Estimation of Continuous Time Models for Stock Returns and Interest Rates [1]

A. Ronald Gallant
Department of Economics
University of North Carolina
Chapel Hill NC 27599-3305 USA

George Tauchen
Department of Economics
Duke University
Durham NC 27708-0097 USA

August 1995
Last revised: January 1997

[1]Supported by the National Science Foundation. The corresponding author is A. Ronald Gallant, Department of Economics, University of North Carolina, CB# 3305, 6F Gardner Hall, Chapel Hill NC 27599-3305 USA; Phone 1-919-966-5338; FAX 1-919-966-4986; e-mail ron_gallant@unc.edu. Updated versions of this paper are available by anonymous ftp at site ftp.econ.duke.edu in directory pub/arg/papers as file ct.psz, which is compressed PostScript.

Abstract

Efficient Method of Moments (EMM) is used to estimate and test continuous time diffusion models for stock returns and interest rates. For stock returns, a four-state, two-factor diffusion with one state observed can account for the dynamics of the daily return on the S&P composite index, 1927–1987. This contrasts with results indicating that discrete-time, stochastic volatility models cannot explain these dynamics. For interest rates, a trivariate yield factor model is estimated from weekly, 1962–1995, Treasury rates. The yield factor model is sharply rejected, although extensions permitting convexities in the local variance come closer to fitting the data.

Key words: Diffusions, Efficient Method of Moments, EMM, Stochastic Differential Equations, Stock Returns, Term Structure of Interest Rates, Yield Curve, Yield Factor Model.

1 Introduction and Summary

In this article, we estimate and test several multi-state, multi-factor, diffusion models of stock returns and interest rates that are expressed as systems of stochastic differential equations (SDE). Using daily returns on the S&P Composite Index, 1927–1987, we sequentially determine a four-state, continuous-time, asymmetric, stochastic volatility model by means the informative Efficient Method of Moments (EMM) diagnostics proposed by Gallant and Tauchen (1996a). This model has two Brownian forcing terms, or factors, and observables are generated by sampling the last of the four states at a daily frequency. Using weekly, 1962–1995, observations on the term structure at three months, one year, and ten years, we fit several variants of the Yield Factor Model proposed by Duffie and Kan (1993). This is a three-state model with three Brownian forcing terms. The three states are sampled at weekly frequency to generate observables. All variants of the Yield Factor Model are sharply rejected by the data, although the variants that have convex local variance function do better than those that do not.

We use and extend the Gallant and Long (1997) adaptation of the EMM estimator proposed by Bansal, Gallant, Hussey, and Tauchen (1993, 1995). The estimator is a minimum chi-square, method of moments estimator. The moment function that enters the chi-squared criterion is the expectation — with respect to the invariant measure determined by discretely sampling the continuous-time system — of the score of a transition density proposed by Gallant and Tauchen (1989) in connection with nonparametric time series analysis. The value of the optimized objective function is asymptotically chi-squared, and thus it can be used for forming confidence intervals and testing system adequacy.

The nonparametric density has three identifiable components: a linear location function, which is a function of past observations, an ARCH scale function, which is a function of past observations, and a Hermite polynomial, which is a function of both contemporaneous and past observations. If a fitted stochastic differential equation is rejected by the diagnostic tests, then studentized scores associated with the parameters of the three components suggest how the system can be modified to improve the fit.

The estimator is similar in some respects to the dynamic simulation estimators proposed by Duffie and Singleton (1993), Ingram and Lee (1991), and others. Very long simulations are used to compute expectations given a candidate value of the parameter vector. It differs from these in recommending specific moment functions that guarantee efficiency rather than permitting ad hoc selection. EMM is less computationally demanding than the simulation estimator proposed by Gourieroux, Monfort, and Renault (1993) and Broze, Scaillet, Zakoian (1995) because it circumvents computation of a binding function and a Hessian at each candidate parameter value.

Related estimation strategies for stochastic differential equations are due to Aït-Sahalia (1996) and Hansen and Scheinkman (1995). Both strategies rely on moment functions computed directly from the data, rather than moment functions computed by simulation, and they require the state to be fully observed in order to estimate all of the parameters. Hansen and Scheinkman (1995) obtain conditions on the infinitesimal generator determined by the system that guarantee a strong law of large numbers and a central limit theorem for discretely sampled observations. When a strong law holds, the minimum chi-square estimator proposed here is consistent. The smoothness conditions discussed in Gallant and Long (1997) are required for asymptotic normality and efficiency, but not for consistency.

Direct application of maximum likelihood, in general, is more difficult than the methods proposed here, even when the state is completely observed, because determining the discrete time transition density, in general, requires solving partial differential equations numerically (Lo, 1988). The numerical methods required here are similar to Runge-Kutta methods for ordinary differential equations and are therefore much easier to implement. In addition, direct maximum likelihood estimation does not afford suggestive diagnostics.

The paper is organized as follows. In Section 2 we describe the minimum chi-square estimator and diagnostic tests. In Section 3 we describe the Gallant-Tauchen nonparametric time-series estimator. In Section 4 we discuss methods for simulating systems of stochastic differential equation. In Section 5 we review recent work on fitting discrete time stochastic volatility models and then estimate continuous time models of stock returns. In Section 6 we estimate various continuous time models of the term structure of interest rates, starting with the Yield-Factor Model of Duffie and Kan (1993) and proceeding on to various extensions suggested by model diagnostics.

2 Efficient Method of Moments

We should like to estimate the parameter ρ that appears in the system of stochastic differential equations

$$dU_t = A(U_t, \rho)dt + B(U_t, \rho)dW_t \qquad 0 \le t < \infty.$$

The parameter ρ has dimension p_ρ, the state vector U_t has dimension d, W_t is a k-dimensional vector of independent Wiener processes, $A(\cdot, \rho)$ maps $\Re^d$ into $\Re^d$, and $B(\cdot, \rho)$ is a $d \times k$ matrix comprised of the column vectors $B_1(\cdot, \rho), \ldots, B_k(\cdot, \rho)$, each of which maps $\Re^d$ into $\Re^d$. U_t is interpreted as the solution of the integral equations

$$U_t = U_0 + \int_0^t A(U_s, \rho)\, ds + \sum_{i=1}^k \int_0^t B_i(U_s, \rho)\, dW_{is},$$

where U_0 is the initial condition at time $t = 0$, and $\int_0^t B_i(U_s, \rho)\, dW_{is}$ denotes the Ito stochastic integral (Gihman and Skorohod, 1972).

The system is observed at equally spaced time intervals $t = 0, 1, \ldots$ and selected characteristics

$$y_{t-L} = T(U_t) \qquad t = 0, 1, \ldots \tag{1}$$

of the state are recorded, where y_t is an M-dimensional vector and $L > 0$ is the number of lagged variables that enter formulas which follow. It will be important throughout to distinguish data, simulations, and random variables: Data are denoted by $\{\tilde{y}_t\}_{t=-L}^n$, simulations by $\{\hat{y}_t\}_{t=-L}^N$, and the random variables to which they correspond by $\{y_t\}_{t=-L}^\infty$.

An example is the continuous time version of the stochastic volatility model that has been proposed by Clark (1973), Tauchen and Pitts (1983), and others, as a description of speculative markets. For daily price observations on two securities the model is

$$\left.\begin{aligned} dU_{1t} &= (\rho_1 - \rho_2 U_{1t})dt + \rho_3 dW_{1t} \\ dU_{2t} &= (\rho_4 - \rho_5 U_{2t})dt + \rho_6 \exp(U_{1t})dW_{2t} \\ dU_{3t} &= (\rho_7 - \rho_8 U_{3t})dt + \rho_9 \exp(U_{1t})dW_{3t} \end{aligned}\right\} \quad 0 \le t < \infty$$

$$\left.\begin{aligned} y_{1t} &= U_{2t} \\ y_{2t} &= U_{3t} \end{aligned}\right\} \quad t = 0, 1, ..., n$$

U_{1t} represents an unobserved flow of new information to the market that influences the volatility of asset prices U_{2t} and U_{3t} by changing the instantaneous conditional variances of U_{2t} and U_{3t}. The observed data $\tilde{y}_{1t}$ and $\tilde{y}_{2t}$ are prices at the end of each trading day. To achieve identification in estimation, a normalization rule such as $\rho_9 = 1$ should be imposed. We should remark that one does not have to restrict attention to models where latent variables affect only volatility. They can affect the drift as well; see, for instance, Andersen and Lund (1996a, 1996b).

We assume that U_t, and hence y_t, is stationary and ergodic. We further assume that the stationary distribution of y_t is absolutely continuous. Thus, for each setting of parameter ρ and lag length L, there exists a time invariant density $p(y_{-L}, \ldots, y_0|\rho)$ such that

$$\lim_{N\to\infty} \frac{1}{N} \sum_{t=0}^{N} g(\hat{y}_{t-L}, \ldots, \hat{y}_t) = \int \cdots \int g(y_{-L}, \ldots, y_0)\, p(y_{-L}, \ldots, y_0|\rho)\, dy_{-L} \cdots dy_0$$

where $\{\hat{y}_t\}_{t=-L}^{N}$ is realization of length $N + L + 1$ from the system. This assumes that g is integrable and that either U_0 is a sample from the stationary distribution of U_t or that a longer realization was observed and enough initial observations were discarded for transients to have dissipated. Methods for generating $\{\hat{y}_t\}_{t=-L}^{N}$ are described in Section 4.

Gallant and Tauchen (1996a) proposed an estimator that is applicable in this instance which they termed Efficient Method of Moments (EMM). It would be regarded a minimum chi-square estimator in the statistics literature and a generalized method of moments (GMM) estimator in the econometrics literature.

The method requires a preliminary summary of the data that takes the form of a projection of the data onto a transition density $f(y_t|y_{t-L}, \ldots, y_{t-1}, \theta)$. Projection is accomplished by estimating θ from the data by means of quasi maximum likelihood and using the estimated transition density $f(y_t|y_{t-L}, \ldots, y_{t-1}, \tilde{\theta}_n)$ to represent the data in subsequent computations. The transition density used for data summary is called the auxiliary model. This is very much like a sufficiency reduction of the data and would, in fact, be a sufficiency reduction if the joint density $f(y_{-L}, \ldots, y_0|\theta)$ implied by the auxiliary model encompasses the joint density $p(y_{-L}, \ldots, y_0|\rho)$ implied by the SDE. As the analogy with sufficiency suggests, if the auxiliary model is a good description of the data, then the EMM estimator will have high efficiency (Tauchen, 1996); and if the auxiliary model actually encompasses the family of transition densities implied by (1), then the EMM estimator is fully efficient (Gallant and Tauchen, 1996a). In Section 3 we describe the SNP model proposed by Gallant and Nychka (1987). Gallant and Long (1997) showed that the EMM estimator is asymptotically fully efficient when auxiliary model is SNP. This is the auxiliary model that we use in our analysis of stock returns and interest rates.

