Reprojecting Partially Observed Systems With Application to Interest Rate Diffusions

A. Ronald GALLANT and George TAUCHEN

We introduce reprojection as a general purpose technique for characterizing the dynamic response of a partially observed nonlinear system to its observable history. Reprojection is the third step of a procedure wherein first data are summarized by projection onto a Hermite series representation of the unconstrained transition density for observables; second, system parameters are estimated by minimum chi-squared, where the chi-squared criterion is a quadratic form in the expected score of the projection; and third, the constraints on dynamics implied by the nonlinear system are imposed by projecting a long simulation of the estimated system onto a Hermite series representation of the constrained transition density for observables. The constrained transition density can be used to study the response of the system to its observable history. We utilize the technique to assess the dynamics of several diffusion models for the short-term interest rate that have been proposed and to compare them to a new model that has feedback from the interest rate into both the drift and diffusion coefficients of a volatility equation.

KEY WORDS: Efficient method of moments; Nonlinear dynamic models; Partially observed state; Stochastic differential equations.

1. INTRODUCTION

1.1 Interest Rate Diffusions

The data used in this article are observations on the weekly 3-month Treasury bill rate from January 5, 1962 to August 30, 1996, yielding 1,809 observations. Figure 1 plots the data; Table 1 provides summary descriptive statistics. The rate of interest over a short time interval (called the short rate) is a fundamental time series in economics and finance. Its dynamics describe the equilibrium substitution possibilities of goods, services, and wealth across time. Among other things, these dynamics play a central role in determining longer term bond prices and interest rates for various horizons.

The finance literature normally treats the short rate as a diffusion, usually expressed as a stochastic differential equation. Our data would thus be regarded as having resulted from discretely sampling a diffusion. For example, a well-known scalar diffusion model proposed for the short rate is the square root model of Cox, Ingersoll, and Ross (1985):

$$dU_t = (a_0 + a_1 U_t)dt + b_0 U_t^{1/2} dW_t, \qquad 0 \le t < \infty$$

and

$$y_t = U_t, \qquad t = 0, 1, \ldots \tag{1}$$

where $U_t, t \in [0, \infty)$, is the continuous-time short rate, W_t is a continuous-time Brownian motion, and $y_t, t = 0, 1, \ldots$, is the discretely sampled series. As discussed later, empirical evidence from this and other studies strongly discredits models like (1). The difficulties lie not so much with the form of the drift and diffusion as functions of U_t in (1), but rather with inherent limitations imposed by presumption of a scalar $U_t \in \Re^1$. The evidence suggests that a more appropriate specification is a stochastic volatility model with $U_t \in \Re^2$; an elementary version of this model is

$$dU_{1t} = (\alpha_{10} + \alpha_{11} U_{1t})dt + e^{U_{2t}} dW_{1t}, \qquad 0 \le t < \infty,$$

$$dU_{2t} = (\alpha_{20} + \alpha_{22} U_{2t})dt + \beta_{20} dW_{2t},$$

and

$$y_t = U_{1t}, \qquad t = 0, 1, \ldots \tag{2}$$

In (2) U_{1t} is the continuous-time short rate; U_{2t} is the logarithm of instantaneous volatility, which cannot be observed; W_{1t} and W_{2t} are independent, continuous-time Brownian motions; and $y_t = U_{1t}$ for $t = 0, 1, \ldots$ represents discrete sampling. The sequence $\{y_t\}$ is the observed Treasury bill series at the weekly frequency. More elaborate versions of these systems are discussed later in Section 3. The interpretation of systems such as (1) and (2) as diffusions follows that of Karatzas and Shreve (1991), but of more relevance here is the fact that such systems can be conveniently simulated using algorithms from Kloeden and Platen (1992).

1.2 Statistical Methods for Partially Observed Systems

The specification (2) leads us to consider statistical methods for the analysis of dynamic nonlinear models that have unobserved variables. Although the motivating problem (2) is a diffusion described by nonlinear stochastic differential equations, dynamic nonlinear models—expressed as diffusions, differential equations, or difference equations—in which the state vector is partially observed pervade science. For example, in epidemiology the SEIR model determines those susceptible to, exposed to, infected by, and recovered from a disease, whereas data usually are from case reports that report only those infected (Olsen and Schaffer 1990). The SEIR model is expressed as a system of differential equations. Noise can enter as random perturbation of coefficients, as noise that feeds back into the system, as observational error, or as some combination of these three sources.

A. Ronald Gallant is Latané Professor of Economics, Department of Economics, University of North Carolina, Chapel Hill, NC 27599. George Tauchen is Glasson Professor of Economics, Department of Economics, Duke University, Durham, NC 27708. This work was supported by the National Science Foundation.

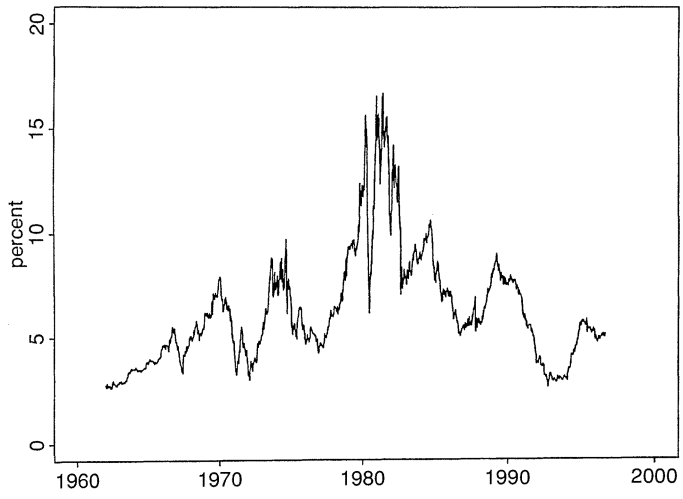

Figure 1. Weekly 3-Month Treasury Bill Rates from January 5, 1962 to August 30, 1996. Plotted are Friday's quote from the secondary market with Thursday's quote substituted if Friday was a holiday.

An example of the use of the estimation method described here with the SEIR model, but with a neural net score instead of a Hermite score, has been provided by Ellner, Gallant, and Theiler (1995). Other examples include discrete time stochastic volatility models of speculative markets from finance (Ghysels, Harvey, and Renault 1995), general equilibrium models from economics (Gennotte and Marsh 1993), and compartment models from pharmacokinetics (Mallet, Mentré, Steimer, and Lokiec 1988).

To estimate the parameters of nonlinear models such as (2) that have observed variables $y_t = U_{1t}$ and unobserved variables U_{2t}, standard statistical methods, both classical and Bayesian, are usually not applicable, because one lacks a convenient closed-form expression for the density of y_t given a history $x_{t-1} = (y_{t-L}, \ldots, y_{t-1})$ of L observations from the past. This density, denoted by $p(y_t|x_{t-1}, \rho)$, is called the transition density and depends on the parameters of the system, which are $\rho = (\alpha_{10}, \alpha_{11}, \alpha_{20}, \alpha_{22}, \beta\alpha_{20})$ in the case of (2). Typically, a transition density is not available for a nonlinear dynamical system either because no tractable expression for the joint density of $U_{1,t-L}, U_{2,t-L}, \ldots, U_{1t}, U_{2t}$ can be found or because the integration to eliminate $U_{2,t-L}, \ldots, U_{2t}$ from the expression is too difficult. Without the transition density, a likelihood is not available. Statistical methods are sometimes available for diffusions, differential equations, or difference equations whose structure happens to lend itself to Kalman filtering, Markov chain Monte Carlo, numerical quadrature, or clever analytical manipulation. Usually, however, one's model does not have one of these special structures, and if it does, alterations suggested by a scientific theory or a statistical test will often destroy it. For instance, (1) can be estimated by maximum likelihood, but the statistical evidence favors (2), which cannot. Our purpose here is to describe methods that are generally applicable.

Although computing the likelihood is often difficult if not infeasible, simulating the evolution of a state vector such as (U_{1t}, U_{2t}) of (2) is often practicable. Our methods rely on this. Briefly, we project the observed data onto a Hermite series representation of the transition density for observables. Given a parameter setting for the system, we use simulation to compute the expected score of the projection. The expected score thus computed depends on system parameters and is used to form a chi-squared criterion function that is a quadratic form in the expected score. A nonlinear optimizer is used to find the parameter setting that minimizes the criterion. This is a method-of-moments–type procedure that is as efficient in large samples as if maximum likelihood had been used and is thus termed efficient method of moments (EMM). Diagnostic tests are available to assess system adequacy; among them are informative t ratios that suggest reasons for model failure. Subsequent reprojection of the estimated dynamical system onto a Hermite series provides a facility for model elucidation that is as convenient as if the transition density for observables were available.

The use of method of moments together with simulation to estimate the parameters of dynamic models with unobserved variables is not new. Previous work includes that of Duffie and Singleton (1993), Gourieroux and Monfort (1996), Gourieroux, Monfort, and Renault (1993), Ingram and Lee (1991), and Smith (1993). The particular methods that we use are due to earlier work (Gallant and Tauchen 1996a), in which we discussed the procedure, diagnostics, and asymptotics. The theoretical support for the projection that we use was provided by Gallant and Long (1997), who showed that it achieves the same efficiency as maximum likelihood. We sketch the ideas from this literature with the intent of making this article self-contained.

A forceful criticism of simulation-based method-of-moments estimation has been that the method does not provide a representation of the observables in terms of their own past as do maximum likelihood based on a conditional density and time series methods such as autoregressive integrated moving average (ARIMA), autoregressive conditional heteroscedasticity (ARCH), and generalized autoregressive conditional heteroscedasticity (GARCH) modeling

Table 1. Descriptive Statistics for Weekly 3-Month Treasury Bill Rates From January 5, 1962 to August 30, 1996

Mean	6.28
Standard Deviation	2.68
Variance	7.20
Skewness	1.24
Kurtosis	1.73
Number of observations	1,809
Min	2.63
Q01	2.73
Q05	2.93
Q10	3.35
Q25	4.46
Median	5.66
Q75	7.70
Q90	9.54
Q95	12.08
Q99	15.15
Max	16.68

NOTE: Units are percentages. Kurtosis is $\kappa = \mu_4/\mu_2^2 - 3$, where μ_i is the ith central moment. QX denotes the X% quantile.

(Jacquier, Polson, and Rossi 1994). Thus the methodology cannot be used for model elucidation by providing, for example, descriptions of the volatility of the observed process as a function of its own past. This has provided motivation for ad hoc methods on a case-by-case basis. The primary methodological contribution of this article is to overcome this criticism. We introduce the notion of reprojection to get a representation of the observed process in terms of observables that incorporates the dynamics implied by the nonlinear system under consideration. Once done, the methods of model elucidation introduced by Gallant, Rossi, and Tauchen (1992, 1993) may be applied. If required by the application, one can as easily get a representation for unobservables in terms of the observed past and present that incorporates system dynamics.

