0.736	0.112	0.051		0.011	1.075	0.037	-0.094	0.304	-0.315						0.484
0.046	0.724	0.110	0.052		0.012	1.149	0.083	-0.190	0.351	-0.510	0.164					0.485
Spline-Asymmetric & Long Memory																
0.035	0.829	0.119	0.054	-0.431	0.021	0.975	-0.037	0.031	0.179							0.541
0.031	0.849	0.121	0.052	-1.030	0.024	0.931	-0.067	0.085	0.060	0.664						0.496
0.031	0.850	0.121	0.052	-1.034	0.025	0.932	-0.066	0.084	0.056	0.677	0.111					0.501

The rows of Table 4 correspond to the rows of Table 3. Due to identification restrictions across parameters, the number of parameters in a row do not necessarily correspond to the number of free parameters shown in Table 3.

Table 5. Trivariate Series: Optimized Value of the Criterion

Score Generator (SNP)								SV Model			Objective Function		
L_u	L_r	L_p	K_z	I_z	K_x	I_x	ℓ_θ	L_y	L_w	ℓ_ρ	χ^2	df	p-val
4	16	1	8	7	0	0	101	2	1	44	490.306	57	< 0.0001
4	16	1	8	7	0	0	101	2	2	47	329.603	54	< 0.0001
4	16	1	8	7	2	1	251	2	3	47	4168.470	204	< 0.0001

L_u is the number of lags in the linear part of the SNP model; L_r is the number of lags in the ARCH part; L_p the number of lags in the polynomial part, $P(z, x)$. The polynomial $P(z, x)$ is of degree K_z in z, with interactions of degree exceeding $K_z - I_z$ suppressed; likewise, $P(z, x)$ is of degree K_x in x, with interactions of degree exceeding $K_x - I_x$ suppressed. By convention, $L_p = 1$ if $K_x = 0$. ℓ_θ is the number of free parameters associated with the SNP model. L_y is the number of lags in the linear conditional mean specification of the stochastic volatility model, and L_w is the number of lags in the volatility specification. ℓ_ρ is the number of free parameters of the stochastic volatility model. χ^2 is the EMM objective function scaled to be distributed $\chi^2(df)$ under the maintained assumption of correct specification of the stochastic volatility model.

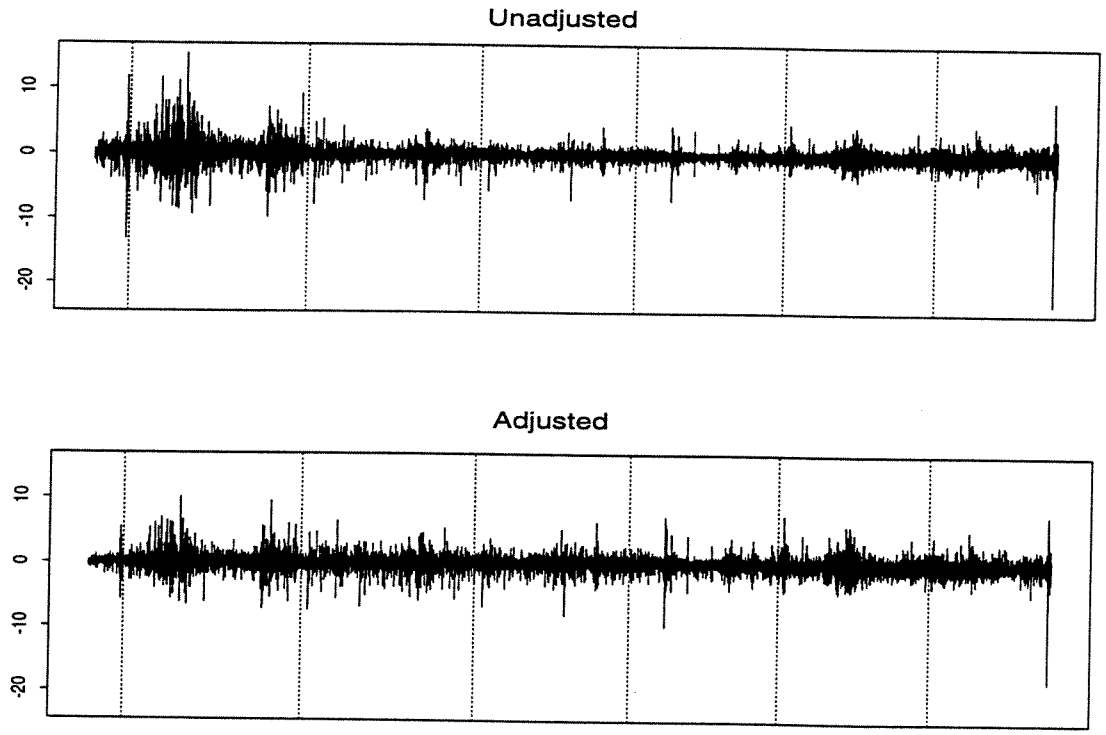

Figure 1. Time series of unadjusted and adjusted stock price movements. The top panel shows a time series plot of the daily unadjusted price movement series, $100(\log P_t - \log P_{t-1})$. The data are daily from 1928 to 1987, 16,127 observations. The bottom panel shows the adjusted price movement series. The adjustments remove calendar effects and long-term trend on the basis of least squares regressions. The adjusted series can reasonably be taken as stationary, which is required for use of the SNP estimator. See Section 1 of Gallant, Rossi, and Tauchen (1992) for a description of the adjustment procedure.

