

New Minimum Chi-Square Methods in Empirical Finance

George Tauchen
Duke University

Initial Draft: June 1995
Last Revised: April 1996

Abstract

The paper reviews recently developed simulation-based minimum chi-square estimators for structural models. Particular attention is paid to selection of the auxiliary model that defines the GMM-type criterion used in the minimum chi-square estimation. Considerations of statistical efficiency and behavior under misspecification make a strong case for using a very flexible, nonparametric approach to select the auxiliary model. To avoid a numerically ill-behaved GMM criterion function, the dynamic stability of the auxiliary model must also be verified, though, interestingly, the dynamic stability of the structural model itself is automatically enforced and need not be imposed in estimation. The empirical application involves estimation of a single-factor diffusion model for the 30-day Eurodollar interest rate.

1 Introduction

1.1 Background

A structural financial model typically defines a stochastic data generator. Using the data generator, it is relatively easy to compute expectations of nonlinear functions, normally by simulation, but the likelihood function itself is intractable. This is the setup for Simulated Method of Moments (SMM), as set forth in Ingram and Lee (1991) and Duffie and Singleton (1993). SMM is the extension to the time series context of the simulation estimators of McFadden (1989) and Pakes and Pollard (1989).

This estimation context arises across a broad range of areas in finance, including estimation of models of market microstructure (Foster and Viswanathan, 1995), estimation of equilibrium asset pricing models (Gennotte and Marsh, 1993), and estimation of continuous time models of interest rates and stock prices, as proposed by Melino (1994) in his address for the Sixth World Congress in 1990. The essential feature of each of these applications is the presence of unobserved stochastic processes that enter the structural model nonlinearly. In the microstructure application, the latent processes comprise the random information flow to informed traders and the stochastic behavior of noise traders. For asset pricing, the latent processes are endowment processes along with taste and technology shocks. In continuous time work, the latent processes are the underlying continuous factor processes which may be only partially observed at discrete time intervals.

Recent research develops a new class of minimum chi-square estimators that provide a systematic way of developing moment conditions for estimation by simulation. One approach, developed in Bansal, Gallant, Hussey, and Tauchen (1993, 1995), and Gallant and Tauchen (1996), and termed Efficient Method of Moments (EMM), uses the score function of an auxiliary model, called the score generator, to define a criterion function for GMM estimation (Hansen, 1982). The other approach, developed by Smith (1990, 1993) and extended by Gouriéroux, Monfort, and Renault (1993), and termed Indirect Inference, uses the parameters of the auxiliary model itself to define the GMM criterion. In either approach, so long as the underlying structural model is correctly specified, the auxiliary model need not nest the structural model and in fact may be misspecified. Subject only to minimal regularity

and identifiability conditions, the estimators are still consistent and asymptotically normal. The score-based approach has some computational advantages, as it circumvents the need to refit the auxiliary model (evaluate the binding function) for each candidate value of the parameter vector, and it eliminates the need to estimate the Hessian matrix of the auxiliary model.

Using an auxiliary model to define the GMM criterion brings to bear on the task of estimation the accumulated body of knowledge regarding the statistical characteristics of the relevant data. This knowledge is acquired from the years of experience of many investigators using statistical models to describe the data. The statistical modeling effort typically provides a family of statistical models that are known to fit the data quite well. A case in point pertains to models in the ARCH/GARCH class (Engle, 1982; Bollerslev, Chou, and Kroner, 1992) which provide excellent first approximations to the conditional distributions of financial returns. Another is the evidence from Hamilton (1989) and the follow-up literature that Markov-switching models provide excellent descriptions of the nonlinear dynamics of macro aggregate series. ARCH/GARCH models and switching-regime models are candidate auxiliary models for any of the three classes of estimation mentioned above. ARCH-GARCH or switching models provide readily available statistical descriptions of the observed data in empirical microstructure research, in asset pricing contexts, and in estimation of diffusion models as well.