We use the proposed methodology for estimation and diagnostic assessment of several diffusion models of the short rate expressed as a partially observed system of stochastic differential equations. We undertake this application because, as discussed earlier, interest rate diffusions are of considerable interest to economics and finance (see, e.g., Aït-Sahalia 1996a; Andersen and Lund 1996, 1997; Hansen and Scheinkman 1995; Lo 1988; Melino 1994), and we know of no other general-purpose method for elucidating the dynamics of a discretely sampled system of stochastic differential equations with a partially observed state.

2. EFFICIENT METHOD OF MOMENTS

2.1 Projection

Let $\{y_t\}_{t=-\infty}^{\infty}, y_t \in \Re^M$, be a discrete stationary time series. The stationary distribution of a contiguous subsequence $y_{t-L}, \ldots, y_t$ is presumed to have a density $p(y_{-L}, \ldots, y_0)$ defined over $\Re^l, l = M(L+1)$. Put $y = y_0, x = x_{-1} = (y_{-L}, \ldots, y_{-1})$, and write the stationary, marginal, and conditional densities as $p(x,y) = p(y_{-L}, \ldots, y_0), p(x) = \int p(y_{-L}, \ldots, y_0)\, dy_0$, and $p(y|x) = p(x,y)/p(x)$. Let $\{\tilde{y}_t\}_{t=-L}^{n}$ denote the realization from the process $\{y_t\}_{t=-\infty}^{\infty}$ that is available for analysis. We require estimates of the conditional density $p(y|x)$.

We describe an expanding class of conditional densities

$$\mathcal{H}_K = \{f_K(y|x,\theta) : \theta = (\theta_1, \theta_2, \ldots, \theta_{p_K})\}$$

proposed in earlier work (Gallant and Tauchen 1989), which we termed SNP for seminonparametric, that has two properties:

- The union $\mathcal{H} = \cup_{K=1}^{\infty} \mathcal{H}_K$ is quite rich, and it is reasonable to assume that the density $p(y|x)$ is contained in the closure of $\mathcal{H}$ under a weighted Sobolev norm.
- If θ is estimated by quasi-maximum likelihood,

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

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) nonparametric estimator of $p(y|x)$ with desirable qualitative features (Fenton and Gallant 1996).

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 reduce 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 given by

$$\mu_{x_{t-1}} = b_0 + Bx_{t-1} \tag{3}$$

for a matrix B; $\mu_{x_{t-1}}$ is presumed to depend on $L_u \le L$ lags, which is accomplished by putting leading columns of B to 0 as required.

The scale function R_x is given by

$$\operatorname{vech}(R_{x_{t-1}}) = \rho_0 + P|e_{t-1}^*|, \tag{4}$$

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 matrix, $e_{t-1}^* = [(y_{t-L_r} - \mu_{x_{t-L_r-1}}), \ldots, (y_{t-1} - \mu_{x_{t-2}})]$, and $|\cdot|$ denotes elementwise absolute value. The scale function depends on L_r lagged (unnormalized) innovations $(y_t - \mu_{x_{t-1}})$ and $L_r + L_u \le L$ lagged y_t in total. This is an ARCH-type process akin to that proposed by Nelson (1991). Later, we require the derivative of R_x with respect to (b, B) and thus replace $|e_{t-1}^*|$ in (4) by the twice continuously differentiable approximation $a(e_{t-1}^*)$, where

$$a(u) = \begin{cases} (|100u| - \pi/2 + 1)/100 & |100u| \ge \pi/2 \\ (1 - \cos(100u))/100 & |100u| < \pi/2. \end{cases}$$

For α with nonnegative integer elements, let $z^\alpha = z_1^{\alpha_1} \ldots z_M^{\alpha_M}$ and $|\alpha| = \sum_{k=1}^{M} \alpha_k$; similarly for x^β. Consider the density

$$h_K(z|x) = \frac{[P_K(z,x)]^2 \phi(z)}{\int [P_K(u,x)]^2 \phi(u)\, du} \tag{5}$$

formed from the polynomial

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

where $\phi(z) = (2\pi)^{-M/2} e^{-z'z/2}$. In (5) the term $P_K(z,x)\sqrt{\phi(z)}$ is a Hermite polynomial of degree K_z in z whose coefficients are polynomials of degree K_x in x; it enters (5) as the square to enforce positivity. 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 $A = [a_{\beta\alpha}]$ to 0, $P_K(z,x)$ can be made to depend on only $L_p \le L$ lags from x. In applications where M is large, the

coefficients $a_{\beta\alpha}$, corresponding to monomials z^α that represent high-order interactions, can be set to 0 with little effect on the adequacy of approximations. Let $I_z = 0$ indicate that no interaction coefficients are set to 0, let $I_z = 1$ indicate that coefficients corresponding to interactions z^α of order larger than $K_z - 1$ are set to 0, and so on; similarly for x^β and I_x.

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

$$f_K(y_t|x_{t-1},\theta) = \frac{\{P_K[R_{x_{t-1}}^{-1}(y_t - \mu_{x_{t-1}}), x_{t-1}]\}^2 \times \phi[R_{x_{t-1}}^{-1}(y_t - \mu_{x_{t-1}})]}{|\det(R_{x_{t-1}})|^{1/2} \int [P_K(u, x_{t-1})]^2 \phi(u)\, du} \quad (6)$$

completes the description of the SNP density. The vector θ 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. The tuning parameters are $L_u, L_r, L_p, K_z, I_z, K_x$, and I_x, which determine K and the dimension p_K of θ.

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

How best to select the tuning parameters L_u, L_r, L_p, K_z, I_z, K_x, and I_x is an open question. A strategy found to work well is to move upward along an expansion path using the Bayes information criterion (BIC),

$$\text{BIC} = s_n(\tilde\theta) + (1/2)(p_K/n)\log(n),$$

$$s_n(\theta) = -\frac{1}{n}\sum_{t=0}^{n} \log[f_K(\tilde y_t|\tilde y_{t-L},\ldots,\tilde y_{t-1},\theta)],$$

(Schwarz 1978) to guide the search, with models having small values of BIC preferred.

The expansion path has a tree structure. Rather than examining the full tree, the strategy is to expand first in L_u with $L_r = L_p = K_z = K_x = 0$ until the BIC turns upward. Next, expand L_r with $L_p = K_z = K_x = 0$; then expand K_z with $K_x = 0$; and finally expand L_p and K_x. It is useful to expand in K_z, L_p, and K_x at a few intermediate values of L_r, because it sometimes happens that the smallest value of BIC lies elsewhere within the tree.

When the estimated transition density $f_K(y_t|x_{t-1},\tilde\theta_n)$ is used in connection with the estimator described in Section 2.2, it is essential that $f_K(y_t|x_{t-1},\tilde\theta_n)$ not represent an explosive process, as discussed in detail by Tauchen (1997b). When estimated from persistent processes such as Figure 1, this can happen. We examine long simulations from candidate estimates $f_K(y_t|x_{t-1},\tilde\theta_n)$ and exclude from consideration those that are explosive. This is only a partial solution because, as discussed in Section 2.2, $f_K(y_t|x_{t-1},\tilde\theta_n)$ must be evaluated at simulations from systems such as (2), which can be unstable at trial parameter settings determined by the numerical optimizer. We find that transforming each component of x_{t-1} by the logarithmic spline

$$\hat x_i = \begin{cases} \frac{1}{2}[x_i - x^c - \log(1 - x_i - x^c)] & x_i < -x^c \\ x_i & -x^c \le x_i \le x^c \\ \frac{1}{2}[x_i + x^c + \log(1 + x_i - x^c)] & x^c < x_i \end{cases} \quad (7)$$

with $x^c = 4$ substantially enhances numerical stability and has a negligible effect on either evaluation of $f_K(y_t|x_{t-1},\tilde\theta_n)$ over the data or on the value computed for $\tilde\theta_n$. To avoid cluttered notation, (7) presumes that the data $\{\tilde y_t\}_{t=-L}^n$ have been scaled so each component of the vector $\tilde y_t$ (hence of x_{t-1}) has a sample mean of 0 and a sample variance of 1, so $x^c = 4$ means that the logarithmic spline has effect only 4 standard deviations beyond the mean. One should reexpress (7) in raw units for application.

2.2 Estimation

We now suppose that a dynamic system such as (2) defines the density $p(y_{-L},\ldots,y_0|\rho)$ for observables of Section 2.1, where $\rho \in \Re^{p_\rho}$ is a vector of unknown system parameters to be estimated. We are interested in the situation where an analytic expression for $p(y_{-L},\ldots,y_0|\rho)$ is not available but yet an expectation of the form

$$\mathcal{E}_\rho(g) = \int\cdots\int g(y_{-L},\ldots,y_0) \times p(y_{-L},\ldots,y_0|\rho)\, dy_{-L}\ldots dy_0$$

can be computed by simulation, quadrature, or other numerical means for given ρ. We focus on the case where simulation is used to compute $\mathcal{E}_\rho(g)$. That is, for given ρ, one generates the simulation $\{\hat y_t\}_{t=-L}^N$ from the system and puts

$$\mathcal{E}_\rho(g) = \frac{1}{N}\sum_{t=0}^{N} g(\hat y_{t-L},\ldots,\hat y_t)$$

with N large enough so that Monte Carlo error is negligible. Examples of this situation include the dynamical systems from finance, epidemiology, and economics discussed in Section 1.

Our objective is threefold: (1) estimate ρ; (2) test the hypothesis that the dynamical system under consideration generated the observed data $\{\tilde y_t\}_{t=-L}^n$; and (3) provide diagnostics that indicate how a rejected system should be modified to better describe the distribution of the observable process $\{y_t\}_{t=-\infty}^\infty$.

In earlier work (Gallant and Tauchen 1996a), we proposed an estimator for ρ in the situation where $\mathcal{E}_\rho(g)$ is all that is available. Termed the EMM estimator, it would be regarded as a minimum chi-squared estimator in the statistics literature and as a generalized method-of-moments (GMM)

estimator in the econometrics literature. Because it is minimum chi-squared, the optimized chi-squared criterion can be used to test system adequacy. Also, as shown later the moments that enter the criterion provide diagnostics that indicate how a tentatively maintained system should be modified when it is rejected by the test of system adequacy.

The moment equations for the EMM estimator are obtained from the score vector $(\partial/\partial\theta)\log f(y|x,\theta)$ of an auxiliary model $f(y|x,\theta)$ termed the score generator. We have shown (Gallant and Tauchen 1996a) that if the score generator $f(y|x,\theta)$ encompasses the maintained model $p(y|x,\rho)$, then the estimator is as efficient as maximum likelihood. Tauchen (1997a) set forth formulas that would lead one to expect that the EMM estimator will be nearly as efficient as maximum likelihood when the score generator $f(y|x,\theta)$ is a good statistical approximation to the process $\{y_t\}_{t=-\infty}^{\infty}$ in the sense of passing diagnostic tests, and so on. Gallant and Long (1997) supported this conjecture by showing that if the score generator is the seminonparametric (SNP) density $f_K(y|x,\theta)$ described in Section 2.1, then the efficiency of the EMM estimator can be made as close to that of maximum likelihood as desired by taking K large enough.