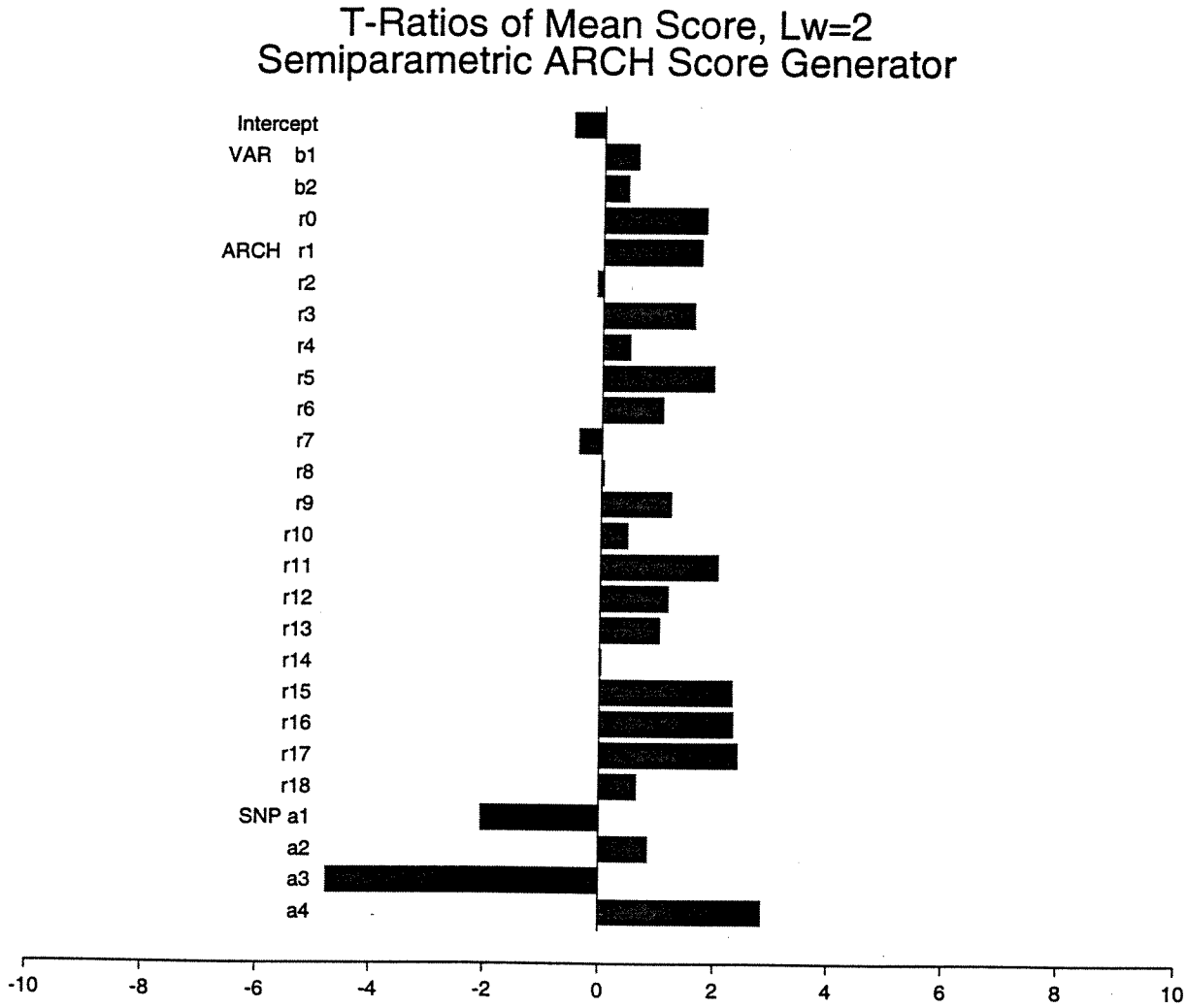

Figure 2. EMM Quasi-t-Ratios for the Stochastic Volatility Model Matched to the Semiparametric ARCH Score. The Semiparametric ARCH score is an SNP specification with $L_u = 2$, $L_r = 18$, $L_p = 1$, $K_z = 4$, $I_z = 0$, $K_x = 0$, and $I_x = 0$. The VAR t-ratios and ARCH t-ratios shown in the plot correspond to the equations $\mu_x = b_o + b'x$ and $r_x = \rho_o + \rho'x$ of the SNP specification, respectively. The SNP t-ratios correspond to the coefficients of the polynomial $P(z, x)$ of the SNP specification where the subscript indicates degree. The stochastic volatility specification is $y_t - \mu_y = c_1(y_{t-1} - \mu_y) + c_2(y_{t-2} - \mu_y) + \exp(w_t) r_y z_t$, $w_t - \mu_w = a_1(w_{t-1} - \mu_w) + a_2(w_{t-2} - \mu_w) + r_w \tilde{z}_t$.