The Gallant-Tauchen EMM estimator $\hat{\rho}_n$ is computed as follows. Use the auxiliary model

$$f(y_t|y_{t-L},\ldots,y_{t-1},\theta) \qquad \theta \in R^{p_\theta}$$

and the data $\{\tilde{y}_t\}_{t=-L}^{n}$ to compute the quasi maximum likelihood estimate

$$\tilde{\theta}_n = \underset{\theta\in\Theta}{\operatorname{argmax}}\ \frac{1}{n}\sum_{t=0}^{n}\log[f(\tilde{y}_t|\tilde{y}_{t-L},\ldots,\tilde{y}_{t-1},\theta)]$$

and the corresponding estimate of the information matrix

$$\tilde{\mathcal{I}}_n = \frac{1}{n}\sum_{t=0}^{n}\Big[\frac{\partial}{\partial\theta}\log f(\tilde{y}_t|\tilde{x}_{t-1}\tilde{\theta}_n)\Big]\Big[\frac{\partial}{\partial\theta}\log f(\tilde{y}_t|\tilde{x}_{t-1},\tilde{\theta}_n)\Big]',$$

where

$$\tilde{x}_{t-1} = (\tilde{y}_{t-L},\ldots,\tilde{y}_{t-1}).$$

Define

$$m(\rho,\theta) = \int\cdots\int\frac{\partial}{\partial\theta}\log[f(y_0|y_{-L},\ldots,y_{-1},\theta)]\,p(y_{-L},\ldots,y_0|\rho)\,dy_{-L}\cdots dy_0$$

which is computed by averaging over a long simulation

$$m(\rho,\theta) \doteq \frac{1}{N}\sum_{t=0}^{N}\frac{\partial}{\partial\theta}\log[f(\hat{y}_t|\hat{y}_{t-L},\ldots,\hat{y}_{t-1},\theta)]$$

as described above. The estimator is

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

The asymptotics of the estimator, which are derived in Gallant and Tauchen (1996a), are as follows. If ρ^o denotes the true value of the parameters in the system

$$dU_t = A(U_t,\rho)dt + B(U_t,\rho)dW_t,$$

and θ^o is an isolated solution of the moment equations $m(\rho^o, \theta) = 0$, then

$$\lim_{n\to\infty} \hat{\rho}_n = \rho^o \quad \text{a.s.}$$

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

$$\lim_{n\to\infty} \hat{M}_n = M^o \quad \text{a.s.}$$

$$\lim_{n\to\infty} \tilde{\mathcal{I}}_n = \mathcal{I}^o \quad \text{a.s.}$$

where $\hat{M}_n = M(\hat{\rho}_n, \tilde{\theta}_n)$, $M^o = M(\rho^o, \theta^o)$, $M(\rho,\theta) = (\partial/\partial\rho')m(\rho,\theta)$, and

$$\mathcal{I}^o = \int\cdots\int \Big[\frac{\partial}{\partial\theta}\log f(y_0|x_{-1},\theta^o)\Big]\Big[\frac{\partial}{\partial\theta}\log f(y_0|x_{-1},\theta^o)\Big]' p(y_{-L},\ldots,y_0|\rho^o)\, dy_{-L}\cdots dy_0.$$

Under the null hypothesis that

$$dU_t = A(U_t,\rho)dt + B(U_t,\rho)dW_t,$$

is the correct model,

$$C = n\, m'(\hat{\rho}_n, \tilde{\theta}_n)(\tilde{\mathcal{I}}_n)^{-1}m(\hat{\rho}_n, \tilde{\theta}_n)$$

is asymptotically chi-squared on $p_\theta - p_\rho$ degrees freedom. Under the null hypothesis $H: h(\rho^o) = 0$, where h maps $\Re^{p_\rho}$ into $\Re^q$,

$$L = n\Big[m'(\hat{\hat{\rho}}_n, \tilde{\theta}_n)(\tilde{\mathcal{I}}_n)^{-1}m(\hat{\hat{\rho}}_n, \tilde{\theta}_n) - m'(\hat{\rho}_n, \tilde{\theta}_n)(\tilde{\mathcal{I}}_n)^{-1}m(\hat{\rho}_n, \tilde{\theta}_n)\Big]$$

is asymptotically chi-squared on q degrees freedom where

$$\hat{\hat{\rho}}_n = \underset{h(\rho)=0}{\operatorname{argmin}}\, m'(\rho, \tilde{\theta}_n)(\tilde{\mathcal{I}}_n)^{-1}m(\rho, \tilde{\theta}_n).$$

A Wald confidence interval on an element ρ_i of ρ can by constructed in the usual way from an asymptotic standard error $\sqrt{\hat{\sigma}_{ii}}$. A standard error may be obtained by computing the Jacobian $M_n(\rho,\theta)$ numerically and taking the estimated asymptotic variance $\hat{\sigma}_{ii}$ to be the i-th diagonal element of $\hat{\Sigma} = (1/n)[(\hat{M}_n)'(\tilde{\mathcal{I}}_n)^{-1}(\hat{M}_n)]^{-1}$. These intervals, which are symmetric, are somewhat misleading because they do not reflect the rapid increase in the EMM objective function $s_n(\rho) = m'(\rho, \tilde{\theta}_n)(\tilde{\mathcal{I}}_n)^{-1}m(\rho, \tilde{\theta}_n)$ when ρ_i approaches a value for which simulations from $dU_t = A(U_t,\rho)dt + B(U_t,\rho)dW_t$ become explosive. Confidence intervals obtained by inverting the criterion difference test L do reflect this phenomenon and are therefore more useful. To invert the test one puts in the interval those ρ_i^* for which L for the hypothesis $\rho_i^o = \rho_i^*$ is less than the critical point of a chi-squared random variable on one degree freedom. To avoid re-optimization one may use the approximation

$$\hat{\hat{\rho}}_n = \hat{\rho}_n + \frac{\rho_i^* - \hat{\rho}_{in}}{\hat{\sigma}_{ii}} \hat{\Sigma}_{(i)}$$

in the formula for L where $\hat{\Sigma}_{(i)}$ is the i-th column of $\hat{\Sigma}$.

When $C = n\, m'(\hat{\rho}_n, \tilde{\theta}_n)(\tilde{\mathcal{I}}_n)^{-1} m(\hat{\rho}_n, \tilde{\theta}_n)$ exceeds the chi-squared critical point one would like some suggestions as to what is wrong. Because

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

inspection of the t-ratios

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

where $S_n = \Big(\mathrm{diag}\{\tilde{\mathcal{I}}_n - (\hat{M}_n)[(\hat{M}_n)'(\tilde{\mathcal{I}}_n)^{-1}(\hat{M}_n)]^{-1}(\hat{M}_n)'\}\Big)^{1/2}$, can suggest reasons for failure. Different elements of the score correspond to different characteristics of the data. Large t-ratios reveal the characteristics that are not well approximated. For this purpose, the quasi-t-ratios $\tilde{T}_n = [(\mathrm{diag}\, \tilde{\mathcal{I}}_n)^{1/2}]^{-1} \sqrt{n}\, m(\hat{\rho}_n, \tilde{\theta}_n)$, which are under-estimates, are usually adequate and are cheaper to compute because they avoid numerical approximation to $\hat{M}_n$.

Regularity conditions sufficient to imply the results above and identification issues are discussed in Gallant and Long (1997) and references therein. Our experience has been that a reliable indicator of a lack of identification is numerical instability due to poor conditioning of the Hessian approximation when using a good quality quasi-Newton optimization algorithm. In our experience, the cause of identification problems has usually been an auxiliary model that is not rich enough to completely represent the dynamics rather than problems with the specification of the continuous time system. For example, a continuous time stochastic volatility model such as that above cannot be estimated from an auxiliary model that does not have an ARCH-like component. If the auxiliary model is mean reverting, then this forces mean reversion on the estimated continuous time system (Tauchen, 1997). Thus, ergodicity is essentially forced upon the estimated continuous time system by the EMM methodology.

Fortran code and a User's Guide in PostScript that implement the EMM estimator using $f(y_t|x_{t-1}, \theta)$ as described in Section 3 as the auxiliary model are available by anonymous ftp at site ftp.econ.duke.edu in directory pub/get/emm.

3 The Auxiliary Model

Put $y = y_0$, $x = x_{-1} = (y_{-L}, \ldots, y_{-1})$, $\ell = M(L+1)$, and $p(x, y|\rho) = p(y_{-L}, \ldots, y_0|\rho)$, $p(x|\rho) = \int p(y_{-L}, \ldots, y_0|\rho)\, dy_0$, and $p(y|x, \rho) = p(x, y|\rho)/p(x|\rho)$.

Under regularity conditions, Gallant and Nychka (1987) prove that $p(x, y|\rho^o)$ has the representation

$$\begin{aligned} p(x, y|\rho^o) &= [g(x, y|\rho)]^2 \phi(x, y) + \epsilon_0 \phi(x, y) \\ g(v|\rho) &= \sum_{|\lambda| < \infty} a_\lambda v^\lambda. \end{aligned}$$

where $\lambda = (\lambda_1, \ldots, \lambda_\ell)$ is a nonnegative multi-index (a vector of nonnegative integers), $|\lambda| = \sum_{k=1}^{\ell} |\lambda_k|$,

$$v^\lambda = v_1^{\lambda_1} \cdots v_\ell^{\lambda_\ell},$$

$\phi(x, y) = (2\pi)^{-\ell/2} e^{-x'x/2 - y'y/2}$, and ϵ_0 is a small positive number. Under regularity conditions, Gallant and Long (1997) prove that if the truncated expansion

$$\begin{aligned} f_K(x, y|\theta) &= [g_K(x, y|\theta)]^2 \phi(x, y) + \epsilon_0 \phi(x, y) \\ g_K(v|\theta) &= \sum_{|\lambda| < K} \bar{a}_\lambda v^\lambda, \end{aligned}$$

is used as the auxiliary model

$$f_K(y|x, \theta) = \frac{f_K(x, y|\theta)}{\int f_K(x, y|\theta) dy}$$

of the minimum chi-square estimator $\hat{\rho}_n$ of Section 2, then the efficiency of $\hat{\rho}_n$ can be made as close as desired to that of the maximum likelihood estimator by taking the truncation point K and lag length L suitably large. Here $\theta = \{\bar{a}_\lambda : 1 \leq |\lambda| \leq K\}$ and $\bar{a}_\lambda$ denotes the leading coefficients $\{a_\lambda : 1 \leq |\lambda| \leq K\}$ of the infinite expansion renormalized so that $f_K(x, y|\theta)$ integrates to one.

In applications, certain algebraic modifications facilitate computations and improve finite sample performance. The term $\epsilon_0 \phi(v)$ in the truncated expansion $[g_K(v|\theta)]^2 \phi(v) + \epsilon_0 \phi(v)$ may be deleted provided that those $\log g_K(y_t|x_{t-1}, \theta)$ that become too small to have machine representation during an optimization of the log likelihood are put to the smallest number that the machine can represent. In tests, optimizations using this strategy have been more efficient and stable than methods that retained the term $\epsilon_0 \phi(v)$. With the term deleted, a change of location and scale prior to conditioning, and some rearrangement of polynomial coefficients, the conditional density can be put in the form (Gallant, Hsieh, and Tauchen, 1991)

$$h_K(y_t|x_{t-1},\theta) = \frac{\{P[R^{-1}(y-\mu_{x_{t-1}}),x_{t-1}]\}^2 \phi[R^{-1}(y-\mu_{x_{t-1}})]}{|\det(R)|^{1/2} \int [P(z,x_{t-1})]^2 \phi(z)\,dz}$$

where

$$\mu_{x_{t-1}} = b_0 + B x_{t-1}$$

$$P(z,x) = \sum_{\alpha=0}^{K_z} \left(\sum_{\beta=0}^{K_x} a_{\beta\alpha} x^\beta \right) z^\alpha$$

and R is an upper triangular matrix. We permit $P(z,x)$ to depend on $L_p \leq L$ lags and $\mu_{x_{t-1}}$ to depend on $L_u \leq L$ lags, which means some of the coefficients of $P(z,x)$ and of μ_x may be zero. We refer to μ_x as the location function and to $P^2(z,x)\phi(z)$ as the Hermite polynomial.

The model's ability to approximate conditionally heteroskedastic data is much improved if the Hermite polynomial is relieved of the task. This may be done by replacing R above by $R_{x_{t-1}}$ with

$$\operatorname{vech}(R_{x_{t-1}}) = \rho_0 + P|e_{t-1}|$$

where $\operatorname{vech}(R)$ denotes a vector of length $M(M+1)/2$ containing the elements of the upper triangle of R, P is a $M(M+1)/2 \times L_r$ matrix,

$$e_{t-1} = \left[(y_{t-L_r} - \mu_{x_{t-L_r-1}}), \ldots, (y_{t-1} - \mu_{x_{t-2}}) \right]',$$

$L_r + L_u \leq L$, and $|e|$ denotes elementwise absolute value.