The EMM estimator $\hat{\rho}_n$ is computed as follows. Use the score generator

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

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,

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

This estimator presumes that the score generator $f(y|x,\theta)$ provides an adequate statistical approximation to the transition density of the data, so that $\{(\partial/\partial\theta)\log f(\tilde{y}_t|\tilde{x}_{t-1},\tilde{\theta}_n)\}$ is essentially serially uncorrelated. If $f(y|x,\theta)$ is not adequate, then one of the more complicated expressions for $\tilde{\mathcal{I}}_n$ set forth previously (Gallant and Tauchen 1996a) must be used, although the EMM estimator is still consistent and asymptotically normal. If the SNP density $f_K(y|x,\theta)$ of Section 2.1 is used as the score generator, and the model selection protocol of Section 2.1 is used to determine $(L_u,L_r,L_p,K_z,I_z,K_x,I_x)$, then one may expect $\tilde{\mathcal{I}}_n$ as earlier to be appropriate (Gallant and Long 1997).

Define

$$m(\rho,\theta) = \mathcal{E}_\rho\left\{\frac{\partial}{\partial\theta}\log[f(y_0|y_{-L},\ldots,y_{-1},\theta)]\right\},$$

which is computed by averaging over a long simulation:

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

The estimator is

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

The asymptotics of the estimator are as follows. If ρ^0 denotes the true value of ρ and θ^0 is an isolated solution of the moment equations $m(\rho^0,\theta)=0$, then

$$\lim_{n\to\infty}\hat{\rho}_n = \rho^0 \quad \text{a.s.}$$

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

$$\lim_{n\to\infty}\hat{M}_n = M^0 \quad \text{a.s.}$$

$$\lim_{n\to\infty}\tilde{\mathcal{I}}_n = \mathcal{I}^0 \quad \text{a.s.} \tag{8}$$

where $\hat{M}_n = M(\hat{\rho}_n,\tilde{\theta}_n), M^0 = M(\rho^0,\theta^0), M(\rho,\theta) = (\partial/\partial\rho')m(\rho,\theta)$, and

$$\mathcal{I}^0 = \mathcal{E}_{\rho^0}\left[\frac{\partial}{\partial\theta}\log f(y_0|x_{-1},\theta^0)\right]\left[\frac{\partial}{\partial\theta}\log f(y_0|x_{-1},\theta^0)\right]'.$$

As a referee noted, there are instances where the EMM technique is applicable but increasing N to reduce the error in computing $\mathcal{E}_\rho(g)$ is too costly to be feasible. Quadrature can be used instead in some instances (Gallant and Tauchen 1996a). Another approach is to incorporate the Monte Carlo variance into reported standard errors (Gourieroux et al. 1993). As the referee remarked, in some application areas finding an efficient means for computing $\mathcal{E}_\rho(g)$ is an active area of research.

Under the null hypothesis that $p(y_{-L},\ldots,y_0|\rho)$ is the correct model,

$$L_0 = nm'(\hat{\rho}_n,\tilde{\theta}_n)(\tilde{\mathcal{I}}_n)^{-1}m(\hat{\rho}_n,\tilde{\theta}_n) \tag{9}$$

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

$$L_h = n[m'(\hat{\rho}_n,\tilde{\theta}_n)(\tilde{\mathcal{I}}_n)^{-1}m(\hat{\rho}_n,\tilde{\theta}_n) - m'(\hat{\rho}_n,\tilde{\theta}_n)(\tilde{\mathcal{I}}_n)^{-1}m(\hat{\rho}_n,\tilde{\theta}_n)] \tag{10}$$

is asymptotically chi-squared on q df, 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 be 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 ith 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 the system under consideration is explosive. Confidence intervals obtained by inverting the criterion difference test L_h do reflect this phenomenon and thus

are more useful. To invert the test, one puts in the interval those ρ_i^* for which L_h for the hypothesis $\rho_i^0 = \rho_i^*$ is less than the critical point of a chi-squared on 1 df. To avoid reoptimization, 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_h, where $\hat{\Sigma}_{(i)}$ is the ith column of $\hat{\Sigma}$.

As a referee pointed out, the foregoing remarks should be taken to imply only that confidence intervals obtained by inverting the criterion difference test have more desirable structural characteristics than those obtained by inverting the Wald test and not that they have more accurate coverage probabilities. It seems that the only way to assess coverage probabilities is by Monte Carlo. Simulations for a discrete-time stochastic volatility model by Andersen, Chung, and Sorensen (1997) and Chumacero (1997) and suggest that Wald intervals have coverage probabilities reasonably close to their nominal values. Their work does not cover the criterion difference test. We know of no other relevant Monte Carlo results.

When L_0 exceeds the chi-squared critical point, diagnostics that suggest improvements to the system are desirable. Because

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

inspection of the t ratios

$$T_n = S_n^{-1}\sqrt{n}m(\hat{\rho}_n, \tilde{\theta}_n), \tag{11}$$

where $S_n = (\mathrm{diag}\{\tilde{\mathcal{I}}_n - (\hat{M}_n)[(\hat{M}_n)'(\tilde{\mathcal{I}}_n)^{-1}(\hat{M}_n)]^{-1}(\hat{M}_n)'\})^{1/2}$, can suggest reasons for failure. Different elements of the score correspond to different characteristics of the data, and large t ratios reveal those characteristics that are not well approximated.

2.3 Reprojection

Having the EMM estimate of system parameters $\hat{\rho}_n$ in hand, we should like to elicit the dynamics of the implied conditional density for observables

$$\hat{\rho}(y_0|y_{-L}, \ldots, y_{-1}) = p(y_0|y_{-L}, \ldots, y_{-1}, \hat{\rho}_n). \tag{12}$$

Although analytic expressions are not available, an unconditional expectation,

$$\mathcal{E}_{\hat{\rho}_n}(g) = \int \cdots \int g(y_{-L}, \ldots, y_0) \times p(y_{-L}, \ldots, y_0|\hat{\rho}_n)\, dy_{-L} \ldots dy_0,$$

can be computed by generating a simulation $\{\hat{y}_t\}_{t=-L}^N$ from the system with parameters set to $\hat{\rho}_n$ and using

$$\mathcal{E}_{\hat{\rho}_n}(g) = \frac{1}{N}\sum_{t=0}^{N} g(\hat{y}_{t-L}, \ldots, \hat{y}_t).$$

With respect to unconditional expectation so computed, define

$$\hat{\theta}_K = \underset{\theta \in \Re^p K}{\mathrm{argmax}}\, \mathcal{E}_{\hat{\rho}_n} \log f_K(y_0|y_{-L}, \ldots, y_{-1}, \theta)$$

where $f_K(y_0|y_{-L}, \ldots, y_{-1}, \theta)$ is the SNP density given by (6). Let

$$\hat{f}_K(y_0|y_{-L}, \ldots, y_{-1}) = f_K(y_0|y_{-L}, \ldots, y_{-1}, \hat{\theta}_K). \tag{13}$$

Theorem 1 of Gallant and Long (1997) states that

$$\lim_{K\to\infty} \hat{f}_K(y_0|y_{-L}, \ldots, y_{-1}) = \hat{p}(y_0|y_{-L}, \ldots, y_{-1}).$$

Convergence is with respect to a weighted Sobolev norm that they describe. Of relevance here is that convergence in their norm implies that $\hat{f}_K$, as well as its partial derivatives in $(y_{-L}, \ldots, y_{-1}, y_0)$, converge uniformly over $\Re^l$, $l = M(L+1)$ to those of $\hat{p}$. We propose to study the dynamics of $\hat{p}$ by using $\hat{f}_K$ as an approximation. This result provides the justification for our approach.

To approximate $\hat{p}$ by $\hat{f}_K$, values of $(L_u, L_r, L_p, K_z, I_z, K_x, I_x)$ must be chosen. It seems natural to reuse the values of the projection that determined $\hat{\rho}_n$ because, among other things, that choice facilitates a comparison of the constrained dynamics determined by the estimated system with the unconstrained dynamics determined by the data. However, if the estimated nonlinear system is to be sampled at a different frequency than the data, then it will be necessary to redetermine $(L_u, L_r, L_p, K_z, I_z, K_x, I_x)$ by the methods described in Section 2.1. We anticipate that the dynamics at a different sampling frequency often will not be of interest, and we presume in what follows that the sampling frequency of the nonlinear system is the same as that of the data. The modifications required when it differs are mentioned as they occur.

Of immediate interest in eliciting the dynamics of observables are the first two one-step-ahead conditional moments:

$$\mathcal{E}(y_0|y_{-L}, \ldots, y_{-1}) = \int y_0 f_K(y_0|x_{-1}, \hat{\theta}_K)\, dy_0$$

and

$$\mathrm{var}(y_0|y_{-L}, \ldots, y_{-1}) = \int [y_0 - \mathcal{E}(y_0|x_{-1})] \times [y_0 - \mathcal{E}(y_0|x_{-1})]' f_K(y_0|x_{-1}, \hat{\theta}_K)\, dy_0,$$

where $x_{-1} = (y_{-L}, \ldots, y_{-1})$. Owing to the form of a Hermite expansion, expressions for these integrals as linear combinations of high-order moments of the normal distribution are available (Gallant and Tauchen 1992). The moments themselves may be obtained from standard recursions for the moments of the normal (Johnson and Kotz 1970).

Filtered volatility is the one-step-ahead conditional standard deviation evaluated at data values; viz.

$$\sqrt{\mathrm{var}(y_{k0}|y_{-L}, \ldots, y_{-1})}|_{(y_{-L}, \ldots, y_{-1}) = (\tilde{y}_{t-L}, \ldots, \tilde{y}_{t-1})} \tag{14}$$

for $t = 0, \ldots, n$. In (14) $\tilde{y}_t$ denotes data and y_{k0} denotes the kth element of the vector $y_0, k = 1, \ldots, M$. Because filtered volatility is a data-dependent concept, the dynamical system must be sampled at the same frequency as the data to determine $\hat{f}_K$. As mentioned earlier, it had been thought

that filtered volatility could not be recovered from method-of-moments estimates of a nonlinear dynamical system with partially observed state and that this has been a criticism of such estimates. However, as just seen, filtered volatility is easily computed using the reprojection notion.

As a referee has pointed out, we are using the term "filtered volatility" with a purely ARCH-type meaning, as in the nonlinear impulse-response literature. Another usage of filtering, perhaps the predominant one, involves estimating an unobserved state variable conditional on all past and present observables. Filtering according to this notion (for L lags rather than back to the first observation) can be accomplished through reprojection. This may be seen by noting that one can repeat the derivation of Section 2.1 with y taken to be a contemporaneous unobserved variable and x taken to be contemporaneous and lagged observed variables. Denote y and x thus modified by y^* and x^*. The result is a density $f_K(y^*|x^*,\theta)$ of the same form as (6) but with altered dimensions. One can simulate $\{y_t^*, x_t^*\}$ from the structural modal and perform the reprojection step to get $\hat{f}_K(y^*|x^*)$ as described earlier. The proof of Gallant and Long (1997) can be altered to justify these modifications. How one uses $\hat{f}_K(y^*|x^*)$ will be application specific. The application mentioned by the referee is to obtain an estimate of U_{2t} in a system such as (2) for the purpose of pricing an option. In this instance $y_t^* = U_{2t}, x_t^* = (U_{1,t-L},\ldots,U_{1t})$, and $\hat{U}_{2t}(x^*) = \int y^* \hat{f}_K(y^*|x^*)\,dy^*$. To avoid any confusion, hereafter we refer to (14) as reprojected volatility. We now return to the main discussion.