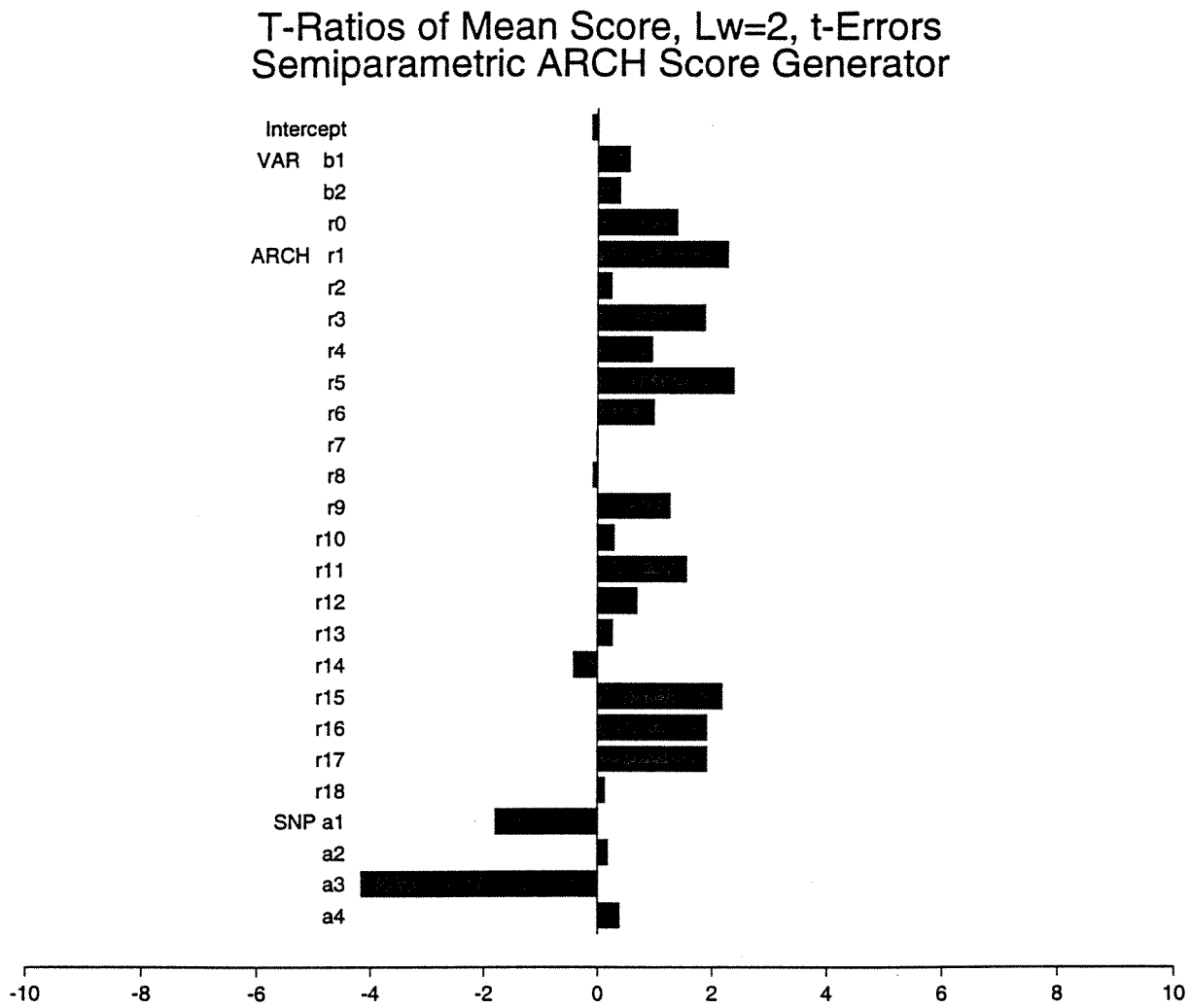

Figure 3. EMM Quasi-t-Ratios for the t-Innovations Stochastic Volatility Model Matched to the Semiparametric ARCH Score. The Semiparametric ARCH score is an SNP specification with $L_u = 2$, $L_r = 18$, $L_p = 1$, $K_z = 4$, $I_z = 0$, $K_x = 0$, and $I_x = 0$. The VAR t-ratios and ARCH t-ratios shown in the plot correspond to the equations $\mu_x = b_o + b'x$ and $r_x = \rho_o + \rho'x$ of the SNP specification, respectively. The SNP t-ratios correspond to the coefficients of the polynomial $P(z, x)$ of the SNP specification where the subscript indicates degree. The stochastic volatility specification is $y_t - \mu_y = c_1(y_{t-1} - \mu_y) + c_2(y_{t-2} - \mu_y) + \exp(w_t)r_y\tau_{15,t}$, $w_t - \mu_w = a_1(w_{t-1} - \mu_w) + a_2(w_{t-2} - \mu_w) + r_w\tilde{z}_t$, where $\tau_{15,t}$ follows the t-distribution on 15 degrees freedom.