The model with $R_{x_{t-1}}$ substituted for R is the auxiliary model used in applications

$$f(y_t|x_{t-1},\theta) = \frac{\{P[R_{x_{t-1}}^{-1}(y_t-\mu_{x_{t-1}}),x_{t-1}]\}^2 \phi[R_{x_{t-1}}^{-1}(y_t-\mu_{x_{t-1}})]}{|\det(R_{x_{t-1}})|^{1/2} \int [P(z,x_{t-1})]^2 \phi(z)\,dz}.$$

We refer to R_x as the scale function and to $f(y_t|x_{t-1},\theta)$ as the SNP density.

The vector θ of the SNP density contains the coefficients $A = [a_{\beta\alpha}]$ of the Hermite polynomial, the coefficients $[b_0, B]$ of the location function, and the coefficients $[\rho_0, P]$ of the scale function. To achieve identification, the coefficient $a_{0,0}$ is set to 1.

When M is large, coefficients $a_{\beta\alpha}$ corresponding to monomials z^α that represent high order interactions can be set to zero with little effect on the adequacy of approximations.

Let $I_z = 0$ indicate that no interaction coefficients are set to zero, $I_z = 1$ indicate that coefficients corresponding to interactions z^α of order larger than $K_z - 1$ are set to zero, and so on; similarly for L_p, x^β, and I_x.

We also find that setting the elements of P to zero that correspond to the off-diagonal elements of R_x can improve the stability of optimizations with little effect on the adequacy of approximations.

Lastly, outliers y_{t-j} entering the variance function R_x can be destabilizing; to some extent this is true for $P(z, x)$ and μ_x as well. Stability can be improved without affecting theoretical results by replacing $x_{i,t-j}$ with $\hat{x}_{i,t-j} = 8\exp[2(x_{i,t-j} - \bar{y}_i)/s_i]/\{1+\exp[2(x_{i,t-j} - \bar{y}_i)/s_i]\} - 4$ in the formulas for μ_x, R_x, and $P(z, x)$, where $\bar{y}_i$ denotes the sample mean of the data, i.e., $\bar{y}_i = [1/(n + L + 1)] \sum_{t=-L}^{n} \tilde{y}_{it}$ for $i = 1, \ldots, M$, and s_i denotes the sample standard deviation. We find this modified logistic transformation to be beneficial in applications involving strongly mean reverting data. However, with persistent data, we find that a sensitivity to outliers seems to help keep ρ out of explosive regions during EMM optimizations so that using the logistic transform with persistent data is ill-advised. The currently distributed SNP code contains a spline transformation to be used instead of the logistic with persistent data that seems to work well. However, that transform was developed after the work reported here.

Some structural characteristics of $f(y_t|x_{t-1}, \theta)$ might be noted. If K_z, K_x, and L_r and are put to zero then $f(y_t|x_{t-1}, \theta)$ is a Gaussian vector autoregression. If K_x and L_r are put to zero then $f(y_t|x_{t-1}, \theta)$ is a non-Gaussian vector autoregression with homogeneous innovations. If K_z and K_x are put to zero then $f(y_t|x_{t-1}, \theta)$ is a Gaussian ARCH model. If K_x is put to zero then $f(y_t|x_{t-1}, \theta)$ is a non-Gaussian ARCH model with homogeneous innovations. If $K_z > 0$, $K_x > 0$, $L_p > 0$, $L_\mu > 0$, and $L_r > 0$ then $f(y_t|x_{t-1}, \theta)$ is a general nonlinear process with heterogeneous innovations.

Fortran code, a User's Guide in PostScript, and PC executables for maximum likelihood estimation of the parameters of the SNP density are available by anonymous ftp at site ftp.econ.duke.edu in directory pub/arg/snp.

4 Simulations from Stochastic Differential Equations

A simulation $\{\hat{y}_t\}_{t=-L}^{N}$ for computing

$$m(\rho, \theta) \doteq \frac{1}{N} \sum_{t=0}^{N} \frac{\partial}{\partial \theta} \log[f(\hat{y}_t | \hat{y}_{t-L}, \ldots, \hat{y}_{t-1}, \theta)]$$

may be obtained by first generating the sequence

$$\hat{U}_0,\ \hat{U}_\Delta,\ \hat{U}_{2\Delta},\ \hat{U}_{3\Delta},\ \ldots,\ \hat{U}_{N+L-\Delta},\ \hat{U}_{N+L}$$

with step-size $\Delta = 1/n_0$ from the system

$$dU_t = A(U_t, \rho)dt + B(U_t, \rho)dW_t \qquad 0 \leq t < N + L$$

by means of an explicit order 2 weak scheme and then transforming and retaining every n_0 th element to get

$$\hat{y}_{t-L} = T(\hat{U}_{\Delta(n_0 t)}) = T(\hat{U}_t) \qquad t = 0,\ 1,\ \ldots,\ N + L.$$

A simulation $\{\hat{y}_t\}_{t=-L}^{N}$ for graphical display or for computation of statistics such as density estimates may be generated similarly using an explicit order 1 strong scheme. For the results in Subsection 5.4, time t is in days and $n_0 = 24$, which implies that the simulation step-size, $\Delta = 1/24$, is one hour.

The formulae for the two schemes used here are given in Gallant and Tauchen (1996b) and are available by anonymous ftp from ftp.econ.duke.edu in directory pub/arg/libf as files weak2.f and stng1.f; also at the same site is weak2s.f, which is a stopped-process variant of weak2.f that is useful in connection with EMM estimation with compilers that do not handle NaN's according to IEEE standards. These routines represent certain choices of tuning parameters on our part to schemes from Kloeden and Platen (1992, p. 347, p. 376, p. 486). The weak scheme is valid for autonomous systems $dU_t = A(U_t, \rho)dt + B(U_t, \rho)dW_t$ only. The strong scheme is valid for non-autonomous systems $dU_t = A(t, U_t, \rho)dt + B(t, U_t, \rho)dW_t$ as well. Discussions of the numerical properties of these schemes and some experience with them are in Gallant and Long (1997).

A reader has remarked that since we are averaging over the stationary distribution, there is no need to discard the intermediate sample points. That is, for the example above, we could have averaged over the hourly points rather than the daily. In principle the reader is correct, and we intend future experiments to determine if retaining the intermediate sampled points leads to improved numerical accuracy without inducing undesirable side effects.

5 Stock Prices

5.1 Data and Auxiliary Models

The data is a long time series comprised of 16,127 daily observations on adjusted movements of the Standard and Poor's Composite Price Index, 1928–87. This series is exactly the same as the univariate stock series used in Gallant, Rossi, and Tauchen (1992). As described in Gallant, Rossi, and Tauchen (1992), the raw series $\{S_t\}_{t=0}^{16,127}$ is converted to a price movements series $\{\Delta S_t = 100[\log(S_t) - \log(S_{t-1})]\}_{t=1}^{16,127}$ and then is adjusted for systematic calendar effects in location and scale to obtain the series $\{\tilde{y}_t\}_{t=1-L}^{16,127-L}$ for analysis. 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 be accounted for in order not to adversely affect analysis.

We estimated the parameters θ of the SNP auxiliary model $f(y|x,\theta)$ from these data by quasi maximum likelihood, selecting the values $L_u = 2$, $L_r = 18$, $L_p = 2$, $K_z = 4$, $I_z = 0$, $K_x = 1$, and $I_x = 0$ for the tuning parameters as in Gallant, Rossi, and Tauchen (1992). (Older code forced the constraint $L_u = L_r$ in their work which has been relaxed here.) This specification has the following functional form.

Nonlinear Nonparametric Score (NN)

$$
\begin{aligned}
f(y|x,\theta) &= \frac{\{P[(y-\mu_x)/r_x, x]\}^2 \phi[(y-\mu_x)/r_x]}{|r_x|^{1/2} \int [P(v,x)]^2 \phi(v)\, dv} \\
P(v,x) &= \sum_{\alpha=0}^{4} \left[\sum_{\beta_1=0}^{1} \sum_{\beta_2=0}^{1} a_{\beta\alpha} (y_{-2})^{\beta_2} (y_{-1})^{\beta_1} \right] v^{\alpha} \\
\mu_x &= b_0 + b_1 y_{-1} + b_2 y_{-2} \\
r_x &= r_0 + \sum_{j=1}^{18} r_j |y_{-j} - b_0 - b_1 y_{-j-1} - b_2 y_{-j-2}| \\
x &= (y_{-20}, y_{-19}, \ldots, y_{-1}).
\end{aligned}
$$

If the logistic transformation is applied to x as described in Section 3 then the model is denoted as NN-L.

If we impose $K_x = 0$ then the SNP density is a form of ARCH model with conditionally homogeneous, non-Gaussian innovations that is similar to the semiparametric ARCH model considered by Engle and Gonzales-Rivera (1991). We term this variant the

Semiparametric ARCH Score (SA)

$$
\begin{aligned}
f(y|x,\theta) &= \frac{\{P[(y-\mu_x)/r_x]\}^2 \phi[(y-\mu_x)/r_x]}{|r_x|^{1/2} \int [P(v)]^2 \phi(v)\, dv} \\
P(v) &= \sum_{\alpha=0}^{4} a_\alpha v^\alpha \\
\mu_x &= b_0 + b_1 y_{-1} + b_2 y_{-2} \\
r_x &= r_0 + \sum_{j=1}^{18} r_j |y_{-j} - b_0 - b_1 y_{-j-1} - b_2 y_{-j-2}| \\
x &= (y_{-20}, y_{-19}, \ldots, y_{-1}).
\end{aligned}
$$

If the logistic transformation is applied to x then the model is denoted as SA-L.

5.2 Discrete Time Stochastic Volatility Models

Gallant, Hsieh, and Tauchen (1997) fit a large number of the discrete time stochastic volatility models taken from the literature using the EMM method with both the Semiparametric ARCH Score and the Nonlinear Nonparametric Score. Their work is summarized in two tables reproduced here as Tables 1 and 2.

Table 1 about here

Table 2 about here

Table 1 shows the optimized values of the EMM objective function scaled to follow the chi-squared distribution, as described in Section 2. From the top panel of the table it is seen that the standard stochastic volatility model with Gaussian errors

Gaussian

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

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. These variants are as follows.

t-Errors

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

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 the value of the objective function is successively fixed at $\nu = 10, 15, 20, 25$.

Spline

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

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 $(2\pi)^{-1/2} \int v T_z(v) \exp(-v^2/2) dv = 0$ and $(2\pi)^{-1/2} \int v^2 T_z(v) \exp(-v^2/2) dv = 1$ are imposed on the b_{zj}.

Bollerslev and Mikkelsen (1996), Ding, Granger, and Engle (1993), and Breidt, Crato, 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. The long-memory stochastic volatility model is

Long Memory

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

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 $1/2 \le d < 1$, $(1-\mathcal{L})^d v_t = \epsilon_t$, 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.

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

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

where g is a free parameter to be estimated. The model that results is

Asymmetric

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

Table 2 displays the objective function surface for versions of the stochastic volatility model against the Nonlinear Nonparametric Score. The standard model is overwhelmingly rejected. The various extensions provide much improvement over the standard Gaussian model, but nothing comes as close as did the spline variants against the Semiparametric ARCH Score.

5.3 Continuous-Time Models

All the continuous time models that were estimated from stock prices may be obtained from the following general model by putting certain parameters to zero or one.