One-step-ahead dynamics may be studied by means of plots of (the elements of) $\mathcal{E}(y_0|y_{-L},\ldots,y_{-1}+\Delta)$, $\mathrm{var}(y_0|y_{-L},\ldots,y_{-1}+\Delta)$, or other conditional moments against δ, where Δ is an M vector with δ in the ith element and 0s elsewhere. More general perturbation strategies may be considered such as $\Delta = \delta\tilde{y}_\tau$, where $\tilde{y}_\tau$ is a point chosen from the data such that perturbations in the direction $\delta\tilde{y}_\tau$ take into account contemporaneous correlations among the components of y_t. Perturbations to a single element of y_{-1} in a multivariate setting may represent a movement that is improbable according to the dynamics of the system. Some thought must be given to the perturbation scheme in multivariate applications if plots of conditional moments against δ are to be informative. This issue has been discussed by Gallant et al. (1993).

Two methods for choosing $(y_{-L},\ldots,y_{-1})$ for these plots suggest themselves. The first method is to put $y_{-L},\ldots,y_{-1}$ to the sample mean—that is, put $(y_{-L},\ldots,y_{-1}) = (\bar{y},\ldots,\bar{y})$; where $\bar{y} = (1/n)\sum_{t=0}^{n}\tilde{y}_t$—and plot, for instance,

$$\mathrm{var}(y_0|\bar{y},\ldots,\bar{y}+\Delta) \tag{15}$$

against δ. The second method is to average over the data and plot, for instance,

$$(1/n)\sum_{t=0}^{n}\mathrm{var}(y_t|\tilde{y}_{t-L},\ldots,\tilde{y}_{t-1}+\Delta) \tag{16}$$

against δ. If the estimated system is sampled at a different frequency than the data, then the average $(1/N)\sum_{t=0}^{n}\mathrm{var}(y_t|\hat{y}_{t-L},\ldots,\hat{y}_{t-1}+\Delta)$ is plotted over a simulation $\{\hat{y}_t\}_{t=-L}^{N}$ at the correct frequency instead.

In an economic system, the graphics just described are interpreted as representing the consequences of a shock to the system that comes as a surprise to the economic agents involved, and similar interpretations hold in other contexts. If one wants to consider the consequences of forcing the system to a different equilibrium, then the graphic obtained by plotting $\mathrm{var}(y_0|y_{-L}+\Delta,\ldots,y_{-1}+\Delta)$ against δ is relevant. We provide illustrations of both in Section 3. They can be quite different.

Multistep-ahead dynamics may be studied by considering plots of the trajectories

$$\mathcal{E}[g(y_{j-L},\ldots,y_{j-1})|y_{-L},\ldots,y_{-1}+\Delta] \tag{17}$$

against $j = 0,1,\ldots,J$, where $g(y_{-L},\ldots,y_{-1})$ is a time-invariant function whose choice is discussed next. As discussed by Gallant et al. (1993), if one sets the initial condition to $(y_{-L},\ldots,y_{-1}+\Delta) = (\bar{y},\ldots,\bar{y}+\Delta)$, then it is helpful to net out transients by plotting either

$$\mathcal{E}[g(y_{j-L},\ldots,y_{j-1})|\bar{y},\ldots,\bar{y}+\Delta] - \mathcal{E}[g(y_{j-L},\ldots,y_{j-1})|\bar{y},\ldots,\bar{y}] \tag{18}$$

or

$$\frac{1}{n}\sum_{t=0}^{n}\mathcal{E}[g(y_{t+j-L},\ldots,y_{t+j-1})|\tilde{y}_{t-L},\ldots,\tilde{y}_{t-1}+\boldsymbol{\Delta}] \tag{19}$$

against $j = 0,1,\ldots,J$ instead of (17). Although (19) is conceptually superior, in the examples considered by Gallant et al. (1993), plots of (18) had nearly the same appearance and are much cheaper to compute.

To compute (17), one exploits the fact that there are efficient algorithms for sampling the density $\hat{f}_K(y_0|y_{-L},\ldots,y_{-1}+\Delta)$ recursively to obtain R simulated futures

$$\{\hat{y}_{0,i},\ldots,\hat{y}_{J,i}\}, \qquad i = 1,\ldots,R,$$

each conditional on $y_{-L},\ldots,y_{-1}+\Delta$ (Gallant and Tauchen 1992). Prepend $\{y_{-L},\ldots,y_{-1}+\Delta\}$ to each future to obtain the sequences

$$\{\hat{y}_{-L,i},\ldots,\hat{y}_{-1,i},\hat{y}_{0,i},\ldots,\hat{y}_{J,i}\}, \qquad i = 1,\ldots,R.$$

Then $\mathcal{E}[g(y_{j-L},\ldots,y_{j-1})|y_{-L},\ldots,y_{-1}+\Delta]$ can be computed as

$$\mathcal{E}[g(y_{j-L},\ldots,y_{j-1})|y_{-L},\ldots,y_{-1}+\Delta] = \frac{1}{R}\sum_{i=1}^{R}g(\hat{y}_{j-L,i},\ldots,\hat{y}_{j-1,i}).$$

A general discussion of appropriate choice of $g(y_{-L},\ldots,y_{-1})$ for nonlinear impulse-response analysis, the analysis of turning points, and so on has been provided by Gallant et al. (1993). Of these, the more routinely useful are conditional mean profiles and conditional volatility profiles. Conditional mean profiles are plots of

$$\mu_j(y_{-L},\ldots,y_{-1}+\Delta) = \mathcal{E}[\mathcal{E}(y_{k,j}|y_{j-L},\ldots,y_{j-1})|y_{-L},\ldots,y_{-1}+\Delta]$$

against $j = -1, \ldots, J$ for the components $k = 1, \ldots, M$ of y, which extend the impulse-response profiles of Doan, Litterman, and Sims (1984) and Sims (1980) to nonlinear systems. Conditional volatility profiles are plots of

$$\sigma_j^2(y_{-L}, \ldots, y_{-1} + \Delta) = \mathcal{E}[\text{var}(y_{k,j}|y_{j-L}, \ldots, y_{j-1})|y_{-L}, \ldots, y_{-1} + \Delta]$$

against $j = 0, \ldots, J$ for the components $k = 1, \ldots, M$ of y, which extend the volatility impulse-response profiles of Bollerslev and Engle (1993) and Engle, Ito, and Lin (1990) to nonlinear systems. Plots of the conditional mean profile reveal the future dynamic response of system forecasts to a contemporaneous shock to the system. These generally will be nonlinear and can differ markedly when the sign of δ changes. This holds similarly for volatility.

Persistence can be studied by inspection of profile bundles, which are overplots for $t = 0, \ldots, n$ of the profiles

$$\{\mu_j(\tilde{y}_{t-L}, \ldots, \tilde{y}_{t-1}), j = -1, \ldots, J\}. \tag{20}$$

That is, one overplots profiles conditional on each observed datum. If the thickness of the profile bundle tends to collapse to 0 rapidly, then the process is mean reverting. If the thickness tends to retain its width, then the process is persistent. Similarly, the profile bundles

$$\{\{\sqrt{\sigma_j^2}(\tilde{y}_{t-L}, \ldots, \tilde{y}_{t-1}), j = 0, \ldots, J\}, t = 0, \ldots, n\} \tag{21}$$

can be used to examine volatility for persistence. These are extensions to nonlinear systems of notions of persistence due to Bollerslev and Engle (1993). Rather than comparing plots, one can instead compare half-lives. A half-life $\hat{j}$ can be obtained by computing the range R_j at each ordinate $j = 0, \ldots, J$ of either (20) or (21), regressing $\log R_j$ on $j\beta$, and using $(-\log 2)/\hat{\beta}$ as an estimate of half-life.

Extensive examples of the use of the methods described here for elucidating the joint dynamics of stock prices and volume have been provided by Gallant et al. (1992, 1993). Here we also provide illustrations in Section 3.

3. THE SHORT RATE

The diffusion specifications that we consider are contained within the general setup

$$\begin{pmatrix} dU_1 \\ dU_2 \end{pmatrix} = \begin{pmatrix} \alpha_{10} + \alpha_{11}U_1 \\ \alpha_{20} + \alpha_{21}U_1 + \alpha_{22}U_2 \end{pmatrix} dt + \begin{pmatrix} (\beta_{10} + \beta_{11}U_1^{\gamma})e^{U_2} & 0 \\ 0 & \beta_{20} + \beta_{21}U_1 \end{pmatrix} \times \begin{pmatrix} dW_1 \\ dW_2 \end{pmatrix} \tag{22}$$

$$y_t = U_1 t, \qquad t = 0, 1, 2, \ldots$$

Here U_{1t} is the continuous record of the short-term rate of interest whereas U_{2t} is an unobserved volatility factor; y_t is the discretely sampled short-term interest rate. The restriction $\alpha_{20} = -\alpha_{22}$ is imposed to achieve identification; it implies a steady-state value of $U_{2t} = 1$ when $\alpha_{21} = 0$. The initial condition U_0 is a draw from the stationary distribution, which is normally accomplished by letting the system run until transients dissipate. We interpret $\alpha_{21} = \alpha_{22} = \beta_{20} = \beta_{21} = 0$ to mean $U_{2t} = 1$ for all $t \geq 0$, so that the volatility factor becomes irrelevant and the setup defines a one-factor scalar diffusion model for the short rate. The general setup encompasses several different models, which are defined in Table 2:

- **OU:** α_{10}, α_{11}, and β_{10} free; all others set to 0, γ is irrelevant. This is the basic Ornstein–Uhlenbeck process used by Vasicek (1977) in one of the earliest continuous-time models of the yield curve.
- **SQRT:** α_{10}, α_{11}, and β_{11} free and $\gamma = \frac{1}{2}$; all other parameters set to 0. This is the well-known square root model used by Cox et al. (1985) to model the yield curve.
- **SQRT0:** $\alpha_{10}, \alpha_{11}, \beta_{10}$, and β_{11} free and $\gamma = \frac{1}{2}$; all other parameters set to 0. This is the SQRT model with the intercept β_{10} included in the diffusion function. Aït-Sahalia (1996a) found that more flexibility near the origin in the diffusion function is empirically important.
- **CKLS:** α_{10}, α_{11}, and β_{11} free and $\gamma > \frac{1}{2}$; all other parameters set to 0. This is the constant elasticity of variance model of Chan, Karolyi, Longstaff, and Sanders (1992).
- **CKLS0:** The CKLS specification with an intercept β_{10} in the diffusion function.
- **SQRT-SV, SQRT0-SV, CKLS0-SV:** The corresponding preceding specifications with α_{22} and β_{20} allowed to be free, which activates the volatility process U_2 and generates a two-factor model. The stochastic volatility process U_2 evolves autonomously.
- **CKLS0-SV-FB:** Similar to CKLS0-SV, but α_{21} and β_{21} are free, so that the interest rate process U_1 feeds back into the drift and diffusion of the volatility process U_2.