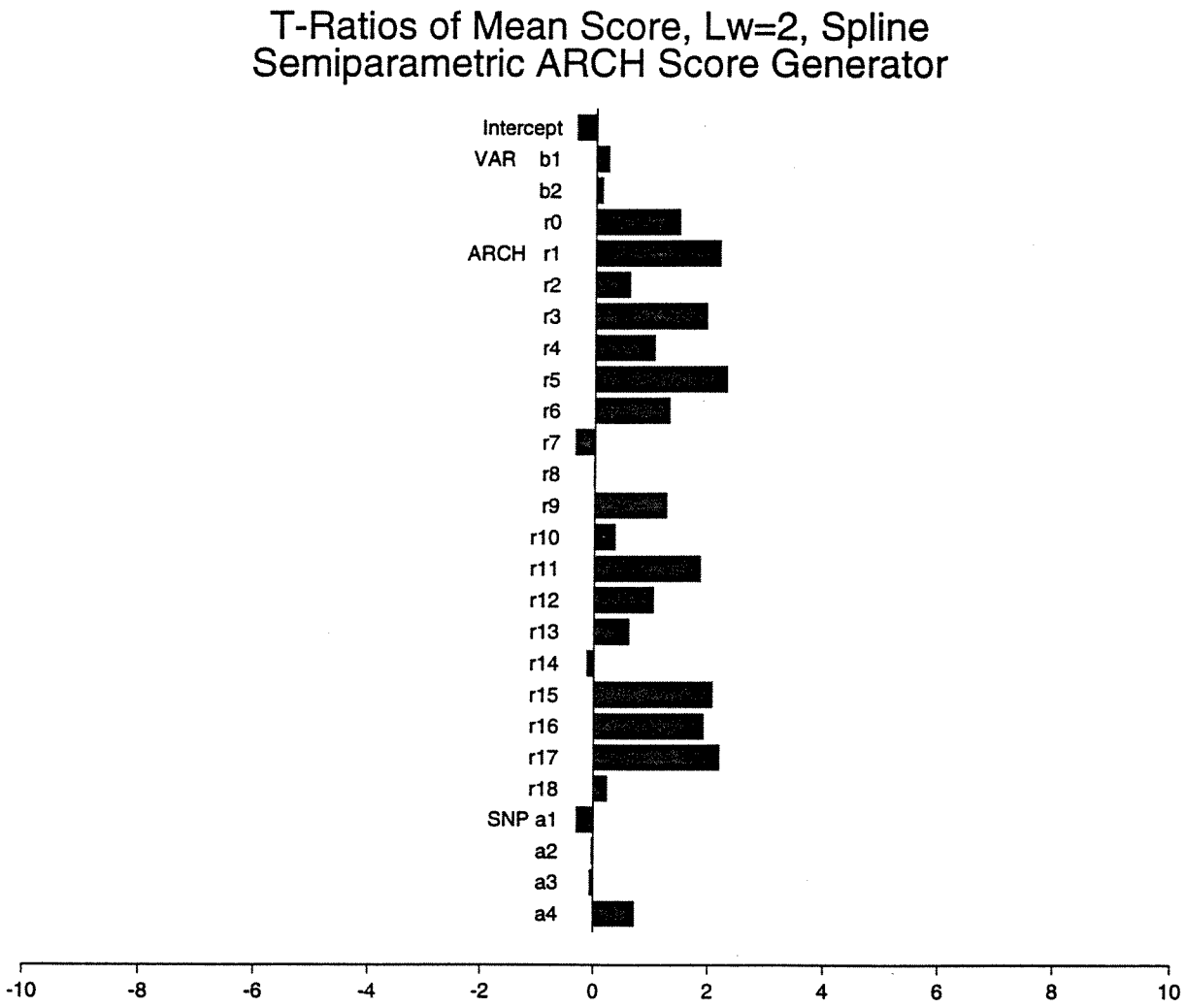

Figure 4. EMM Quasi-t-Ratios for the Spline-Transform Stochastic Volatility Model Matched to the Semiparametric ARCH Score. The Semiparametric ARCH score is an SNP specification with $L_u = 2$, $L_r = 18$, $L_p = 1$, $K_z = 4$, $I_z = 0$, $K_x = 0$, and $I_x = 0$. The VAR t-ratios and ARCH t-ratios shown in the plot correspond to the equations $\mu_x = b_o + b'x$ and $r_x = \rho_o + \rho'x$ of the SNP specification, respectively. The SNP t-ratios correspond to the coefficients of the polynomial $P(z, x)$ of the SNP specification where the subscript indicates degree. The stochastic volatility specification is $y_t - \mu_y = c_1(y_{t-1} - \mu_y) + c_2(y_{t-2} - \mu_y) + \exp(w_t) r_y T_z(z_t)$, $T_z(z_t) = b_{z0} + b_{z1} z_t + b_{z2} z_t^2 + b_{z3} I_+(z_t) z_t^2$, $w_t - \mu_w = a_1(w_{t-1} - \mu_w) + a_2(w_{t-2} - \mu_w) + r_w \tilde{z}_t$.