General Model

$$
\begin{aligned}
dH_t &= (a_h + a_{hh} H_t + a_{hu} U_t)dt + (b_{h1} + b_{hh1} H_t)dW_{1t} + (b_{h2})dW_{2t} && 0 \le t < \infty \\
dU_t &= (a_{uu} U_t + a_{uh} H_t)dt \\
dX_t &= (a_x + a_{xx} X_t + a_{xv} V_t)dt + (b_{x2} + b_{xx2}|X_t|^{\gamma} + b_{xh2} e^{H_t})dW_{2t} \\
dV_t &= (a_{vv} V_t + a_{vx} X_t)dt \\
y_{t-L} &= T(X_t, b_2, b_3) && t = 0, 1, \ldots
\end{aligned}
$$

where $T(z, b_2, b_3) = b_0 + b_1 z + b_2 z^{2+\delta} I_{z \ge 0}(z) + b_3(-z)^{2+\delta} I_{z<0}(z)$ with b_0 and b_1 determined by $\int T(z, b_2, b_3)\phi(z)\,dz = 0$ and $\int T^2(z, b_2, b_3)\phi(z)\,dz = 1$.

The stock price data is in terms of price movements from yesterday's close to today's. This is for two reasons. The first is that using price movement data permits direct comparison with the results of Gallant, Hsieh, and Tauchen (1997) and Gallant, Rossi, and Tauchen (1992, 1993). The second is that the critical assumptions of stationarity and ergodicity are questionable for stock market data in either levels or logs. Modification to the EMM theory and methodology would be required if these assumptions are violated.

Within-day price movement may be obtained from a continuous time model such as above using $\Delta S_{t+s} = y_{t+s} - h_{ts}$, where $\Delta S_{t+s} = \log(S_{t+s}) - \log(S_t)$, $0 < s \le 1$, $t = 0, 1, \ldots,$ and h_{ts} is a conditionally deterministic function such that $h_{t0} = y_t$ and $h_{t1} = 0$. Two examples for h_{ts} are $h_{ts} = S_t - S_{t-1+s}$ and $h_{ts} = (1 - s)y_t$. Any choice of h_{ts} is an assumption as to how prices evolve during the day, and $h_{t0} = y_t$ is the assumption that prices do not change overnight. If within day price data were available then overnight price changes could be permitted and the functional form of h_{ts} identified. To obtain level data at time $t + s$ use

$\log(S_{t+s}) = \log(S_{-1}) + (\sum_{j=0}^{t} y_j)/100 + (\Delta S_{t+s})/100$ with $0 < s \leq 1$, $t = 0, 1, \ldots$. For some purposes it may also be necessary to apply the inverse of the calendar effects adjustment that was applied to the data to the series $\{y_j\}_{j=0}^{t}$ prior to computing the sum $\sum_{j=0}^{t} y_j$.

The models considered may be thought of as variants on a model with first order drift and constant diffusion

FD

$$\begin{aligned} dX_t &= (a_x + a_{xx} X_t)dt + (b_{x2})dW_{2t} \\ y_{t-L} &= X_t \qquad\qquad t = 0, 1, \ldots \end{aligned}$$

or on a model with second order drift and constant diffusion

SD

$$\begin{aligned} dX_t &= (a_x + V_t)dt + (b_{x2})dW_{2t} \\ dV_t &= (a_{vv} V_t + a_{vx} X_t)dt \\ y_{t-L} &= X_t \qquad\qquad t = 0, 1, \ldots \end{aligned}$$

Note that the forcing terms are applied to the first equation (velocity) rather than the second (acceleration) as one might expect from the control literature. This is done to preserve the Brownian motion flavor of sample paths – continuous and nondifferentiable – of the first order model. If forcing were added to the second equation rather than the first, then sample paths would be continuous and differentiable.

As we shall see later, both a second order drift and some transformation to X_t are required in order to match the moments implied by the Nonlinear Nonparametric Score. Accordingly, we consider second order drift models with spline-transformed output hereafter.

$SD - S\delta$

$$\begin{aligned} dX_t &= (a_x + V_t)dt + (b_{x2})dW_{2t} \\ dV_t &= (a_{vv} V_t + a_{vx} X_t)dt \\ y_{t-L} &= T(X_t, b_2, b_3) \qquad\qquad t = 0, 1, \ldots \end{aligned}$$

where $T(z, b_2, b_3) = b_0 + b_1 z + b_2 z^{2+\delta} I_{z \geq 0}(z) + b_3(-z)^{2+\delta} I_{z<0}(z)$ with b_0 and b_1 determined by $\int T(z, b_2, b_3)\phi(z)\, dz = 0$ and $\int T^2(z, b_2, b_3)\phi(z)\, dz = 1$.

The diffusion can be modified by adding an term that will induce conditional heteroskedasticity

$SD - S\delta - CH$

$$
\begin{aligned}
dX_t &= (a_x + V_t)dt + (b_{x2} + b_{xx2}|X_t|^{\gamma})dW_{2t} \\
dV_t &= (a_{vv}V_t + a_{vx}X_t)dt \\
y_{t-L} &= T(X_t, b_2, b_3) \qquad\qquad t = 0, 1, \ldots
\end{aligned}
$$

Allowing powers γ other than $\gamma = 1$ permits consideration such models as the square root model $\gamma = 1/2$, which is popular in the finance literature.

Alternatively, conditional heteroskedasticity can be induced by means of either first order stochastic volatility

$SD - S\delta - FSV$

$$
\begin{aligned}
dH_t &= (a_h + a_{hh}H_t)dt + (b_{h1} + b_{hh1}H_t)dW_{1t} \qquad 0 \le t < \infty \\
dX_t &= (a_x + V_t)dt + (b_{x2} + e^{H_t})dW_{2t} \\
dV_t &= (a_{vv}V_t + a_{vx}X_t)dt \\
y_{t-L} &= T(X_t, b_2, b_3) \qquad\qquad t = 0, 1, \ldots
\end{aligned}
$$

or second order stochastic volatility

$SD - S\delta - SSV$

$$
\begin{aligned}
dH_t &= (a_h + U_t)dt + (b_{h1} + b_{hh1}H_t)dW_{1t} \qquad 0 \le t < \infty \\
dU_t &= (a_{uu}U_t + a_{uh}H_t)dt \\
dX_t &= (a_x + V_t)dt + (b_{x2} + e^{H_t})dW_{2t} \\
dV_t &= (a_{vv}V_t + a_{vx}X_t)dt \\
y_{t-L} &= T(X_t, b_2, b_3) \qquad\qquad t = 0, 1, \ldots
\end{aligned}
$$

Lastly, we can induce a correlation between the volatility and output along the lines of the asymmetric discrete-time stochastic volatility model described in Subsection 5.2

$SD - S\delta - FSV - A$

$$
\begin{aligned}
dH_t &= (a_h + a_{hh}H_t)dt + (b_{h1} + b_{hh1}H_t)dW_{1t} + (b_{h2})dW_{2t} \qquad 0 \le t < \infty \\
dX_t &= (a_x + V_t)dt + (b_{x2} + e^{H_t})dW_{2t} \\
dV_t &= (a_{vv}V_t + a_{vx}X_t)dt \\
y_{t-L} &= T(X_t, b_2, b_3) \qquad\qquad t = 0, 1, \ldots
\end{aligned}
$$

5.4 Continuous-Time Estimation

The first specifications that we considered were a few variations on the first order drift class of models, FD. Using the procedure described in Section 2 we obtained the chi-squared values shown in the first block of Table 3; Tables 4 and 5 show the values of the drift and diffusion parameters corresponding to these fits. All FD specifications are sharply rejected.

Table 3 about here

Figure 1 about here

Table 4 about here

Table 5 about here

The quasi-t-ratios displayed as a bar chart in the upper panel of Figure 1 suggest why first order drift class of models are rejected. The large quasi-t-ratio b_2 in the location function implies that first order drift is not adequate: matching the moments of a discrete-time second order autoregression will require second order drift in continuous time. The large quasi-t-ratios for the Hermite polynomial suggest the need of a transformation to the output X_t. The quantile-quantile plot to the right of the bar chart in the upper panel of Figure 1 implies that this transformation should have the effect of thickening the tails by decreasing the magnitude of observations to the left of the median and increasing those to the right. The spline transform $T(z, b_2, b_3) = b_0 + b_1 z + b_2 z^{2+\delta} I_{z \geq 0}(z) + b_3 (-z)^{2+\delta} I_{z<0}(z)$ with b_0 and

b_1 determined by $\int T(z, b_2, b_3)\phi(z)\,dz = 0$ and $\int T^2(z, b_2, b_3)\phi(z)\,dz = 1$ accomplishes this. If $\delta > 0$ the transform is twice continuously differentiable which allows application of Ito's lemma in order to determine a final expression for the model as a stochastic differential equation.

We next tried a second order drift with second order stochastic volatility model with spline transform and asymmetry, $SD - S\delta - SSV - A$. Using the procedure described in Section 2, we obtained the chi-squared values shown in the second block of Table 3. From Tables 3 and 5 it appears that increasing δ would produce a satisfactory fit, but during the course of the optimizations it became apparent that the second order stochastic volatility is not needed and that first order would suffice. We therefore continued as shown in the last block of Table 3. The value $\delta = 1.4$ produces satisfactory fits as seen from the last panels of Tables 3 and 5 and last panel of Figure 1.

Our preferred model is, therefore,

Preferred Model

$$
\begin{aligned}
dH_t &= (a_h + a_{hh}H_t)dt + (b_{h1} + b_{hh1}H_t)dW_{1t} + (b_{h2})dW_{2t} && 0 \le t < \infty \\
dX_t &= (a_x + V_t)dt + (b_{x2} + b_{xx2}|X_t| + e^{H_t})dW_{2t} \\
dV_t &= (a_{vv}V_t + a_{vx}X_t)dt \\
y_{t-L} &= T(X_t, b_2, b_3) && t = 0, 1, \ldots
\end{aligned}
$$

where $T(z, b_2, b_3) = b_0 + b_1 z + b_2 z^{3.4} I_{z \ge 0}(z) + b_3(-z)^{3.4} I_{z<0}(z)$ with b_0 and b_1 determined by $\int T(z, b_2, b_3)\phi(z)\,dz = 0$ and $\int T^2(z, b_2, b_3)\phi(z)\,dz = 1$. Upon application of Ito's lemma, we obtain a representation of our preferred model as a stochastic differential equation in four state variables with one component observed.

$$
\begin{aligned}
dH_t &= (a_h + a_{hh}H_t)dt + (b_{h1} + b_{hh1}H_t)dW_{1t} + (b_{h2})dW_{2t} && 0 \le t < \infty \\
dX_t &= (a_x + V_t)dt + (b_{x2} + b_{xx2}|X_t| + e^{H_t})dW_{2t} \\
dV_t &= (a_{vv}V_t + a_{vx}X_t)dt \\
dZ_t &= [(a_x + V_t)T'(X_t, b_2, b_3) + \tfrac{1}{2}(b_{x2} + b_{xx2}|X_t| + e^{H_t})T''(X_t, b_2, b_3)]dt \\
&\quad + (b_{x2} + b_{xx2}|X_t| + e^{H_t})T'(X_t, b_2, b_3)dW_{2t} \\
y_{t-L} &= Z_t && t = 0, 1, \ldots
\end{aligned}
$$

Table 6 presents Wald and criterion difference 95% confidence intervals for the parameters of the preferred fit.

———————————————

Table 6 about here

———————————————

A reader has remarked that the quasi-t-ratios shown in Figure 1 have not been orthogonalized so that, due to multicolinearity, there is no guarantee that a modification to the model suggested by the largest quasi-t-ratio is necessarily the best possible modification one might make. The reader is correct in this, some other modification might be better and perhaps orthogonalization, which is numerically feasible, might be beneficial. Nonetheless, our experience to date has been that a model augmentation strategy as outlined above does lead us to improved fits. Moreover, even if the quasi-t-ratios turn out to be misleading, the diagnostics shown in Table 3 protect one from actual error. The worst that can happen is that some time is wasted in fitting a models that are eventually rejected.