The leading special cases are the SQRT and SQRT0 specifications, which are widely used in modeling the yield curve. The specifications offer manipulative convenience because they imply analytically tractable expressions for the interest rates and bond prices at different horizons in single-factor models of the yield curve. The task of solving the usual partial differential equation of bond pricing (Ingersoll 1987, p. 396) can be reduced to solving an ordinary differential equation of the Riccatti type (as in Duffie and Kan 1996). Also, unlike the OU specification, which permits negative interest rates and thereby violates arbitrage, certain parameter constraints ensure that the interest rate cannot reach 0. As can be easily checked for the SQRT0 specification, if $-\beta_{10}/\beta_{11} \geq 0$ and $\alpha_{10} + \alpha_{11}U_1^* > 0$, where $U_1^* = (-\beta_{10}/\beta_{11})^2$, then U_1^* is a reflecting barrier; if $U_{10} > U_1^*$, then $U_{1t} > U_1^*$ for all $t \geq 0$. Another feature is that the SQRT and SQRT0 specifications are consistent with higher volatility at higher levels, which is a prominent characteristic of U.S. interest rates, at least during the monetarist experiment of 1978–1983 (see Fig. 1). Finally, because the discrete time transition density has a closed form, SQRT and SQRT0 are estimable by classical maximum likelihood (see, e.g., Duffie and Singleton 1994).

Table 2. Model Definitions and Minimized Chi-Squared Criterion

Specification	α_{10}	α_{11}	α_{21}	α_{22}	β_{10}	β_{11}	β_{20}	β_{21}	γ	N	$X^2(\hat{\rho})$	df	p value
OU	*	*	0	0	*	0	0	0	—	10k	73.782	13	0
OU	*	*	0	0	*	0	0	0	—	50k	72.064	13	0
OU	*	*	0	0	*	0	0	0	—	75k	72.616	13	0
SQRT	*	*	0	0	0	*	0	0	1/2	10k	74.447	13	0
SQRT	*	*	0	0	0	*	0	0	1/2	50k	71.292	13	0
SQRT	*	*	0	0	0	*	0	0	1/2	75k	64.863	13	0
CKLS	*	*	0	0	0	*	0	0	1.0	10k	64.989	13	0
CKLS	*	*	0	0	0	*	0	0	1.0	50k	56.097	13	0
CKLS	*	*	0	0	0	*	0	0	1.0	75k	48.468	13	0
SQRT0	*	*	0	0	*	*	0	0	1/2	10k	51.204	12	0
SQRT0	*	*	0	0	*	*	0	0	1/2	50k	32.515	12	.0012
SQRT0	*	*	0	0	*	*	0	0	1/2	75k	32.435	12	.0012
CKLS0	*	*	0	0	*	*	0	0	1.0	10k	41.211	12	0
CKLS0	*	*	0	0	*	*	0	0	1.0	50k	40.962	12	0
CKLS0	*	*	0	0	*	*	0	0	1.0	75k	41.459	12	0
SQRT-SV	*	*	0	*	0	*	*	0	1/2	10k	19.475	11	.0531
SQRT-SV	*	*	0	*	0	*	*	0	1/2	50k	18.463	11	.0714
SQRT-SV	*	*	0	*	0	*	*	0	1/2	75k	18.267	11	.0756
SQRT0-SV	*	*	0	*	*	*	*	0	1/2	10k	10.652	10	.3853
SQRT0-SV	*	*	0	*	*	*	*	0	1/2	50k	16.770	10	.0796
SQRT0-SV	*	*	0	*	*	*	*	0	1/2	75k	14.192	10	.1644
SQRT0-SV	*	*	0	*	*	*	*	0	1/2	75$k\times$ 2	14.288	10	.1603
CKLS0-SV	*	*	0	*	*	*	*	0	1.0	10k	24.189	10	.0071
CKLS0-SV	*	*	0	*	*	*	*	0	1.0	50k	17.008	10	.0742
CKLS0-SV	*	*	0	*	*	*	*	0	1.0	75k	16.340	10	.0903
CKLS0-SV-FB	*	*	*	*	*	*	*	*	1.0	10k	9.556	8	.2976
CKLS0-SV-FB	*	*	*	*	*	*	*	*	1.0	50k	9.833	8	.2769
CKLS0-SV-FB	*	*	*	*	*	*	*	*	1.0	75k	9.544	8	.2985

NOTE: * denotes a free parameter; — means not relevant. 75k × 2 denotes a simulation of length N = 75,000 weeks simulated at $1/\Delta$ = 28 steps per week; all others are simulated at 14 steps per week.

These appealing features notwithstanding, both the SQRT and SQRT0 specifications fare poorly when confronted with postwar U.S. interest rate data. Aït-Sahalia (1996a), Chan et al. (1992), Conley, Hansen, Luttmer, and Scheinkman (1997), and Tauchen (1997a) estimated single-factor diffusion models and uncovered evidence that the diffusion function is not well accommodated with $\gamma = \frac{1}{2}$, although values $\gamma > \frac{1}{2}$ in the CKLS specification can improve the fit.

Apart from OU, SQRT, and SQRT0, none of the aforementioned specifications is estimable by maximum likelihood, because the transition density of y_t is not available in closed form. Methods developed by Aït-Sahalia (1996a) and Hansen and Scheinkman (1995) do not entail simulation and apply directly when the interest rate is a scalar diffusion, as in the CHKLS and CHKLS0 specifications. These methods can also handle certain forms of partial observability (random time-scale changes due to subordination), with separate methods used to estimate the parameters of the subordinating process (Conley, Hansen, and Liu 1997).

Anderson and Lund (1996, 1997) and Tauchen (1997a) have used the EMM technique to estimate diffusion models for the short-term interest rate. Using weekly observations on the 30-day Eurodollar rate, Tauchen (1997a) estimated scalar diffusions and found that a model similar to the CKLS specification with $\gamma = 1$ performed far better than other single-factor models but still did not pass the omnibus chi-squared test defined in (9). The t-ratio diagnostics defined in (11) suggest that the specification fails to accommodate the tail behavior (conditional leptokurtosis) in movements in short-term interest rates. Andersen and Lund (1996, 1997) introduced unobserved stochastic volatility into the specification and found great improvement in the fit. Using weekly observations on the 3-month Treasury bill rate, they found that models very close to the SQRT-SV specification can pass the chi-squared test at conventional levels and provide substantial improvement on the t-ratio diagnostics, although not all t-ratios are reduced to insignificance.

The CKLS0-SV-FB specification, which allows for feedbacks from the level of the interest into volatility dynamics, is new to this article. Much of the literature on modeling the short-term rate tries to capture the relationship between the level of the short-term interest rate and its variability. Here we consider the simple expedient of entering the interest rate U_1 directly into the drift and diffusion functions of the latent volatility process U_2, while the diffusion function of U_1 is linear in U_1.

In what follows we estimate each of these models by EMM and use reprojection to represent the implied transition density of the observed process given its past. We then use this representation to elucidate model characteristics and, among other things, generate new diagnostics for estimated diffusions.

The data are 1,809 weekly observations, from January 5, 1962 to August 30, 1996, on the 3-month Treasury bill

rate from the secondary market. Friday rates are used except when unavailable due to a holiday, in which case the Thursday rate is used. Secondary market rates are annualized using a 360-day year and quoted on a discount basis. These data are plotted in Figure 1, with descriptive statistics shown in Table 1, and are available from STATLIB.

The theoretically appropriate interest rate is an idealized instantaneous rate applicable to a loan over an infinitesimally small interval. There is no such rate in practice. Rates on overnight, 7-day, and 30-day loans are quoted on instruments that occasionally are thinly traded and display large movements due to short-term liquidity effects unrelated to market fundamentals. A common approach is to use the 3-month rate as a proxy for the idealized instantaneous rate, as we do here. Although these data are available at daily frequency, we use a weekly sampling rate because daily data are so highly correlated that little statistical efficiency can be gained by going from weekly to daily frequency, and bias due to statistical corrections to location and scale for day-of-the-week and weekend effects is avoided when weekly data are used.

The first step is to determine the appropriate specification of the SNP score generator. In doing so, the first 26 observations are reserved as the provision for initial lags, thus permitting $L \leq 26$, and $t = -26, \ldots, 1{,}783 = n$ indexes the complete dataset. Following the protocol described in Section 2, we find that $L_u = 1, L_r = 4, L_p = 1, K_z = 4$, and $K_x = 1$ is the preferred specification of the score generator model. (I_z and I_x are irrelevant.) There are $p_K = 16$ components of the score vector.

To implement the simulator for the EMM estimation, the week is divided into 14 subintervals, and an explicit order 2 weak scheme (Kloeden and Platen 1992, pp. 486–487) is used. To generate weekly data, every 14th simulated value is retained. (For SQRT0-SV, as a check on sensitivity, simulations at 28 steps per week were used as well.) Simulation length N (after discarding transients) ranges from 10,000 up to 75,000 retained values. To eliminate transients, 5,000 leading simulated values were discarded for $N \geq 50{,}000$; 500 were discarded for $N = 10{,}000$.

Our simulation schemes are related to linear stochastic differential equations (SDEs) and to theorem 1 of Gallant and Long (1997) as follows. Two main classes of simulation schemes are available: strong and weak. With linear interpolation between the points $\hat{U}_0, \hat{U}_\Delta, \hat{U}_{2\Delta}, \ldots, \hat{U}_{N+L}$, defined by an SDE such as (22), an order-δ strong scheme satisfies $\mathcal{E}(\sup_{0\leq t\leq N+L} |U_t - \hat{U}_t|) \leq B\Delta^\delta$ for some B under regularity conditions (Kloeden and Platen 1992, cor. 10.6.5). An order-δ weak scheme satisfies $\max_{t=0,\Delta,\ldots,N+L} |\mathcal{E}g(U_t) - \mathcal{E}g(\hat{U}_t)| \leq B_g\Delta^\delta$ for some B_g under regularity conditions (Kloeden and Platen 1992, sec. 14.5). An order-δ strong scheme is an order-2δ weak scheme. We use a weak scheme because a strong scheme simulation takes much longer to compute for multivariate system; the difference in run times is two orders of magnitude for nonsparse versions of (22). There is some latitude in the implementation of these schemes. Our implementation is available as an algebraic display in earlier work (Gallant and Tauchen 1996b) and as Fortran code from STATLIB. A discretely sampled linear SDE, such as OU, has a transition density that is a discrete-time autoregression. Thus a discretization scheme can be devised that will simulate from it exactly. Our scheme coincides with this discretization in the scalar case. One may view the use of these schemes as a matter of numerical analysis: A choice of N and Δ that will compute $m_n(\rho, \theta)$ sufficiently accurately can be found with a little common sense and experimentation. Under this view, theorem 1 of Gallant and Long (1997) applies without alteration. Another view is that the SDE is just an abstraction, and it is the scheme used together with the specific choice of Δ that actually defines the nonlinear model under consideration. This is the most conservative view, and with it theorem 1 applies without alteration. A third view is that Monte Carlo noise and discretization bias must be taken into account and some version of theorem 1 must be proved that has K and N tending to infinity and Δ tending to 0 in a specific relationship. This is well beyond the scope of this article. Strong schemes lend themselves better to this sort of analysis. Some results in this direction have been given by Duffie and Glynn (1995).