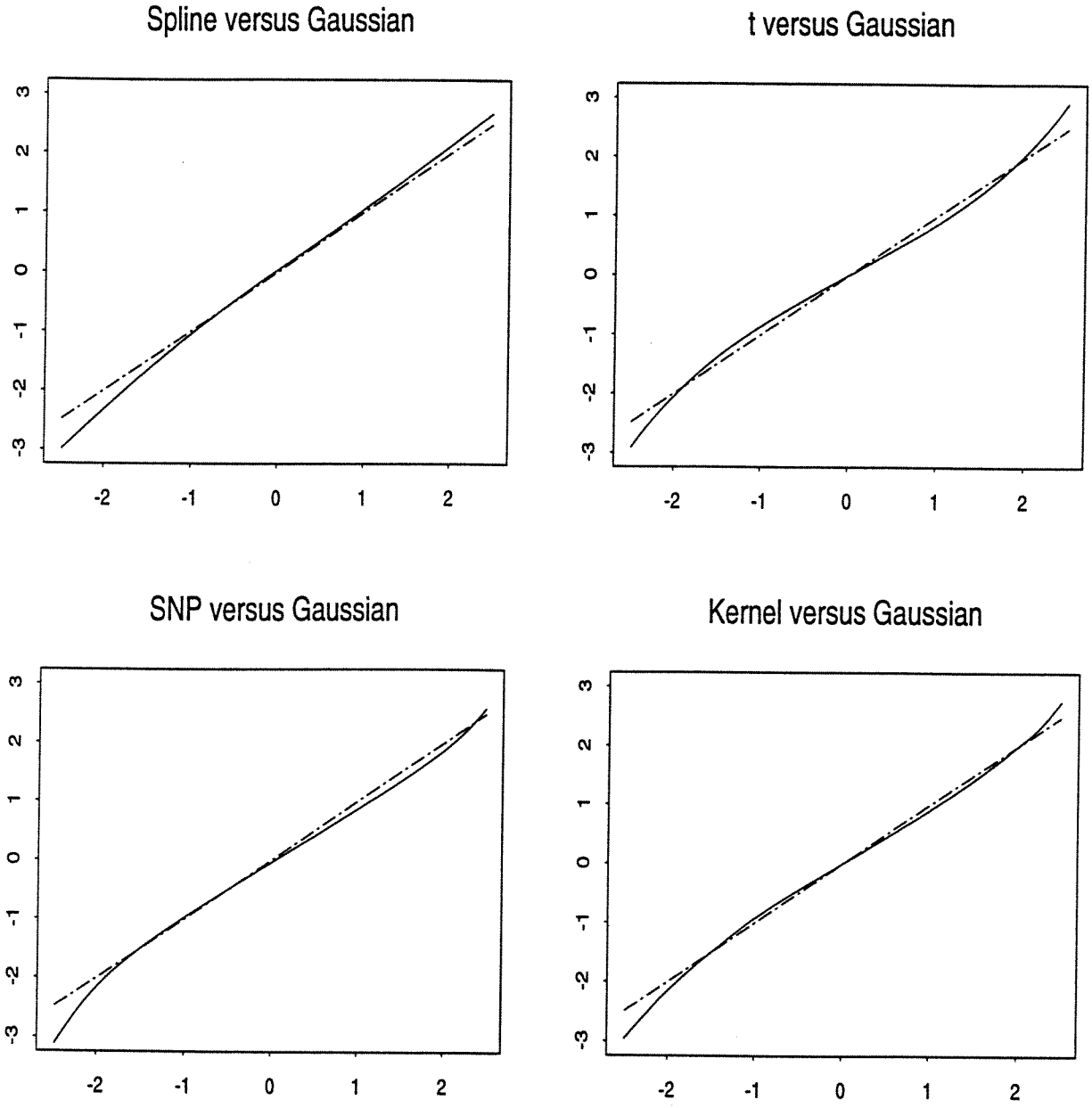

Figure 5. Quantile-Quantile Plots. The solid line in the upper left panel shows the spline transform of Figure 4 which can also be interpreted as a plot of the quantiles of the distribution of the random variable $T_z(z_t)$ on the vertical axis against the quantiles of the standard normal distribution on the horizontal axis. The dashed is a plot of the quantiles of the standard normal against the quantiles of the standard normal. The solid line in the upper right panel is a quantile-quantile plot of the five degree freedom Student t-distribution. The solid line in the lower left panel is a quantile-quantile plot of the innovation distribution of the Semiparametric ARCH Score Generator. The solid line in the lower right panel is a quantile-quantile plot of a kernel density estimate from ARCH residuals.