6 The Term Structure

6.1 Term Structure Data and the Auxiliary Model

The data are 1,735 weekly observations, January 5, 1962 – March 31, 1995, on three interest rates: the three month Treasury Bill rate from the secondary market (TBILL03M), the twelve month Treasury Bill rate from the secondary market (TBILL12M), and a ten year constant maturity Treasury Bond rate (TBOND10Y). Friday rates are used except when unavailable due to a holiday, in which case the Thursday rate is used. We put $\tilde{y}_t = (\text{TBILL03M}_t\ \text{TBILL12M}_t\ \text{TBOND10Y}_t)' \in \Re^3$ as the observed variable. Figure 2 shows plots of the three series.

———————————————

Figure 2 about here

———————————————

For estimation of the auxiliary model, we reserve the observations from the first 26 weeks for forming lags, giving the data set $\{\tilde{y}_t\}_{t=-26}^{1709}$. Using this data set we estimated by maximum likelihood the parameters θ of the auxiliary SNP model $f(y|x,\theta)$. For initial testing, we restrict $K_x = 0$. Following the protocol of Bansal, Gallant, Hussey, and Tauchen (1995), we find $L_u = 3$, $L_r = 4$, $K_z = 4$, $I_z = 3$ for the tuning parameters. This model takes the form of an ARCH model with conditionally homogeneous non-Gaussian errors. The error density is modified Hermite density as discussed in Section 3 above. The polynomial $P(v)$ is a quartic in $v \in \Re^3$ with all interactions suppressed. We call the score from this fit the *Semiparametric ARCH Score*, which is of length $p_\theta = 60$.

We use this auxiliary model for EMM estimation of various continuous time models of the term structure. We emphasize that EMM applied to a score from an SNP model constrained so that $K_x = 0$ is consistent and asymptotically normal, so long as the underlying continuous time model is correctly specified (Gallant and Tauchen, 1996a), albeit with a possible efficiency loss (Gallant and Long, 1997). Some caveats are in order, though. Failure to fit this score can, of course, be construed as sharp evidence against the continuous time model. However, for reasons discussed in Tauchen (1996), successful fitting of this score does not necessarily signal correct specification of the continuous time model.

6.2 The Yield-Factor Model and Extensions

Let X_t denote a vector of fundamental economic factors. For reasons that will become clear as we move along, we take $X_t = (X_{1t}\ X_{2t}\ X_{3t})'$ as 3-dimensional, though almost everything holds for X_t of arbitrary dimension. Duffie and Kan (1993) present the Yield-Factor Model for the dynamics of a vector of fundamental factors. The most general version we write as *YF-General*

$$\begin{aligned} dX_t &= (\alpha_0 + \alpha_1 X_t)dt + \\ & \beta_0 \begin{bmatrix} (\beta_{11} + \beta_{2,[1]}X_t)^{1/2} & 0 & 0 \\ 0 & (\beta_{12} + \beta_{2,[2]}X_t)^{1/2} & 0 \\ 0 & 0 & (\beta_{13} + \beta_{2,[3]}X_t)^{1/2} \end{bmatrix} dW_t \end{aligned}$$

where α_0 is 3×1, α_1 is 3×3, β_0 is 3×3, $\beta_1 = (\beta_{11}\ \beta_{12}\ \beta_{13})'$ is 3×1, β_2 is 3×3 with $\beta_{2,[i]}$ denoting the i^{th} row of β_2, and W_t is a 3×1 vector of independent Wiener processes. The *YF-General* model is a multivariate generalization of the square root model of Cox, Ingersoll, and Ross (1985). A compact way to write the *YF-General* model is to define

$v(X_t) = \beta_1 + \beta_2 X_t$ and then write

$$dX_t = (\alpha_0 + \alpha_1 X_t)dt + \beta_0 \operatorname{diag}(\operatorname{sqrt}[v(X_t)])\, dW_t,$$

where for $v, u \in \Re^3$ $\operatorname{sqrt}(v) = (v_1^{1/2}\ v_2^{1/2}\ v_3^{1/2})'$ and $\operatorname{diag}(u)$ is a diagonal matrix with the u_i, down the main diagonal and zeros elsewhere.

Throughout, we interpret $z^{1/2}$, $z \in \Re^1$, as the signed square root $z^{1/2} = \operatorname{sign}(z)|z|^{1/2}$, rather than follow Duffie and Kan (1993) in restricting the state space to regions where the elements of $v(X_t)$ are nonnegative. The use of the signed square root makes the diffusion function globally defined, which is of great convenience. It does mean that some parameters are identified only up to sign as discussed further below.

A key feature of this setup is that both the local mean function, $\alpha_0 + \alpha_1 X_t$, and the local variance function, $\beta_0 \operatorname{diag}[X_t]\beta_0'$, are affine in X_t. Among other things, this feature simplifies bond pricing calculations. If the elements of X_t are the yields on pure discount bonds, and if we take the yields themselves as the fundamental economic factors, then there exists a function $A(\tau)$ on $[0, \infty)$ into $\Re^1$ and $B(\tau)$ on $[0, \infty)$ into $\Re^3$ such that price at time t of a pure discount bond maturing at time $t + \tau$ is $\exp[A(\tau) + B(\tau)'X_t]$. Evidently, there are consistency relations between the parameters α_0, α_1, α_2, β_0, β_1, and β_2 and the $A(\tau)$ and $B(\tau)$, which are derived in Duffie and Kan (1993).

Our ten year interest rate is not the yield on a pure discount bond, so the pricing relations deduced in Duffie and Kan (1993) do not hold directly. Nonetheless, the three interest rates are quite reasonably taken as fundamental factors determining the term structure. The Yield-Factor Model is a good starting point for fitting continuous time models for these factors.

Imposing restrictions on β_0, β_1, and β_2 leads to intuitively appealing specifications of the diffusion function of the Yield-Factor Model. In our empirical work, we estimate these specifications using the method described in Section 2 above with α_0, α_1 left as as free.

The first, and simplest specification is obtained by letting β_0 be free, setting $\beta_1 = (1\ 1\ 1)'$, and setting $\beta_2 = I_3$, which gives

YF-Diagonal

$$dX_t \quad = \quad (\alpha_0 + \alpha_1 X_t)dt +$$

$$\beta_0 \begin{bmatrix} (1+X_{1t})^{1/2} & 0 & 0 \\ 0 & (1+X_{2t})^{1/2} & 0 \\ 0 & 0 & (1+X_{3t})^{1/2} \end{bmatrix} dW_t.$$

We call this the *YF-Diagonal* specification because the volatility factors depend on the separate interest rates, though β_0 is not restricted to be diagonal. Instead of $\beta_1 = (1\ 1\ 1)'$ we could have set $\beta_1 = (0\ 0\ 0)'$, though in exploratory work it was found that some offset relative to zero fits better. Since the interest rates are measure as percentages (numbers like 7.15, 8.35), the offset of unity implied by $\beta_1 = (1\ 1\ 1)'$ is fairly small.

The second specification is obtained by leaving β_0 free, $\beta_1 = (1\ 1\ 1)'$, and setting

$$\beta_2 = \begin{bmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ -1 & 0 & 1 \end{bmatrix}.$$

These restrictions give

YF-Premium

$$\begin{aligned} dX_t &= (\alpha_0 + \alpha_1 X_t)dt + \\ & \beta_0 \begin{bmatrix} (1+X_{1t})^{1/2} & 0 & 0 \\ 0 & (X_{2t}-X_{1t})^{1/2} & 0 \\ 0 & 0 & (X_{3t}-X_{1t})^{1/2} \end{bmatrix} dW_t \end{aligned}$$

For the *YF-Premium* specification, the overall level of interest rates affects local volatility through X_{1t} as do the term premiums $X_{2t} - X_{1t}$ and $X_{3t} - X_{1t}$. Since β_0 is unrestricted, the level and term premiums affect the local variance of all three yields.

The third is the *YF-General* specification with β_0 and β_2 free along with α_0 and α_1. $\beta_1 = (1\ 1\ 1)'$ remains as a normalization to achieve identification.

The square root function in the the diffusion function of the Yield-Factor Model ensures that the local variance is affine in the elements of X_t, which simplifies bond pricing calculations. However, for short rates, both single-factor models and two-factor time deformation models that have a square root diffusion function fare poorly in empirical applications (Aït-Sahalia, 1996; Conley, Hansen, and Scheinkman, 1994; Koedijk, Nissen, Schotman, and

Wolff, 1995; Tauchen, 1996). There is considerable evidence that values higher than 0.50 for the exponent are needed to accommodate to the data. These findings motivate our considering the specification

YF-Power

$$\begin{aligned} dX_t &= (\alpha_0 + \alpha_1 X_t)dt + \\ & \beta_0 \begin{bmatrix} (\beta_{11} + \beta_{2,[1]} X_t)^{\beta_{31}} & 0 & 0 \\ 0 & (\beta_{12} + \beta_{2,[2]} X_t)^{\beta_{32}} & 0 \\ 0 & 0 & (\beta_{13} + \beta_{2,[3]} X_t)^{\beta_{33}} \end{bmatrix} dW_t \end{aligned}$$

where α_0 is 3×1, α_1 is 3×3, β_0 is 3×3, $\beta_1 = \beta_1 = (1\ 1\ 1)'$, and β_2 is 3×3, and β_3 is 3×1. Setting $\beta_3 = (\frac{1}{2}\ \frac{1}{2}\ \frac{1}{2})'$ gives the *YF-General* model of Subsection 6.2. Analogous to the signed square root we use the signed power function $z^\gamma = \text{sign}(z)|z|^\gamma$ for $z, \gamma \in \Re^1$.

6.3 Estimation

We employ the EMM method described in Section 2 to estimate these specifications of the Yield-Factor Model and its extensions. Although the the likelihood function can be obtained in closed form when $\beta_3 = (\frac{1}{2}\ \frac{1}{2}\ \frac{1}{2})'$ (Feller, 1971; Duffie and Singleton, 1994), it is not available under the extensions we wish to examine. Also, unlike maximum likelihood, EMM provides diagnostics directly.

To compute

$$m(\rho, \theta) \doteq \frac{1}{N} \sum_{t=0}^{N} \frac{\partial}{\partial \theta} \log[f(\hat{y}_t | \hat{y}_{t-L}, \ldots, \hat{y}_{t-1}, \theta)]$$

we use an explicit order 2 weak as described in Section 4 above. For this work, time t is scaled so the interval $[t, t+1]$ is one week and $n_0 = 10$, which implies the simulation step size is $\Delta = 0.10$.

———————————————

Table 7 about here

———————————————

Table 7 summarizes the main results. As can seen be seen from the table, *YF-Diagonal* and *YF-Premium* models fare poorly as does the *YF-General*. The models are sharply rejected.

———————————————

Table 8 about here

———————————————

Table 8 provides insight as to what goes wrong. The Table shows quantiles of the unconditional distributions of the observed three interest rates along with those implied by the estimated models. The specifications *YF-Diagonal*, *YF-Premium*, and *YF-General* can accommodate the left tails of the distributions but fail to account for the skewness in the right tails. The Yield-Factor Model does not generate sufficient periods of high interest rates relative to those of the data.

This, along with previous findings such as Aït-Sahalia (1996) and Tauchen (1996) using univariate short-rate data, suggests that the *YF-Power* specification with exponents above 0.50 should perform better. On the other hand, the results of Conley, Hansen, and Scheinkman (1994) suggest that the exponent is above 0.50 for federal funds, but not Treasury Bills. The middle part of Table 7 reports the concentrated objective function for *YF-Power* specifications with the exponents restricted to a common value, $\beta_{31} = \beta_{32} = \beta_{33} = \beta$, $\beta = 0.60, 0.70, 0.80, 0.90$. These specifications come closer to fitting the Semiparametric ARCH score, with the best fit at $\beta = 0.70$, though the model is still rejected at conventional significance levels. If the common-value restriction on the exponents is maintained but β treated as a free parameter, then the objective function is quite flat in the region 0.70–0.75, with the point estimate being 0.706 and the objective value hardly improves (*YF-Power-Equal* in Table 7). Treating β_1, β_2, and β_3 as three free parameters (*YF-Power-Free*), provides little improvement, suggesting that the separate exponents are not sharply estimated.