Table 2 shows minimized chi-squared criterion from estimation by EMM of the various specifications just discussed. (The line $N = 75k \times 2$ for SQRT0-SV reports the simulations at 28 steps per week.) Consistent with some previous studies, it proved very difficult to estimate γ as a free parameter for many of the specifications, so the table just shows results for the CKLS specifications with γ set to unity.

Table 2, shows that first, a simulation size of $N = 10{,}000$ is too small for reliable results (although useful for getting starting values), whereas the computations stabilize at $N = 50{,}000$. Second, as expected, the OU and SQRT specifications are sharply rejected. Third, SQRT0 and CKLS0 do much better, indicating the importance of the intercept β_{10} in the diffusion specification. Nonetheless, both SQRT0 and CKLS0 are still sharply rejected and, as will be seen, provide poor models of the interest rate. Fourth, introduction of the second stochastic volatility factor dramatically improves the fit, and in the case of SQRT0-SV and CKLS0-SV-FB, the chi-squared statistic is insignificant at conventional levels. This improvement in fit is consistent with the findings of Aït-Sahalia (1996b), who provided nonparametric evidence on the failure of scalar diffusion specifications of the interest rate.

Table 3 shows parameter estimates along with standard errors and the lower and upper limits of 95% confidence intervals computed by (approximately) inverting the criterion difference test (10) as described at the end of Section 2.2. (In Table 3 and all subsequent tables and figures, the fits from Table 2 for the line with $N = 75k$ are used, except for in SQRT0-SV, where line $N = 75k \times 2$ is used.) The standard errors are based on the asymptotic distribution (8) and provide the usual symmetric two-sigma confidence intervals of the Wald theory. The criterion-difference confidence intervals, on the other hand, reflect the asymmetry of the objective function. Generally, the criterion-difference confidence intervals are somewhat tighter than the two-sigma Wald intervals. In some cases, and in particular for the es-

Table 3. Parameter Estimates, Standard Errors, and Criterion-Difference Confidence Intervals

Specification	α_{10}	α_{11}	α_{21}	α_{22}	β_{10}	β_{11}	β_{20}	β_{21}	γ
OU									
Estimate	.026	−.005			.086				
Standard error	.001	.001			.002				
95% lower	.024	−.005			.085				
95% upper	.027	−.004			.090				
SQRT									
Estimate	.030	−.006				.037			.50
Standard error	.004	.001				.002			
95% lower	.023	−.006				.034			
95% upper	.030	−.004				.037			
SQRT0									
Estimate	.306	−.042			1.742	−.583			.50
Standard error	.006	.001			.045	.022			
95% lower	.306	−.042			1.719	−.583			
95% upper	.308	−.042			1.742	−.573			
SQRT0-SV									
Estimate	.274	−.059		−33.161	.094	−.046	8.230		.50
Standard error	.163	.037		71.413	.416	.206	20.178		0
95% lower	.219	−.071		−41.253	.047	−.069	6.044		0
95% upper	.318	−.045		−25.424	.139	−.023	10.507		0
CKLS0-SV-FB									
Estimate	.014	−.002	−.006	−.157	.043	−.018	.593	−.052	1.00
Standard error	.001	.001	.003	.021	.008	.003	.028	.005	
95% lower	.013	−.002	−.006	−.163	.041	−.018	.587	−.053	
95% upper	.014	−.002	−.005	−.157	.043	−.017	.601	−.050	

timate of α_{11} in the SQRT0-SV specification, a two-sigma Wald interval extends well into the unstable region but the criterion-difference confidence interval does not.

Some of the parameter estimates are interpretable. For example, for the drift, α_{10}/α_{11} defines a point where the drift in U_1 vanishes and, in simpler models, defines a steady-state value of the interest rate. For the diffusion, if β_{10} and β_{11} are of opposite sign, then the value $(-\beta_{10}/\beta_{11})^{1/\gamma}$ defines an interest rate at which the diffusion in U_1 vanishes, which is a reflecting barrier that reflects either to the left or right depending on the sign of the drift function evaluated at the barrier. For the fitted SQRT0 model, the barrier is at an interest rate of 8.94% $(= -\beta_{10}/\beta_{11})^2$ and the drift is negative there, so the reflection is to the left (inspection of simulations bears this out). Placement of a left-reflecting barrier at this point partly explains why this model does so poorly on the chi-squared test and why it is not a usable model. On the other hand, for SQRT0-SV there is a right-reflecting barrier at 4.11%, which is within the range of the data but not unreasonably high. For CKLS0-SV-FB, the right-reflecting barrier is at 2.44% $(= -\beta_{10}/\beta_{11})$, which is just outside the range of the data and more reasonable.

Table 4. Diagnostic t Statistics

Coefficient	SQRT	SQRT0-SV	CKLS0-SV-FB
Location function			
b_0	.77	1.73	.82
b_1	−.65	−.07	−.80
Scale function			
r_0	1.44	−.09	1.50
r_1	2.77	.02	.58
r_2	2.64	−1.07	.13
r_3	2.23	−.44	.19
r_4	3.88	.83	1.71
Hermite polynomial			
$a_{1,0}$	.93	−.52	−.22
$a_{0,1}$	1.46	1.56	1.14
$a_{1,1}$	−.79	−.05	.03
$a_{0,2}$	3.28	1.13	1.00
$a_{1,2}$	−.47	−1.42	−.38
$a_{0,3}$	2.68	1.21	.19
$a_{1,3}$	−2.15	−.22	−.13
$a_{0,4}$	6.36	1.64	1.74
$a_{1,4}$	−2.01	−1.74	−.66

Table 4 gives the t-ratio diagnostics for the SQRT, SQRT0-SV, and CKLS0-SV-FB specifications. As to be expected from the large value of the chi-squared statistic in Table 2, many of the t ratios for the SQRT specification are well above 2.0 in magnitude, especially in the Hermite component of the score vector. This suggests that the transition density implied by the SQRT model does not have the appropriate shape. The large t ratios for the scale function suggest that conditional volatility is not well approximated. On the other hand, the low statistics for the location function suggest that the conditional mean is adequately approximated by the SQRT model. For the SQRT0-SV and CKLS0-SV-FB specifications, all t ratios are below 2.0 in magnitude.

Figure 2 shows volatility scatterplots, which are plots of $\tilde{y}_t - \tilde{y}_{t-1}$ against $\tilde{y}_{t-1}$. For the raw data (Fig. 2a), 1,808 points are plotted. For the SQRT, SQRT0-SV, and CKLS0-SV-FB specifications (Figs. 2b–2d), 74,999 simulated values are plotted. One must remain aware of the different series length when comparing the scatter of any fit to that of the raw data, which is inherently more sparse. The SQRT specification fails to generate enough large movements in interest rates. In contrast, the SQRT0-SV specification generates

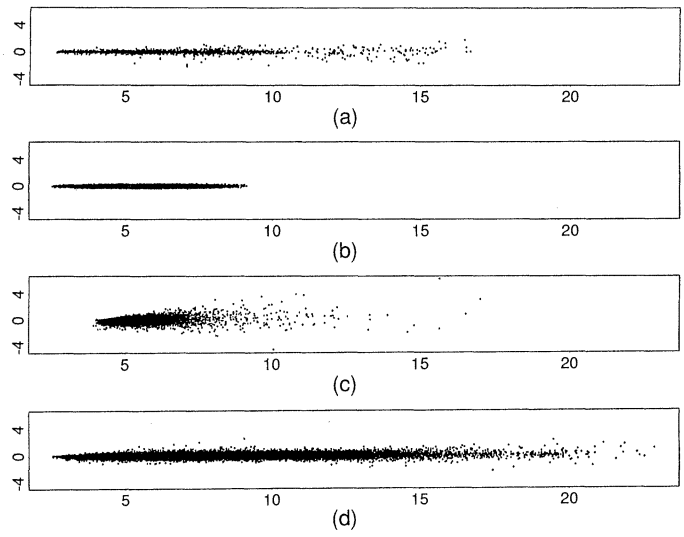

Figure 2. Volatility Scatterplots for $y_t - y_{t-1}$ Versus y_{t-1}. (a) The data, 1,808 points; (b) the SQRT model, 74,999 points; (c) the SQRT0-SV model, 74,999 points; (c) the CKLS0-SV-FB model, 74,999 points.

many more larger movements but fails to generate enough realizations at higher rates. The CKLS0-SV-FB specification appears to better capture the features of the data, provided that one bears in mind that Figure 2a has 1,808 points, whereas 2c has 74,999 points.

Figure 3 shows one-step-ahead conditional volatility computed on the observed data. The Figures 3a and 3b show volatility obtained directly from the data by taking the square root of a moving average of squared residuals, $\{(m+1)^{-1}\sum_{j=0}^{m}\hat{e}_{t-j}^2\}^{1/2}, t = 0,\ldots,1{,}808-m-1$, from estimation of the AR(1) model $y_t = \psi_0 + \psi_1 y_{t-1} + e_t$. Figure 3a is for $m = 4$, and 3b is for $m = 26$. Figure 3c shows the projected volatility from the SNP score generator.

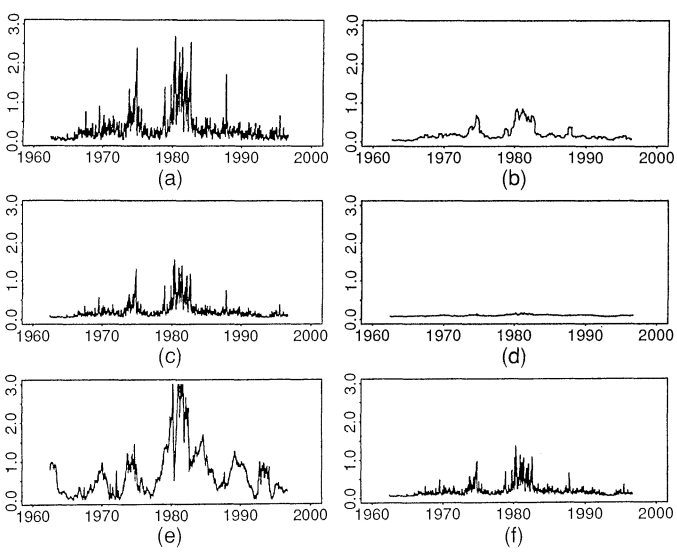

Figure 3. Time Series Plots of One-Step-Ahead Volatility. Plotted is the conditional standard deviation computed as follows: (a) from equally weighted MA(4) of squared AR(1) residuals; (b) from equally weighted MA(26) of squared AR(1) residuals; (c) projected volatility from the SNP score generator; (d) the reprojected volatility from the SQRT specification; (e) the reprojected volatility from the SQRT0-SV specification; and (f) the reprojected volatility from the CKLS0-SV-FB specification.