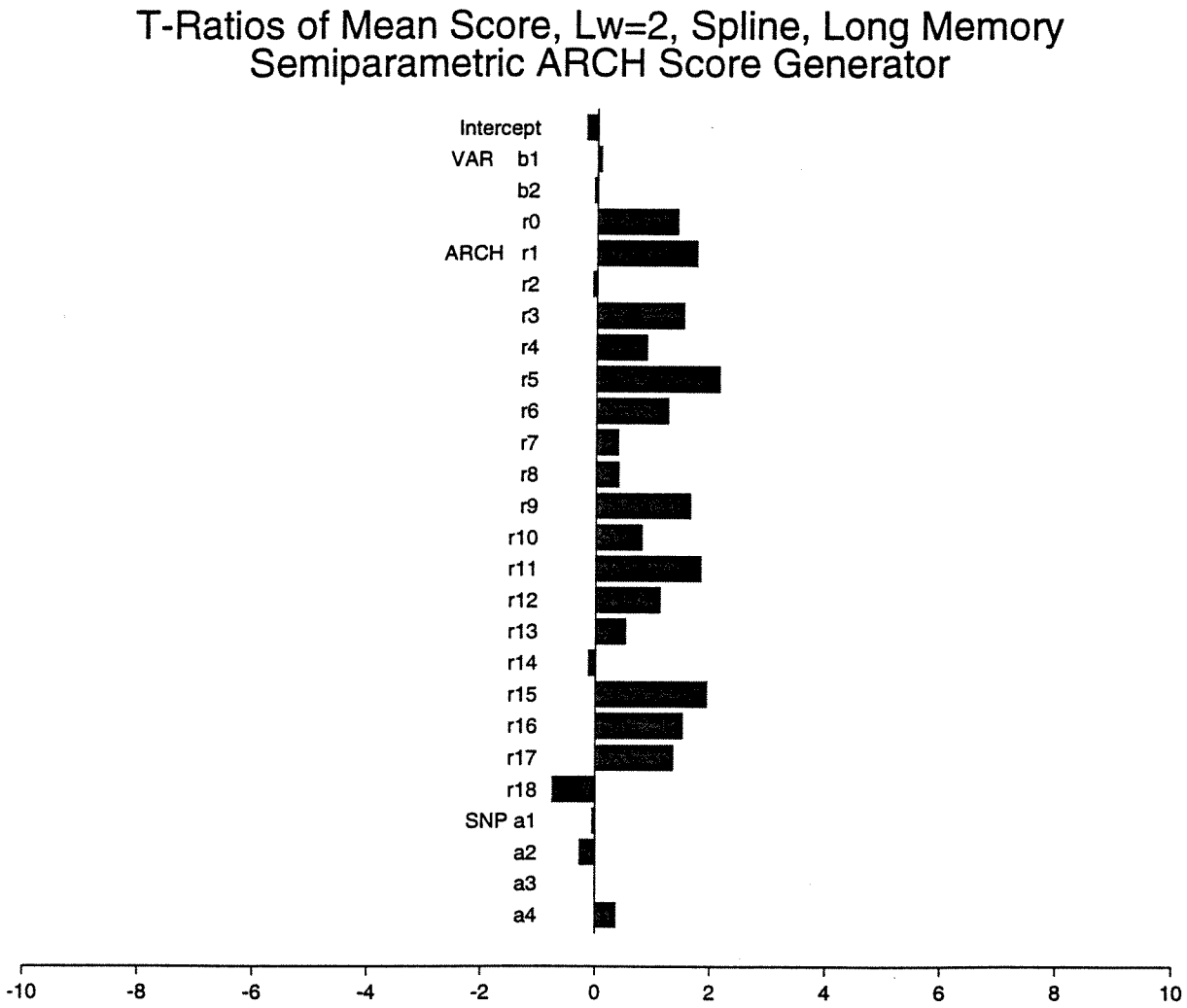

Figure 6. EMM Quasi-t-Ratios for the Spline-Transform Stochastic Volatility Model with a Long-Memory Variance Equation Matched to the Semiparametric ARCH Score. The Semiparametric ARCH score is an SNP specification with $L_u = 2$, $L_r = 18$, $L_p = 1$, $K_z = 4$, $I_z = 0$, $K_x = 0$, and $I_x = 0$. The VAR t-ratios and ARCH t-ratios shown in the plot correspond to the equations $\mu_x = b_o + b'x$ and $r_x = \rho_o + \rho' x$ of the SNP specification, respectively. The SNP t-ratios correspond to the coefficients of the polynomial $P(z, x)$ of the SNP specification where the subscript indicates degree. The stochastic volatility specification is $y_t - \mu_y = c_1(y_{t-1} - \mu_y) + c_2(y_{t-2} - \mu_y) + \exp(w_t^*) r_y T_z(z_t)$, $T_z(z_t) = b_{z0} + b_{z1} z_t + b_{z2} z_t^2 + b_{z3} I_+(z_t) z_t^2$, $w_t^* - \mu_w = (1 - \mathcal{L})^{-d} z_{wt}$, $z_{wt} = a_1 z_{w,t-1} + a_2 z_{w,t-2} + r_w \tilde{z}_t$.