Table 8 indicates that the *YF-Power* specifications do better than the basic Yield Factor models in terms of capturing the right skewness of the unconditional distribution of the interest rates. For exponents of 0.70, 0.80, and 0.90 the specifications do quite well in matching the 95% quantile but overpredict somewhat the 99% quantile.

For the interest rate diffusions considered here, the vector $-\alpha_1^{-1}\alpha_0$ contains the steady-state interest rates where the drift vector vanishes. The left side of Table 9 reports the point estimates of the steady-state rates, $-\hat{\alpha}_1^{-1}\hat{\alpha}_0$, which all appear reasonable relative to the unconditional distribution of the data reported in the top row of Table 8.

The eigenvalues of α_1 govern mean dynamics. A very robust finding is three real negative eigenvalues, with two of the eigenvalues being very close to zero and the third being much more negative. The right side of Table 9 reports the implied half-lives, arranged in ascending order, associated with the three eigenvalues. The two larger half lives lie between 104–232 weeks, i.e., on the order of 2–4 years, while the smallest is around 4 weeks, i.e., a month. This pattern suggests two extremely persistent factors along with a third, very transient factor, that together drive the predictable portion of the term structure.

———————————

Table 9 about here

———————————

7 References

Aït-Sahalia, Y. (1996) Testing continuous-time models of the spot interest rate. *The Review of Financial Studies* 9, 385–426.

Andersen, T. G. & J. Lund (1996a) Estimating continuous time stochastic volatility models of the short term interest rate. *Journal of Econometrics*, forthcoming.

Andersen, T. G. & J. Lund (1996b) Stochastic volatility and mean drift in the short rate diffusion: Sources of steepness, level, and curvature of the yield curve. Manuscript, Northwestern University.

Bansal, R., A. R. Gallant, R. Hussey & G. Tauchen (1993) Computational aspects of nonparametric simulation estimation. In D. A. Belsley (ed.) *Computational Techniques for Econometrics and Economic Analysis*, pp. 3–22. Boston: Kluwer Academic Publishers.

Bansal, R., A. R. Gallant, R. Hussey & G. Tauchen (1995) Nonparametric estimation of structural models for high-frequency currency market data. *Journal of Econometrics* 66, 251–287.

Bollerslev, T. & H. O. Æ. Mikkelsen (1996) Modeling and pricing long-memory in stock market volatility. *Journal of Econometrics* 73, 151-84.

Breidt, F. J., N. Crato & P. de Lima (1994) Modeling long-memory stochastic volatility. Manuscript, Iowa State University.

Broze, L., O. Scaillet & J.-M. Zakoian (1995) Quasi indirect inference for diffusion processes. Manuscript, C.O.R.E, Universite Catholique De Louvain.

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

Conley, T., L. P. Hansen & J. A. Scheinkman (1994) Estimating subordinated diffusions from discrete time data. Manuscript, University of Chicago.

Cox, J. C., J. E. Ingersoll, Jr. & S. A. Ross (1985) A theory of the term structure of interest rates. *Econometrica* 53, 385–407.

Ding, Z., C. W. J. Granger & R. F. Engle (1993) A long memory property of stock market returns and a new model. *Journal of Empirical Finance* 1, 83–108.

Duffie, D. & R. Kan (1993) A yield-factor model of interest rates. Manuscript, Stanford University.

Duffie, D. & K. J. Singleton (1993) Simulated moments estimation of markov models of asset prices. *Econometrica* 61, 929–952.

Duffie, D. & K. J. Singleton (1994) Econometric modeling of term structures of defaultable bonds. Manuscript, Stanford University.

Engle, R. F. & G. Gonzales-Rivera (1991) Semiparametric ARCH models. *Journal of Economic and Business Statistics* 9, 345–360.

Feller, W., (1971) *An Introduction to Probability Theory and Its Applications*, Vol. II, 2nd ed. New York: Wiley.

Gallant, A. R., D. A. Hsieh & G. Tauchen (1991) On fitting a recalcitrant series: The pound/dollar exchange rate, 1974–83. In W. A. Barnett, J. Powell & G. Tauchen (eds.) *Nonparametric and Semiparametric Methods in Econometrics and Statistics, Proceedings of the Fifth International Symposium in Economic Theory and Econometrics,* Cambridge, UK: Cambridge University Press.

Gallant, A. R., D. A. Hsieh & G. Tauchen (1997) Estimation of stochastic volatility models with diagnostics. *Journal of Econometrics*, forthcoming.

Gallant, A. R. & J. Long (1997) Estimating stochastic differential equations efficiently by minimum chi-square. *Biometrika*, forthcoming.

Gallant, A. R. & D. W. Nychka (1987) Semi-nonparametric maximum likelihood estimation. *Econometrica* 55, 363–390.

Gallant, A. R., P. E. Rossi & G. Tauchen (1992) Stock prices and volume. *Review of Financial Studies* 5, 199–242.

Gallant, A. R., P. E. Rossi & G. Tauchen (1993) Nonlinear dynamic structures. *Econometrica* 61, 871–907.

Gallant, A. R. & G. Tauchen (1989) Seminonparametric estimation of conditionally constrained heterogeneous processes: Asset pricing applications. *Econometrica*, 57, 1091–1120.

Gallant, A. R. & G. Tauchen (1996a) Which moments to match. *Journal of Econometric Theory* 12, 657–681.

Gallant, A. R. & G. Tauchen (1996b) Specification analysis of continuous time models in finance. In P. E. Rossi (ed.), *Modeling Stock Market Volatility: Bridging the Gap to Continuous Time*, pp. 357–384. New York: Academic Press.

Gihman, I. I. & A. V. Skorohod (1972) *Stochastic Differential Equations*. Berlin: Springer-Verlag.

Gourieroux, C., A. Monfort & E. Renault (1993) Indirect inference. *Journal of Applied Econometrics* 8, S85–S118.

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

Hansen, L. P. & J. A. Scheinkman (1995) Back to the future: Generating moment implications for continuous time markov process. *Econometrica* 50, 1029–1054.

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

Harvey, A. C. & N. Shephard (1996) Estimation of an asymmetric stochastic volatility model for asset returns. *Journal of Business and Economic Statistics* 14, 429–434.

Ingram, B. F. & B. S. Lee (1991) Simulation estimation of time series models. *Journal of Econometrics* 47, 197–250.

Kloeden, P. E. & E. Platen (1992) *Numerical Solution of Stochastic Differential Equations.* Berlin: Springer-Verlag.

Koedijk, K. G., F. G. J. A. Nissen, P. C. Schotman & C. C. P. Wolff (1995) The Dynamics of short-term interest rate volatility reconsidered. Manuscript, Limburg Institute of Financial Economics.

Lo, A. (1988) Maximum likelihood estimation of generalized Ito process with discreetly sampled data. *Econometric Theory* 4, 231–247.

Sowell, F. (1990) The fractional unit root distribution. *Econometrica* 58, 495–508.

Tauchen, G. (1997) The objective function of simulation estimators near the boundary of the stability set. Manuscript, Duke University.

Tauchen, G. (1996) New minimum chi-square methods in empirical finance. In D. Kreps & K. Wallis (eds.), *Advances in Econometrics, Seventh World Congress.* Cambridge, UK: Cambridge University Press.

Tauchen, G. & M. Pitts (1983) The price variability-volume relationship on speculative markets. *Econometrica* 51, 485–505.

Tables and Figures

Table 1. Discrete Time Stochastic Volatility Models Estimated from Stock Prices: Optimized Value of the Criterion for the Semiparametric ARCH Score.

Model						
L_y	L_w	p_ρ	Score	χ^2	*df*	p-value
Gaussian						
2	1	6	SA-L	104.055	20	< 0.0001
2	2	7	SA-L	100.812	19	< 0.0001
2	3	8	SA-L	88.934	18	< 0.0001
2	4	9	SA-L	77.239	17	< 0.0001
2	5	10	SA-L	76.615	16	< 0.0001
2	6	11	SA-L	64.130	15	< 0.0001
$t(\nu)$, $\nu = 10, 15, 20, 25$						
2	2	8	SA-L	78.456	18	< 0.0001
2	2	8	SA-L	70.210	18	< 0.0001
2	2	8	SA-L	70.999	18	< 0.0001
2	2	8	SA-L	72.435	18	< 0.0001
2	2	8	SA-L	73.752	18	< 0.0001
Spline						
2	1	8	SA-L	55.053	18	< 0.0001
2	2	9	SA-L	54.883	17	< 0.0001
2	3	10	SA-L	44.302	16	0.0002
2	4	11	SA-L	38.213	15	0.0008
2	4	12	SA-L	37.883	14	0.0005
2	6	13	SA-L	37.617	13	0.0003
Gaussian & Long-Memory						
2	1	7	SA-L	84.899	19	< 0.0001
2	2	8	SA-L	84.849	20	< 0.0001
Spline & Long-Memory						
2	1	9	SA-L	43.836	17	0.0004
2	2	10	SA-L	40.923	16	0.0006

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. p_ρ is the number of free parameters of the stochastic volatility model. χ^2 is the EMM objective function scaled to be distributed as a chi-squared on *df* degrees freedom under the maintained assumption of correct specification of the stochastic volatility model.

Table 2. Discrete Time Stochastic Volatility Models Estimated from Stock Prices: Optimized Value of the Criterion for the Nonlinear Nonparametric Score.

Model						
L_y	L_w	p_ρ	Score	χ^2	*df*	p-value
Gaussian						
2	1	6	NN-L	189.451	30	< 0.0001
2	2	7	NN-L	180.569	29	< 0.0001
2	3	8	NN-L	180.331	28	< 0.0001
2	4	9	NN-L	165.722	27	< 0.0001
2	5	10	NN-L	149.236	26	< 0.0001
2	6	11	NN-L	147.984	25	< 0.0001
Spline						
2	1	8	NN-L	159.389	28	< 0.0001
2	2	9	NN-L	151.982	27	< 0.0001
2	3	10	NN-L	146.392	26	< 0.0001
2	4	11	NN-L	145.978	25	< 0.0001
2	5	12	NN-L	137.812	24	< 0.0001
2	6	13	NN-L	134.021	23	< 0.0001
Gaussian-Asymmetric						
2	1	7	NN-L	142.898	29	< 0.0001
2	2	8	NN-L	114.415	28	< 0.0001
2	3	9	NN-L	101.397	27	< 0.0001
2	4	10	NN-L	88.184	26	< 0.0001
2	5	11	NN-L	84.167	25	< 0.0001
2	6	12	NN-L	79.836	24	< 0.0001
Spline-Asymmetric						
2	1	9	NN-L	77.081	27	< 0.0001
2	2	10	NN-L	76.529	26	< 0.0001
2	3	11	NN-L	73.112	25	< 0.0001
2	4	12	NN-L	72.224	24	< 0.0001
2	5	13	NN-L	72.153	23	< 0.0001
2	6	14	NN-L	64.683	22	< 0.0001
Spline & Long Memory						
2	1	9	NN-L	157.306	27	< 0.0001
2	2	10	NN-L	144.321	26	< 0.0001
Spline-Asymmetric & Long Memory						
2	1	10	NN-L	76.274	26	< 0.0001
2	2	11	NN-L	70.766	25	< 0.0001

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. p_ρ is the number of free parameters of the stochastic volatility model. χ^2 is the EMM objective function scaled to be distributed as a chi-squared on *df* degrees freedom under the maintained assumption of correct specification of the stochastic volatility model.

Table 3. Continuous Time Models Estimated from Stock Prices: Optimized Value of the Criterion for the Nonlinear Nonparametric Score.