Projected volatility is the one-step-ahead conditional standard deviation of the SNP projection, and it appears to be a reasonable compromise between 3a and 3b. Figures 3d, 3e, and 3f show reprojected volatility (14) for three estimated specifications of the interest rate diffusion: SQRT, SQRT0-SV, and CKLS0-SV-FB. The SQRT specification (Fig. 3d) appears to over-smooth conditional volatility, whereas on the other hand, the SQRT0-SV (3e) specification overstates fluctuations in volatility. Interestingly, the CKLS0-SV-FB (3f) specification closely mimics the general characteristics of the SNP projection.

We now investigate shape characteristics of the reprojected transition densities (13) of the SQRT, SQRT0-SV, and CKLS0-SV-FB specifications. We want to condition on histories located in various quantiles of the observed data. For this purpose, let t_α be the smallest value of the index t for which $\tilde{y}_{t_\alpha-1}$ equals the αth quantile of the observed data $\{\tilde{y}_t\}_{t=-26}^{1{,}783}$. Figure 4 shows a reprojected transition density $\hat{f}_K(y_t|\tilde{y}_{t_\alpha-4},\ldots,\tilde{y}_{t_\alpha-1})$, as defined in (13), standardized to have mean 0 and unit variance along with a reference standard normal density shown as a dashed line. Figures 4a and 4b show reprojected transition densities for the SQRT specification at $\alpha = .25$ and $\alpha = .75$. Evidently, the transition density is close to the Gaussian, which highlights a chief empirical shortcoming of the basic square-root model. (Inspection of the Hermite coefficients of the reprojection reveal departures from the Gaussian, as should be the case; however, the plots indicate that these departures are quite mild.) Figures 4c and 4d show the reprojected transition density for the SQRT0-SV specification, also at $\alpha = .25, .75$. Activation of the unobserved stochastic volatility process introduces some leptokurtosis into the transition density, which explains the improvement in fit.

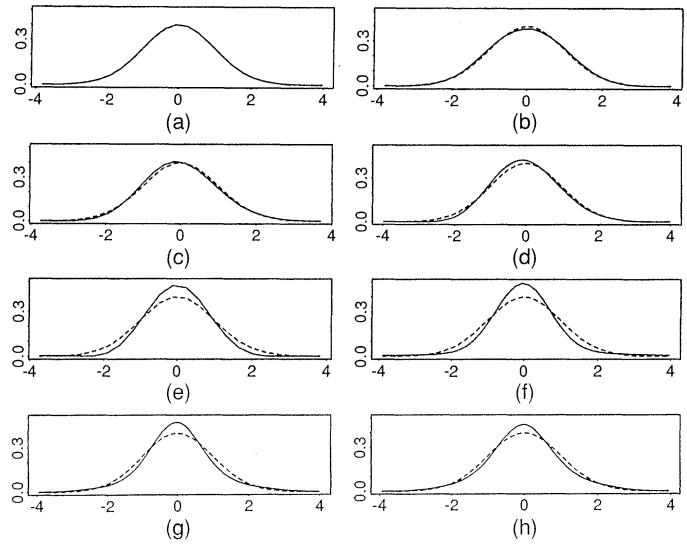

Figure 4. Reprojected Transition Densities. The solid line in (a) is the reprojected transition density for the SQRT specification conditional on being at the 25% quantile of the data. The density has been standardized to have mean zero and variance one. The dashed line is a standard normal density. (b) The same for the 75% quantile; (c) and (d) the same for the SQRT0-SV specification conditional on the 25% and 75% quantiles of the data; (e), (f), (g), and (h) the same for the CKLS0-SV-FB specification conditional on the 5%, 25%, 75%, and 95% quantiles.

Because the support of the unconditional density implied by the SQRT0-SV specification does not fully cover that of the observed data (see Fig. 2), the reprojected transition density for this specification is not well determined conditional on histories $(\tilde{y}_{t_\alpha-4},\ldots,\tilde{y}_{t_\alpha-1})$ with α near 0 or unity. On the other hand, the reprojected transition density for CKLS0-SV-FB is well determined for extreme α. Figures 4e–4h show the reprojected transition density for the CKLS0-SV-FB specification at $\alpha = .05, .25, .75, .95$. The transition density is seen to be leptokurtic relative to the normal density.

Following ideas set forth in Section 2.3, additional insight into one-step-ahead dynamics is available from study of plots of conditional moments against perturbations to the initial conditions. Because all systems considered here are linear in the mean with extreme persistence, one-step-ahead mean dynamics are not interesting. However, one-step-ahead volatility dynamics are quite interesting and reveal further differences in the characteristics of the estimated diffusions.

The volatility structure of U.S. short-term interest rates is known to be quite complicated, as documented and discussed at length by Andersen and Lund (1997) and Brenner, Harjes, and Kroner (1994). These authors pointed out that interest rate volatility contains an ARCH-like component, where recent unforeseen movements in interest rates increase conditional volatility. This component is a common characteristic of financial markets. Unforeseen financial movements in either direction tend to raise volatility, although not necessarily symmetrically (Bollerslev and Engle 1993; Gallant et al. 1993; Nelson 1991). But the short-term interest rates also display a second component known as the "levels effect," in which the higher the overall level of the interest rate, the higher the volatility. Figure 5 plots reprojected one-step-ahead conditional volatility (16) against both anticipated and unanticipated shocks to investigate how well the various diffusion models can accommodate both the levels and the ARCH components of conditional volatility.

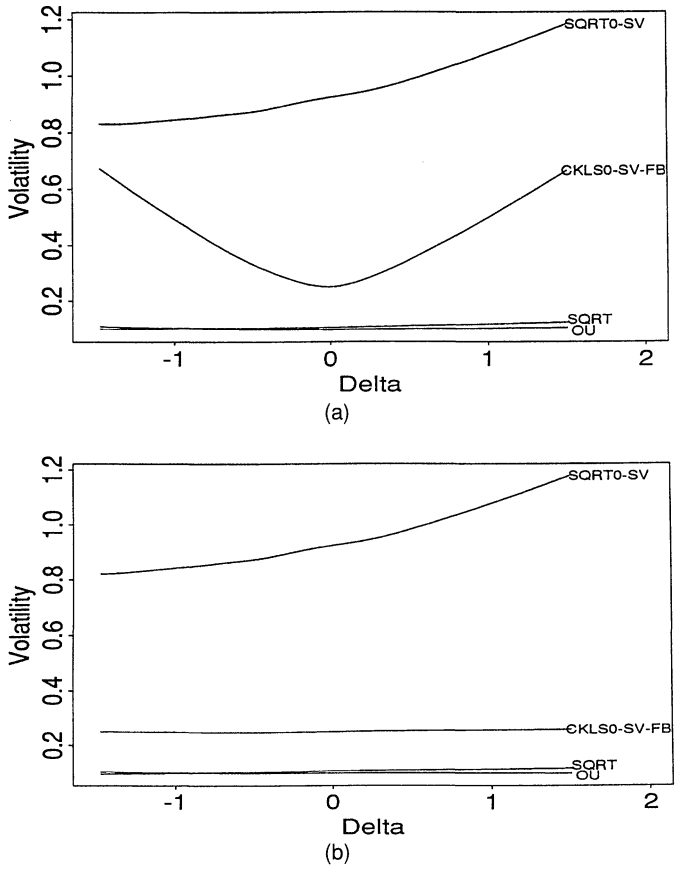

Figure 5. One-Step-Ahead Conditional Volatility. The four lines in (a) show the response to unanticipated movements in the interest rate for the OU, SQRT, SQRT0-SV, and CKLS0-SV-FB specifications. The four lines in (b) show the response to anticipated movements in the interest rate for the OU, SQRT, SQRT0-SV, and CKLS0-SV-FB specifications.

Figure 5a shows plots of

$$\left[(1/n)\sum_{t=0}^{n}\operatorname{var}(y_t|\tilde{y}_{t-4},\ldots,\tilde{y}_{t-1}+\delta)\right]^{1/2}$$

against δ for four estimated diffusions computed from the reprojected density as defined in (13). Because only the most recent value of the interest rate is perturbed, the movement δ is to be interpreted as unforeseen relative to the history of the interest rate. Also, because $\{y_t\}$ is a scalar process, there is no distinction between δ and Δ, as in Section 2.3.

The response for the OU specification is perfectly flat, whereas that for SQRT is slightly upward sloping. These patterns are the only possible ones for these two models. The responses for SQRT0-SV and CKLS0-SV-FB are more interesting, as they show a more U-shaped pattern (although not symmetric), consistent with an ARCH effect.

We now consider systematic movements in the entire history of the interest rate, not just the most recent value. Figure 5b shows plots of

$$\left[(1/n)\sum_{t=0}^{n}\operatorname{var}(y_t|\tilde{y}_{t-4}+\delta,\ldots,\tilde{y}_{t-1}+\delta)\right]^{1/2}$$

against δ for the same four estimated diffusions. Now the movement δ is to be interpreted as a foreseen movement of the interest rate. For OU, SQRT, and SQRT0-SV, there is hardly any difference in the responses relative to those of the upper panels, which reflect empirical shortcomings of these specifications. For CKLS0-SV-FB, there is a sharp contrast in the responses. Comparing Figures 5a and 5b, one sees that CKLS0-SV-FB volatility responds in the ARCH-like and slightly asymmetric manner to the unforeseen movement and responds monotonically and similarly to the SQRT model to the foreseen movement. Apparently, the feedbacks built into the CKLS0-SV-FB specification are needed to separate ARCH and levels components of interest rate volatility.

Finally, we consider multistep dynamics by examining reprojected profile bundles, as defined by (20) and (21). Table 5 shows the half-lives of reprojected mean profile bundles (20) for the OU, SQRT, SQRT0-SV, and CKLS0-SV-FB specifications. Table 6 shows the half-lives of reprojected volatility profile bundles (21) for the same specifications. To prevent plots (not shown) from being excessively dense,

profiles were computed at every tenth datum rather than at each datum. As indicated in Table 5, the OU, SQRT, and CKLS0-SV-FB specifications exhibit the extreme persistence known to be present in interest rate data, whereas the SQRT0-SV specification does not. The half-life of conditional mean profile bundles for the SQRT0-SV specification is 8.5 weeks, compared to half-lives of more than 2 years for the other three models. These half-lives are generally consistent with the point estimates of the parameter α_{11} shown in Table 2. The conditional standard deviation half-lives in Table 6 are quite interesting and show the sharp contrast between the implied dynamics of diffusion models without and with stochastic volatility, and thereby help reveal why the latter fit the data so much better. Interestingly, the CKLS0-SV-FB specification shows sensitivity of volatility to initial conditions that is attenuated relative to that of SQRT0-SV (consistent with Fig. 5) but is much more persistent, with a hard-life of 49.6 weeks, compared to a half-life of 10.6 weeks for SQRT0-SV.