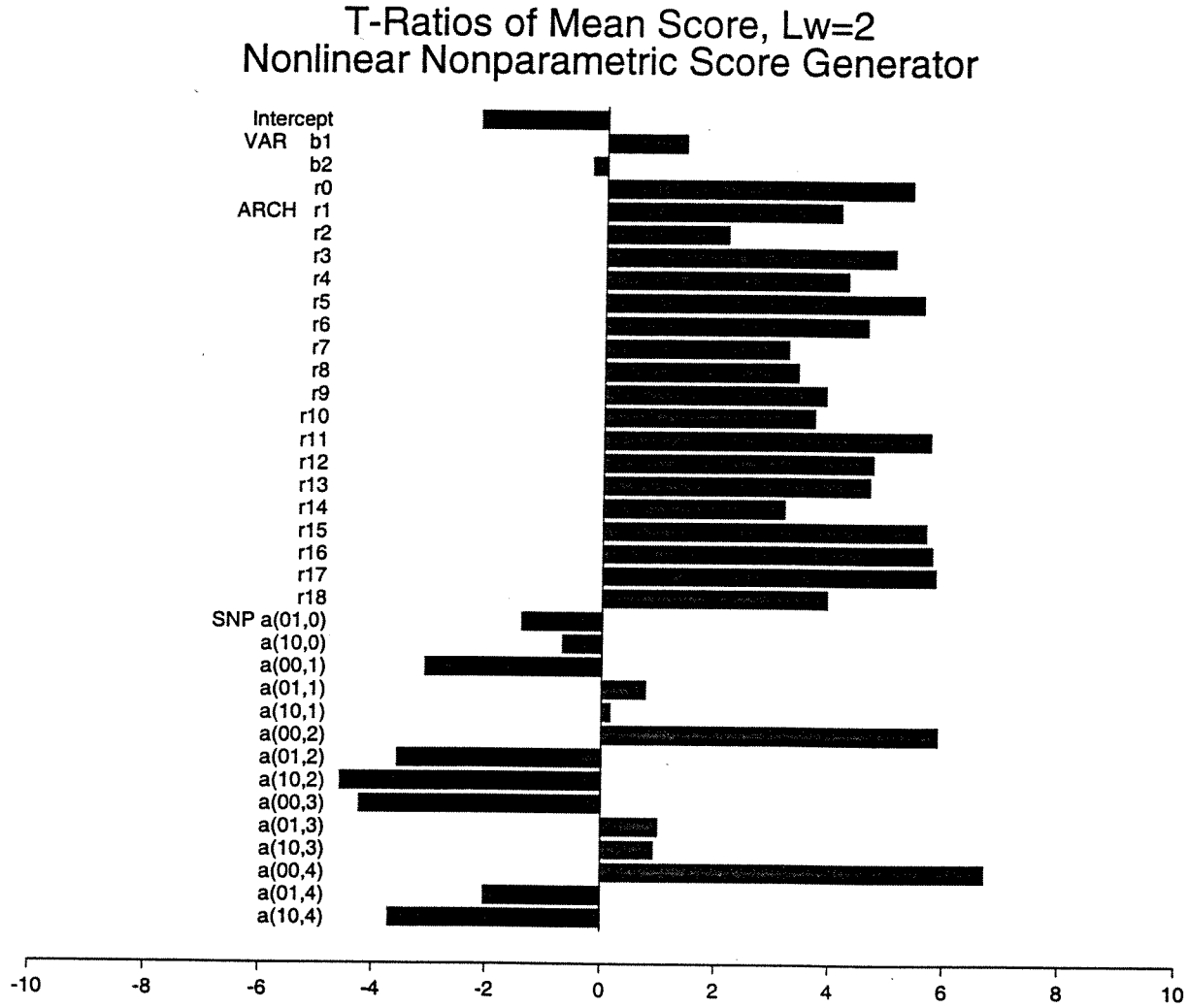

Figure 7. EMM Quasi-t-Ratios for the Stochastic Volatility Model Matched to the Nonlinear Nonparametric Score. The Nonlinear Nonparametric score is an SNP specification with $L_u = 2$, $L_r = 18$, $L_p = 2$, $K_z = 4$, $I_z = 0$, $K_x = 0$, and $I_x = 0$. The VAR t-ratios and ARCH t-ratios shown in the plot correspond to the equations $\mu_x = b_o + b'x$ and $r_x = \rho_o + \rho'x$ of the SNP specification, respectively. The SNP t-ratios correspond to the coefficients of the polynomial $P(z, x)$ of the SNP specification. A coefficient such as $a(00, 2)$ corresponds to the monomial z^2, one such as a(10,2) to z^2x_1, a(01,2) to z^2x_2, and so on. The stochastic volatility specification is $y_t - \mu_y = c_1(y_{t-1} - \mu_y) + c_2(y_{t-2} - \mu_y) + \exp(w_t)r_y z_t$, $w_t - \mu_w = a_1(w_{t-1} - \mu_w) + a_2(w_{t-2} - \mu_w) + r_w \tilde{z}_t$.