						Quantiles					
Model	Score	N	χ^2	df	p-value	1%	5%	25%	75%	95%	99%
Data						-3.1	-1.8	-.56	.62	1.7	3.0
FD	NN-L	50k	376.966	33	< .0001	-1.7	-1.2	-.44	.57	1.3	1.8
FD-CH	NN-L	50k	281.545	32	< .0001	-1.7	-1.2	-.36	.51	1.3	1.9
FD-CH	NN-L	50k	281.432	31	< .0001	-1.8	-1.2	-.38	.52	1.3	2.0
SD-S0.5-SSV-A	NN-L	100k	65.767	25	< .0001	-2.7	-1.5	-.46	.56	1.5	2.6
SD-S0.5-SSV-A	NN-L	100k	59.202	24	.0001	-2.7	-1.4	-.45	.54	1.5	2.5
SD-S0.5-CH-SSV-A	NN-L	100k	57.898	23	.0001	-2.7	-1.5	-.44	.55	1.5	2.5
SD-S0.8-SSV-A	NN-L	100k	66.453	25	< .0001	-2.8	-1.6	-.52	.60	1.6	2.7
SD-S0.8-CH-SSV-A	NN-L	100k	56.522	24	.0002	-2.8	-1.5	-.49	.57	1.5	2.6
SD-S0.5-FSV-A	NN-L	100k	75.405	27	< .0001	-2.6	-1.5	-.50	.59	1.6	2.6
SD-S0.5-FSV-A	NN-L	100k	65.678	26	< .0001	-2.7	-1.6	-.50	.58	1.6	2.6
SD-S0.5-FSV-A	NN-L	100k	59.203	25	.0001	-2.7	-1.4	-.45	.54	1.5	2.5
SD-S0.8-FSV-A	NN-L	100k	57.734	25	.0002	-2.7	-1.4	-.45	.55	1.5	2.5
SD-S0.8-CH-FSV-A	NN-L	100k	54.391	24	.0004	-2.7	-1.5	-.44	.55	1.5	2.5
SD-S0.9-CH-FSV-A	NN-L	100k	54.005	24	.0004	-2.7	-1.4	-.44	.55	1.5	2.5
SD-S1.0-CH-FSV-A	NN-L	100k	53.513	24	.0005	-2.7	-1.4	-.44	.54	1.5	2.5
SD-S1.2-CH-FSV-A	NN-L	100k	49.832	24	.0015	-2.7	-1.4	-.44	.54	1.5	2.6
SD-S1.4-CH-FSV-A	NN-L	100k	41.146	24	.0161	-2.6	-1.4	-.44	.54	1.5	2.5
SD-S1.4-CH-FSV-A	NN	100k	35.622	24	.0597	-2.8	-1.5	-.47	.57	1.5	2.7
SD-S1.4-CH-FSV-A	NN	125k	36.563	24	.0484	-2.8	-1.6	-.50	.57	1.6	2.7
SD-S1.4-CH-FSV-A	NN	150k	31.686	24	.1350	-2.8	-1.6	-.49	.57	1.6	2.7
SD-S1.4-CH-FSV-A	NN	200k	36.239	24	.0520	-2.8	-1.5	-.48	.56	1.5	2.7
SD-S1.6-CH-FSV-A	NN-L	100k	47.157	24	.0032	-2.4	-1.3	-.42	.53	1.4	2.3

χ^2 is the EMM objective function scaled to be distributed as a chi-squared on df degrees freedom under the maintained assumption of correct specification of the continuous time model. Shown are raw quantiles from an order 1 strong scheme simulation of length N, which reflect differences in shape, location, and scale. Quantiles standardized to display only shape differences are shown in Figure 1.

Table 4. Continuous Time Models Estimated from Stock Prices: Estimates of Drift Parameters.

Model	Score	N	a_h	a_{hh}	a_{hu}	a_{uu}	a_{uh}	a_x	a_{xx}	a_{xv}	a_{vv}	a_{vx}
FD	NN-L	50k	0	0	0	0	0	.125	-1.96	0	0	0
FD-CH	NN-L	50k	0	0	0	0	0	.138	-1.95	0	0	0
FD-CH	NN-L	50k	0	0	0	0	0	.137	-1.95	0	0	0
SD-S0.5-SSV-A	NN-L	100k	.00329	0	1	-2.59	-.0677	.0490	0	1	-3.08	-4.09
SD-S0.5-SSV-A	NN-L	100k	-.0212	0	1	-47.9	-1.31	.0471	0	1	-3.08	-4.05
SD-S0.5-CH-SSV-A	NN-L	100k	-.0215	0	1	-47.9	-1.23	.0444	0	1	-3.06	-4.02
SD-S0.8-SSV-A	NN-L	100k	.00366	0	1	-47.1	-1.26	.0492	0	1	-3.05	-4.06
SD-S0.8-CH-SSV-A	NN-L	100k	-.0224	0	1	-47.4	-1.15	.0459	0	1	-3.15	-4.18
SD-S0.5-FSV-A	NN-L	100k	.00279	-.0348	0	0	0	.0545	0	1	-2.94	-3.88
SD-S0.5-FSV-A	NN-L	100k	.00333	-.0264	0	0	0	.0491	0	1	-3.08	-4.09
SD-S0.5-FSV-A	NN-L	100k	.00213	-.0273	0	0	0	.0471	0	1	-3.08	-4.05
SD-S0.8-FSV-A	NN-L	100k	-.0225	-.0263	0	0	0	.0462	0	1	-3.15	-4.17
SD-S0.8-CH-FSV-A	NN-L	100k	-.0249	-.0245	0	0	0	.0430	0	1	-3.10	-4.07
SD-S0.9-CH-FSV-A	NN-L	100k	-.0249	-.0243	0	0	0	.0433	0	1	-3.11	-4.07
SD-S1.0-CH-FSV-A	NN-L	100k	-.0249	-.0241	0	0	0	.0436	0	1	-3.11	-4.08
SD-S1.2-CH-FSV-A	NN-L	100k	-.0231	-.0231	0	0	0	.0444	0	1	-3.12	-4.10
SD-S1.4-CH-FSV-A	NN-L	100k	-.0314	-.0223	0	0	0	.0419	0	1	-3.09	-4.06
SD-S1.4-CH-FSV-A	NN	100k	-.0179	-.0219	0	0	0	.0341	0	1	-3.03	-3.89
SD-S1.4-CH-FSV-A	NN	125k	-.0179	-.0223	0	0	0	.0327	0	1	-2.96	-3.81
SD-S1.4-CH-FSV-A	NN	150k	-.0185	-.0223	0	0	0	.0311	0	1	-2.99	-3.82
SD-S1.4-CH-FSV-A	NN	200k	-.0180	-.0224	0	0	0	.0363	0	1	-2.93	-3.78
SD-S1.6-CH-FSV-A	NN-L	100k	-.0514	-.0172	0	0	0	.0467	0	1	-3.00	-3.98

Table 5. Continuous Time Models Estimated from Stock Prices: Estimates of Diffusion Parameters.

Model	Score	N	b_{h1}	b_{hh1}	b_{h2}	b_{x2}	b_{xx2}	b_{xw2}	γ	b_2	b_3	δ
FD	NN-L	50k	0	0	0	1.47	0	0	.5	0	0	0
FD-CH	NN-L	50k	0	0	0	1.17	.86	0	.5	0	0	0
FD-CH	NN-L	50k	0	0	0	1.28	.276	0	.846	0	0	0
SD-S0.5-SSV-A	NN-L	100k	.0445	-.0728	-.0664	0	0	1	1	-.0308	.0581	.5
SD-S0.5-SSV-A	NN-L	100k	.0489	-.102	-.144	.581	0	1	1	-.0249	.0501	.5
SD-S0.5-CH-SSV-A	NN-L	100k	.0473	-.0895	-.151	.582	.0588	1	1	-.0102	.0300	.5
SD-S0.8-SSV-A	NN-L	100k	.0512	-.0102	-.0635	0	0	1	1	-.0177	.0367	.8
SD-S0.8-CH-SSV-A	NN-L	100k	.0674	-.0891	-.154	.658	0	1	1	-.0129	.0305	.8
SD-S0.5-FSV-A	NN-L	100k	.0612	0	-.0616	0	0	1	1	-.0260	.0498	.5
SD-S0.5-FSV-A	NN-L	100k	.0449	-.0736	-.0671	0	0	1	1	-.0307	.0582	.5
SD-S0.5-FSV-A	NN-L	100k	.0489	-.102	-.144	.581	0	1	1	-.0249	.0503	.5
SD-S0.8-FSV-A	NN-L	100k	.0444	-.118	-.152	.624	0	1	1	-.0138	.0338	.8
SD-S0.8-CH-FSV-A	NN-L	100k	.0375	-.0973	-.168	.640	.0758	1	1	-.00423	.0180	.8
SD-S0.9-CH-FSV-A	NN-L	100k	.0381	-.0991	-.168	.642	.0784	1	1	-.00306	.0156	.9
SD-S1.0-CH-FSV-A	NN-L	100k	.0386	-.101	-.167	.643	.0815	1	1	-.00200	.0138	1.0
SD-S1.2-CH-FSV-A	NN-L	100k	.0226	-.102	-.161	.633	.110	1	1	.00971	.00949	1.2
SD-S1.4-CH-FSV-A	NN-L	100k	.0300	-.106	-.192	.712	.128	1	1	.000791	.00450	1.4
SD-S1.4-CH-FSV-A	NN	100k	.0105	-.0877	-.145	.608	.108	1	1	.000617	.00814	1.4
SD-S1.4-CH-FSV-A	NN	125k	.00660	-.0867	-.144	.614	.104	1	1	.000520	.00904	1.4
SD-S1.4-CH-FSV-A	NN	150k	.00726	-.0871	-.141	.610	.109	1	1	.000560	.00865	1.4
SD-S1.4-CH-FSV-A	NN	200k	.00757	-.0872	-.144	.592	.109	1	1	.000545	.00885	1.4
SD-S1.6-CH-FSV-A	NN-L	100k	.0806	-.120	-.258	.816	.167	1	1	.000652	-.00141	1.6

Table 6. Continuous Time Models Estimated from Stock Prices: Confidence Intervals on the Parameters of the Preferred Model.

				Criterion Difference 95% Confidence Limits	
Parameter	Estimate	Wald S.E.	Wald t-ratio	Lower	Upper
a_h	-.018597	.0047662	-3.9019	-.022778	-.014127
a_{hh}	-.022341	.00069758	-32.027	-.023001	-.022336
a_x	.031144	.010269	3.0329	.031098	.031149
a_{vv}	-2.9878	.12435	-24.027	-3.0660	-2.9878
a_{vx}	-3.8241	.15673	-24.399	-3.9727	-3.7200
b_{h1}	.0072596	.0068358	1.0620	.0072587	.0072602
b_{hh1}	-.0870728	.0029830	-29.190	-.087183	-.084264
b_{h2}	-.14649	.014763	-9.9227	-.16049	-.13247
b_{x2}	.60980	.064340	9.4777	.55500	.66606
b_{xx2}	.10858	.020840	5.2103	.10818	.11169
b_2	.00055988	.00008943	6.2605	.00055988	.00056047
b_3	.0086524	.00081579	10.606	.0080083	.0086524

Wald standard errors are obtained from diagonal elements of the estimated variance-covariance matrix of the asymptotic normal distribution of the EMM estimator. Criterion difference intervals are obtained by inverting the criterion difference test L as described in Section 2. The preferred model is described in Subsection 5.3 and shown as model SD-S1.4-CH-FSV-A, Score=NN, $N = 150$k, in Tables 3 through 5.

Table 7. Continuous Time Models Estimated from Interest Rate Data: Optimized Value of the Criterion for the Semiparametric ARCH Score.