The low persistence results reported for SQRT0-SV do not appear to be attributable to point estimates obtained from a local rather than global minimum of the chi-squared objective function. Similar results were obtained when different simulation increments for the SQRT0-SV specification were used (Table 2) and when the data were extended for all specifications. Higher persistence results would be expected over the period January 5, 1992 to March 31, 1995 (Gallant and Tauchen 1996c), but results reported here are for data extended to August 30, 1996.

4. CONCLUSION

Reprojection is a general-purpose technique for estimating the transition dynamics of the observed variables of an estimated nonlinear dynamic model with latent variables, which are usually due to partial observation of the state vector. Understanding these dynamics is essential for diagnostic analysis and prediction. The key idea is to use a nonparametric series estimator to reestimate the transition dynamics from a long simulation of the estimated model. Using a nonparametric estimator in reprojection is essential for the technique to have generality. To our knowledge, there is no other all-purpose way to determine these transition dynamics, although in some circumstances specialized strategies might work.

We have applied the technique for diagnostic evaluation of continuous-time diffusion models for the short-term interest rate. We set forth additional evidence consistent with work of Aït-Sahalia (1996b) and Andersen and Lund (1996, 1997) that the short-term U.S. interest rate is not well modeled by a scalar diffusion process and is better modeled as a component of a larger system with unobserved stochastic volatility. We also presented evidence on the importance of an intercept term in the diffusion function and introduced a new continuous-time model that permits feedbacks from the interest rate into the drift and diffusion functions of stochastic volatility. This new specification does quite well across a broad range of diagnostic assessments.

Table 5. Half-Lives of Reprojected Mean Profile Bundles

	OU	SQRT	SQRT0-SV	CKLS0-SV-FB
Minimum at $j = -1$	2.63	2.63	2.63	2.63
Maximum at $j = -1$	15.70	15.70	15.70	15.70
Range at $j = -1$	13.07	13.07	13.07	13.07
Half-life (weeks)	179.5	128.7	8.5	264.7
Standard error	1.0	.3	.2	4.6

Table 6. Half-Lives of Reprojected Mean Profile Bundles

	OU	SQRT	SQRT0-SV	CKLS0-SV-FB
Minimum at $j = 0$	.091	.072	.012	.053
Maximum at $j = 0$	.103	.155	3.05	1.039
Range at $j = 0$	.013	.083	3.037	.986
Half-life (weeks)		152.8	10.6	49.6
Standard error		19.1	.3	17.6

[Received July 1996. Revised August 1997.]

REFERENCES

Aït-Sahalia, Y. (1996a), "Testing Continuous-Time Models of the Spot Interest Rate," *The Review of Financial Studies*, 9, 385–426.

——— (1996b), "Do Interest Rates Really Follow Continuous Time Markov Diffusions?" unpublished manuscript, University of Chicago, Graduate School of Business.

Andersen, T. G., Chung, H.-J., and Sorensen, B. E. (1997), "Efficient Method of Moments Estimation of a Stochastic Volatility Model: A Monte Carlo Study," unpublished manuscript, Northwestern University, Kellogg School of Management.

Andersen, T. G., and Lund, J. (1997), "Estimating Continuous Time Stochastic Volatility Models of the Short-Term Interest Rate," *Journal of Econometrics*, 77, 343–378.

——— (1996), "Stochastic Volatility and Mean Drift in the Short-Rate Diffusion: Sources of Steepness, Level, and Curvature of the Yield Curve," unpublished manuscript, Northwestern University, Kellogg School of Management.

Bollerslev, T., and Engle, R. F. (1993), "Common Persistence in Conditional Variances," *Econometrica*, 61, 167–186.

Brenner, R. J., Harjes, R. H., and Kroner, K. F. (1994), "Another Look at Alternative Models of the Short-Term Interest Rate," unpublished manuscript, University of Arizona, Dept. of Economics.

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

Chumacero, R. (1997), "Finite-Sample Properties of the Efficient Method of Moments," unpublished manuscript, University of Chile, Dept. of Economics.

Conley, T., Hansen, L. P., Luttmer, E., and Scheinkman, J. A. (1997), "Estimating Subordinated Diffusions From Discrete Time Data," *Review of Financial Studies*, 10, 525–578.

Conley, T., Hansen, L. P., and Liu, W-F. (1997), "Bootstrapping Diffusion Models," *Macroeconomic Dynamics*, 1, 279–311.

Cox, J. C., Ingersoll, J. E., Jr., and Ross, S. A. (1985), "A Theory of the Term Structure of Interest Rates," *Econometrica*, 53, 385–407.

Doan, T., Litterman, R., and Sims, C. (1984), "Forecasting and Conditional Projection Using Realistic Prior Distributions," *Econometric Reviews*, 3, 1–100.

Duffie, D., and Glynn, P. (1995), "Efficient Monte Carlo Simulation of Security Prices," *Annals of Applied Probability*, 5, 897–906.

Duffie, D., and Kan, R. (1996), "A Yield-Factor Model of Interest Rates," *Mathematical Finance*, 6, 379–406.

Duffie, D., and Singleton, K. J. (1993), "Simulated Moments Estimation of Markov Models of Asset Prices," *Econometrica*, 61, 929–952.

——— (1994), "Econometric Modeling of Term Structures of Defaultable Bonds," unpublished manuscript, Stanford University, Graduate School of Business.

Ellner, S., Gallant, A. R., and Theiler, J. (1995), "Detecting Nonlinear-

ity and Chaos in Epidemic Data," in *Epidemic Models: Their Structure and Relation to Data*, ed. D. Mollison, Cambridge, U.K.: Cambridge University Press, pp. 229–247.

Engle, R. F., Ito, T., and Lin, W-L. (1990), "Meteor Showers or Heat Waves? Heteroskedastic Intra-Daily Volatility in the Foreign Exchange Market," *Econometrica*, 58, 525–542.

Fenton, V. M., and Gallant, A. R. (1996), "Qualitative and Asymptotic Performance of SNP Density Estimators," *Journal of Econometrics*, 74, 77–118.

Gallant, A. R., and Long, J. R. (1997), "Estimating Stochastic Differential Equations Efficiently by Minimum Chi-Squared," *Biometrika*, 84, 125–141.

Gallant, A. R., and Nychka, D. W. (1987), "Seminonparametric Maximum Likelihood Estimation," *Econometrica*, 55, 363–390.

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

——— (1993), "Nonlinear Dynamic Structures," *Econometrica*, 61, 871–907.

Gallant, A. R., and Tauchen, G. (1989), "Seminonparametric Estimation of Conditionally Constrained Heterogeneous Processes: Asset Pricing Applications," *Econometrica*, 57, 1091–1120.

——— (1992), "A Nonparametric Approach to Nonlinear Time Series Analysis: Estimation and Simulation," in *New Dimensions in Time Series Analysis, Part II,* eds. E. Parzen, D. Brillinger, M. Rosenblatt, M. Taqqu, J. Geweke, and P. Caines, New York: Springer-Verlag, pp. 71–92.

——— (1996a), "Which Moments to Match?," *Econometric Theory*, 12, 657–681.

——— (1996b), "Specification Analysis of Continuous Time Models in Finance," in *Modeling Stock Market Volatility: Bridging the Gap to Continuous Time*, ed. P. E. Rossi, New York: Academic Press, pp. 357–384.

——— (1996c), "Reprojecting Partially Observed Systems With Application to Interest Rate Diffusions from January 5, 1992, to March 31, 1995," unpublished manuscript, University of North Carolina, Dept. of Economics.

Gennotte, G., and Marsh, T. A. (1993), "Variations in Economic Uncertainty and Risk Premiums on Capital Assets," *European Economic Review*, 37, 1021–1041.

Ghysels, E., Harvey, A., and Renault, E. (1995), "Stochastic Volatility," in *Handbook of Statistics, Vol. 14, Statistical Methods in Finance*, ed. G. S. Maddala, Amsterdam: North-Holland.

Gourieroux, C., and Monfort, A. (1996), *Simulation-Based Econometric Methods*, New York: Oxford University Press.

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

Hansen, L. P., and Scheinkman, J. A. (1995), "Back to the Future: Generating Moment Implications for Continuous-Time Markov Process," *Econometrica*, 50, 1029–1054.

Ingersoll, J. E., Jr. (1987), *Theory of Financial Decision Making*, Savage, MD: Rowman and Littlefield.

Ingram, B. F., and Lee, B. S. (1991), "Simulation Estimation of Time Series Models," *Journal of Econometrics*, 47, 197–250.

Jacquier, E., Polson, N. G., and Rossi, P. E. (1994), "Bayesian Analysis of Stochastic Volatility Models," *Journal of Business and Economic Statistics*, 12, 371–417.

Johnson, N. L., and Kotz, S. (1970), *Distributions in Statistics: Continuous Univariate Distributions-1*, New York: Wiley.

Karatzas, I., and Shreve, S. E. (1991), *Brownian Motion and Stochastic Calculus* (2nd ed.), Berlin: Springer-Verlag.

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

Lo, A. W. (1988), "Maximum Likelihood Estimation of Generalized Ito Process With Discretely Sampled Data," *Econometric Theory*, 4, 231–247.

Mallet, A., Mentré, F., Steimer, J-L., and Lokiec, F. (1988), "Nonparametric Maximum Likelihood Estimation for Population Pharmacokinetics, With Application to Cyclosporine," *Journal of Pharmacokinetics and Biopharmaceutics*, 16, 311–327.

Melino, A. (1994), "Estimation of Continuous Time Models in Finance," in *Advances in Econometrics, Sixth World Congress, Vol. II*, ed. C. A. Sims, Cambridge, U.K.: Cambridge University Press.

Nelson, D. B. (1991), "Conditional Heteroskedasticity in Asset Returns: A New Approach," *Econometrica*, 59, 347–370.

Olsen, L. F., and Schaffer, W. M. (1990), "Chaos Versus Noisy Periodicity: Alternative Hypotheses for Childhood Epidemics," *Science*, 249, 499–504.

Schwarz, G. (1978), "Estimating the Dimension of a Model," *The Annals of Statistics*, 6, 461–464.

Sims, C. A. (1980), "Macroeconomics and Reality," *Econometrica*, 48, 1–48.

Smith, A. A. (1993), "Estimation Nonlinear Time Series Models Using Simulated Vector Autoregressions," *Journal of Applied Econometrics*, 8, S63–S84.

Tauchen, G. (1997a), "New Minimum Chi-Square Methods in Empirical Finance," in *Advances in Econometrics, Seventh World Congress*, eds. D. Kreps and K. Wallis, Cambridge, U.K.: Cambridge University Press, pp. 279–317.

Tauchen, G. (1997b), "The Objective Function of Simulation Estimators Near the Boundary of the Unstable Region of the Parameter Space," submitted to *Review of Economics and Statistics*.

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