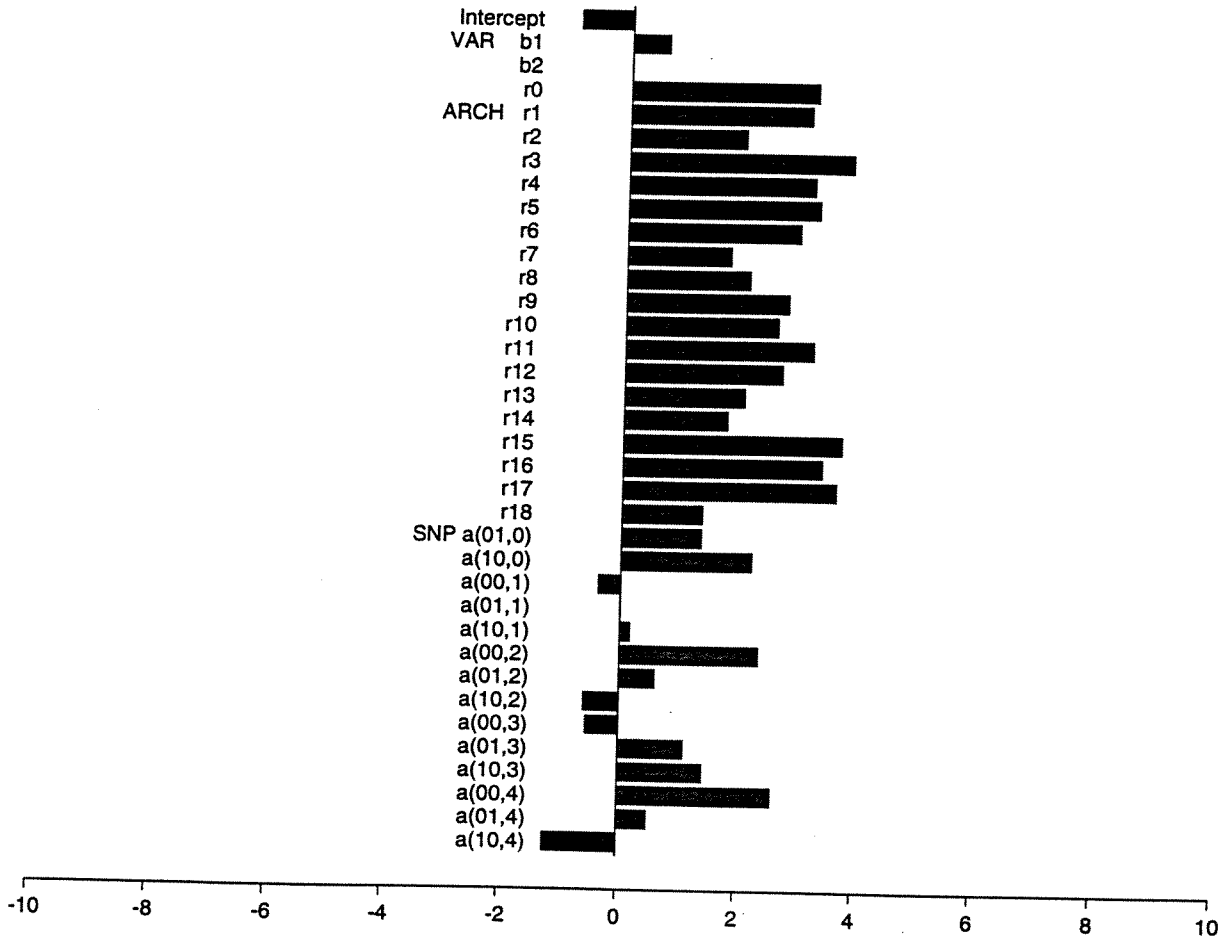

Figure 8. EMM Quasi-t-Ratios for the Asymmetric, Spline-Transform Stochastic Volatility Model with a Long-Memory Variance Equation Matched to the Nonlinear Nonparametric Score. The Nonlinear Nonparametric score is an SNP specification with $L_u = 2$, $L_r = 18$, $L_p = 2$, $K_z = 4$, $I_z = 0$, $K_x = 0$, and $I_x = 0$. The VAR t-ratios and ARCH t-ratios shown in the plot correspond to the equations $\mu_x = b_o + b'x$ and $r_x = \rho_o + \rho' x$ of the SNP specification, respectively. The SNP t-ratios correspond to the coefficients of the polynomial $P(z, x)$ of the SNP specification. A coefficient such as $a(00, 2)$ corresponds to the monomial z^2, one such as a(10,2) to $z^2 x_1$, a(01,2) to $z^2 x_2$, and so on. The stochastic volatility specification is $y_t - \mu_y = c_1(y_{t-1} - \mu_y) + c_2(y_{t-2} - \mu_y) + \exp(w_t^*) r_y T_z(z_t)$, $T_z(z_t) = b_{z0} + b_{z1} z_t + b_{z2}(z_t)^2 + b_{z3} I_+(z_t)(z_t)^2$, $w_t^* - \mu_w = (1 - \mathcal{L})^{-d} z_{wt}$, $z_{wt} = \sum_{j=1}^{L_w} a_j z_{w,t-j} + r_w(\tilde{z}_t + g z_{t-1})$.