Model	p_ρ	Score	N	χ^2	*df*	p-value
Data						
YF-DIAGONAL	21	SA	50k	227.270	39	$< .0001$
YF-PREMIUM	21	SA	50k	236.795	39	$< .0001$
YF-GENERAL	30	SA	50k	133.563	30	$< .0001$
YF-POWER (0.60)	30	SA	50k	73.078	30	$< .0001$
YF-POWER (0.70)	30	SA	50k	56.816	30	0.0022
YF-POWER (0.80)	30	SA	50k	57.809	30	0.0017
YF-POWER (0.90)	30	SA	50k	67.111	30	$< .0001$
YF-POWER-Equal	30	SA	50k	55.940	29	0.0019
YF-POWER-Free	30	SA	50k	54.942	27	0.0012

p_ρ is the number of free parameters in the continuous time model; SA denotes the *Semiparametric ARCH* score; N is the length of the simulation; χ^2 is the EMM objective function scaled to be distributed as a chi-squared on *df* degrees freedom under the maintained assumption of correct specification of the continuous time model.

Table 8. Continuous Time Models Estimated from Interest Rate Data: Quantiles of Unconditional Distributions

			Quantiles						
Model	Score	N	1%	5%	25%	50%	75%	95%	99%
Data									
3-Month			2.73	2.93	4.44	5.81	7.76	12.21	15.15
12-Month			2.93	3.14	4.71	6.15	7.73	12.00	14.12
10-Year			3.88	4.08	6.01	7.48	8.90	13.14	14.65
YF-DIAGONAL	SA	50k							
3-Month			2.38	3.24	4.60	5.74	6.91	8.85	10.35
12-Month			2.49	3.36	4.73	5.89	7.08	9.06	10.58
10-Year			3.49	4.36	6.06	7.35	8.60	10.85	12.29
YF-PREMIUM	SA	50k							
3-Month			3.13	3.96	5.27	6.18	7.13	8.43	9.12
12-Month			3.17	4.03	5.37	6.32	7.33	8.67	9.47
10-Year			3.45	4.71	6.60	7.78	8.97	10.64	11.33
YF-GENERAL	SA	50k							
3-Month			3.53	3.99	4.83	5.75	7.18	9.84	12.44
12-Month			3.65	4.13	4.99	6.00	7.56	10.47	13.28
10-Year			4.48	5.12	6.13	6.92	8.18	10.53	12.69
YF-POWER (0.60)	SA	50k							
3-Month			3.05	3.51	4.16	5.01	6.71	11.43	18.48
12-Month			3.33	3.61	4.23	5.16	7.08	12.32	19.92
10-Year			3.81	4.73	5.72	6.40	7.57	10.84	14.76
YF-POWER (0.70)	SA	50k							
3-Month			3.61	3.98	4.60	5.42	7.12	12.33	22.23
12-Month			3.84	4.09	4.68	5.57	7.48	13.29	24.29
10-Year			4.66	5.54	6.34	7.09	8.48	12.53	19.85
YF-POWER (0.80)	SA	50k							
3-Month			3.19	3.82	4.57	5.48	7.20	12.60	23.66
12-Month			3.46	3.91	4.66	5.64	7.59	13.59	25.96
10-Year			4.28	5.28	6.20	7.06	8.58	13.04	23.14
YF-POWER (0.90)	SA	50k							
3-Month			3.24	3.88	4.82	5.84	7.49	13.26	23.91
12-Month			3.35	3.99	4.96	5.99	7.71	13.85	24.70
10-Year			4.83	5.46	6.32	7.26	8.76	13.56	22.99

Shown are raw quantiles from an order 1 strong scheme simulation of length N, which reflect differences in shape, location, and scale.

Table 9. Implied Steady State Interest Rates and Half Lives

Model	Steady State Interest Rates 3-Month	12-Month	10-Year	Half Life, Weeks		
YF-GENERAL	6.21	6.50	7.43	3.68	94.3	198.2
YF-POWER (0.60)	6.06	6.36	7.06	4.55	171.1	232.1
YF-POWER (0.70)	6.56	6.88	8.03	4.71	179.8	179.8
YF-POWER (0.80)	6.50	6.83	7.97	4.72	197.9	198.0
YF-POWER (0.90)	7.15	7.38	8.39	4.32	104.6	145.0

Steady state interest rates are the components of $-\alpha_1^{-1}\alpha_0$, where the drift is $(\alpha_0 + \alpha_1 X_t)dt$; a half life is $\log(0.50)/\lambda$, where λ is an eigenvalue of α_1.

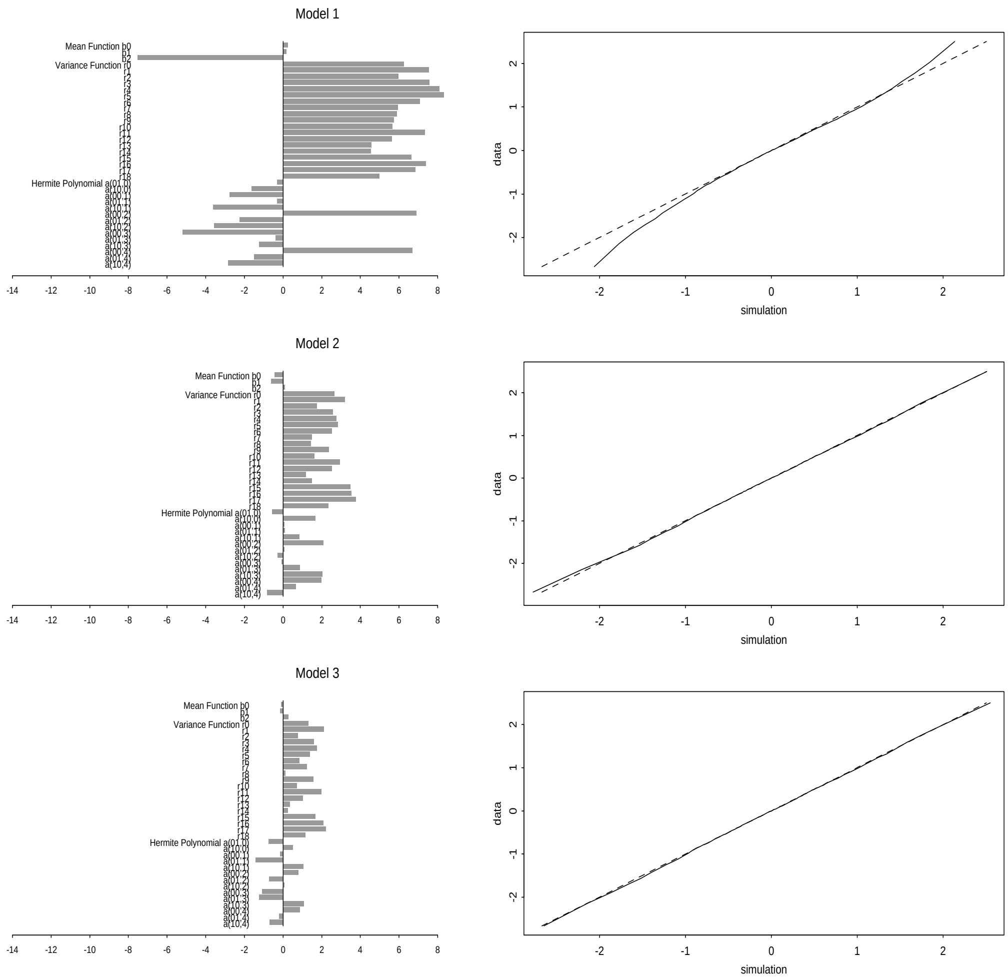

Figure 1. Quasi-t-Ratios and Quantile-Quantile Plots. The upper-left panel displays the vector of quasi-t-ratios $\hat{T}_n$ as a bar chart for the model FD-CH with $N = 50k$, and $\gamma = .846$. The t- ratios correspond to the parameters of the location function μ_x, scale function r_x, and Hermite polynomial $P^2(z, x)\phi(z)$ of the auxiliary model f. The solid line in the upper-right panel are the quantiles of an order 1 strong scheme simulation of length N plotted against the quantiles of the data. Prior to computing quantiles the simulation was standardized by subtracting the median and dividing by the interquartile range; similarly for the data. The dashed line is a 45^o line. The solid line can be interpreted as a transformation that can be applied to the simulation to produce the same shaped distribution as the data. The middle panels are the same for model SD-S0.5-CH-SSV-A with $N = 100k$, and the lower panels are the same for model SD-S1.4-CH-FSV-A with $N = 150k$.

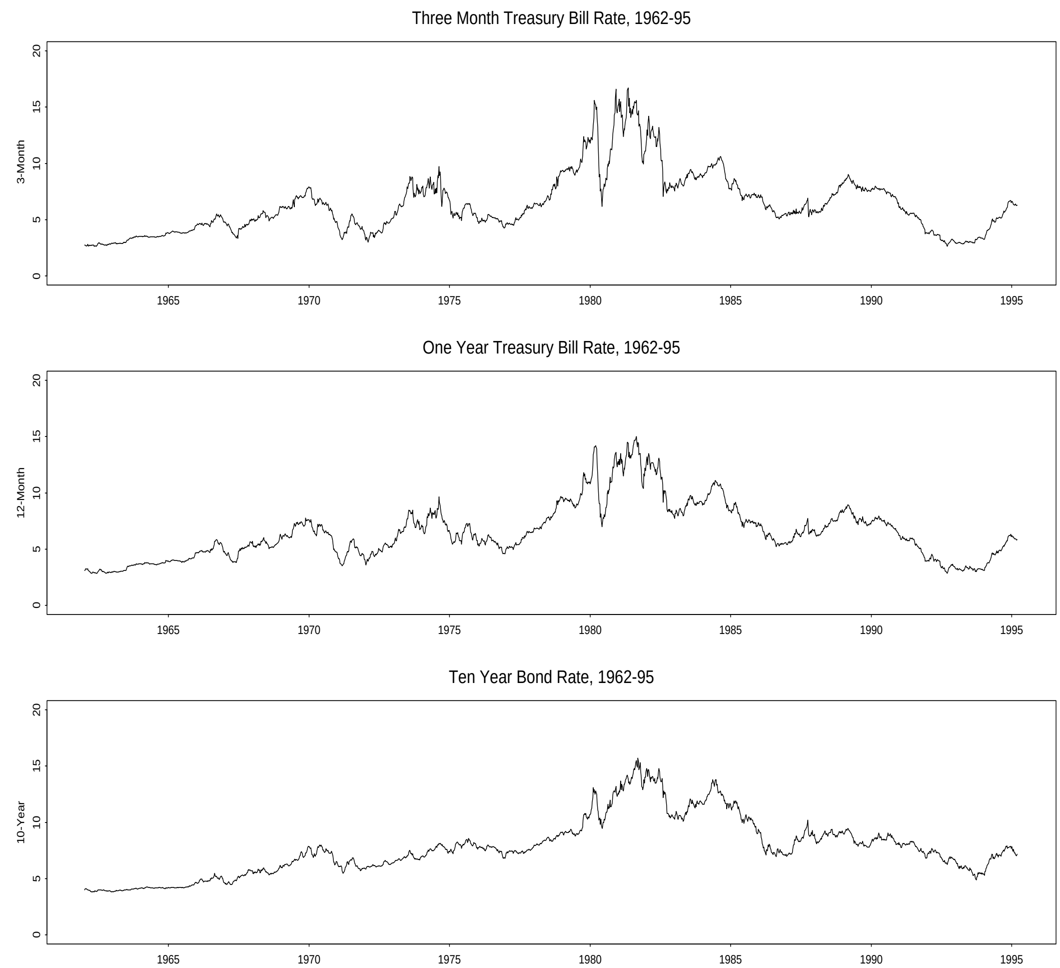

Figure 2. Interest Rate Data. The top panel panel is the Three Month Treasury Bill Rate, the middle is the Twelve Month Treasury Bill Rate, and the bottom is the 10-Year Constant Maturity Treasury Bond Rate, weekly, January 5, 1962 — March 31, 1995.