Applications of these estimators include Smith (1993), who uses a VAR auxiliary model for estimation of a structural macroeconomic model. Engle and Lee (1994) use a GARCH model with Gaussian errors as an auxiliary model for fitting a volatility diffusion to daily stock returns data. Pagan, Hall, and Martin (1995) likewise use a GARCH model with Gaussian errors as the auxiliary model to estimate a diffusion model for monthly interest rates. Andersen and Lund (1995) employ generalized E-GARCH models as auxiliary models in order to estimate continuous time stochastic volatility models for weekly interest rates. Gouriéroux and Monfort (1994, pp. 153-154) use an autoregressive model with a heteroskedastic error structure to estimate a model of geometric Brownian motion for monthly interest rate data. Buraschi (1994) uses an autoregressive model for fitting a model of the monthly term structure. Ghysels and Jasiak (1994) use a flexible, nonparametric auxiliary

model to estimate a time deformation model for stock returns and volume. Hsu and Kugler (1995) and Bansal, Gallant, Hussey, and Tauchen (1995) also use more flexible auxiliary models to estimate term structure models and exchange rate models, while Gallant, Hsieh, and Tauchen (1995) do so for stochastic volatility models.

This paper reviews these minimum chi-square estimators with a particular focus on the selection of the auxiliary model. Section 4 below examines several issues, including statistical efficiency, specification testing, behavior under misspecification, and numerical stability. The investigation indicates that appropriate choice of the auxiliary model must be based on nonparametric considerations, with particular care taken to ensure that the internal dynamics of the selected model are stable. Previous experience and prior knowledge can, of course, provide an indication of a preliminary candidate model. But one must go beyond such a candidate and actually find an auxiliary model that fully describes the data at hand. Only by using a flexible, nonparametric auxiliary model can one ensure that parameter estimates are fully efficient when the underlying structural model is true and that misspecification will be detected when the structural model is false.

The remainder of this paper proceeds as follows. Section 2 sets up the notation and estimation context. Section 3 reviews the minimum chi-square estimators. Section 4 examines in detail the various considerations entailed in selecting the auxiliary model. Section 5 is the empirical application.

2 Stochastic Data Generators

We set out some basic notation where a structural model defines a data generator. Throughout, $\{w_t\}$ denotes a vector process of exogenous forcing variables. The process $\{w_t\}$ is assumed to be strictly stationary and Markovian with transition density $q(w_t|w_{t-L_w}, \dots, w_{t-1}, \gamma)$, where γ is a parameter vector of length ℓ_γ and L_w is the lag length. In many circumstances, w_t is unobserved, though the treatment here is sufficiently general to handle the case when some elements of w_t are observed and others are latent. Let v_t denote a vector outcome process, which is given by the *Solution Function*:

$$v_t = h(v_{t-L_1}, \dots, v_{t-1}, w_{t-L_2}, \dots, w_{t-1}, w_t, \psi), \quad (2.1)$$

where L_1 and L_2 are the longest lags of v_t and w_t appearing in h and ψ is a parameter vector of length ℓ_ψ .

The conditional density q and the solution function h jointly define a data generation process. Set $\rho = (\gamma' \psi)'$, where ρ is of length $\ell_\rho = \ell_\gamma + \ell_\psi$ and contains all of the parameters. For a candidate value ρ of the parameter, one simulates artificial realizations as follows: Given initial values $\hat{w}_{-L_w+1}, \dots, \hat{w}_0$, generate a realization on the forcing process $\{\hat{w}_\tau\}$ by drawing \hat{w}_1 from $q(w_1|\hat{w}_{-L_w+1}, \dots, \hat{w}_0)$, \hat{w}_2 from $q(w_2|\hat{w}_{-L_w+2}, \dots, \hat{w}_1)$, and so forth. Given initial values $\hat{v}_{L_1+1}, \dots, \hat{v}_0$ and having let $\{\hat{w}_\tau\}$ run long enough so there are least L_2 values available, then v 's can be generated iteratively through the solution function (2.1) above.

It is assumed that $\{q(\cdot|\gamma), h(\cdot|\psi)\}_{\rho \in \mathcal{R}}$, $\mathcal{R} \subset \mathfrak{R}^{\ell_\rho}$, defines a strictly stationary Markov data generator. For each $\rho \in \mathcal{R}$, the simulated process $\{\hat{s}_\tau\}_{\tau=1}^N$ (given that ρ) is asymptotically stationary and ergodic with unconditional densities

$$\{p_J(s_{t-J}, s_{t-J+1}, \dots, s_t|\rho)\}_{J=0}^\infty$$

of stretches of length $J + 1$, $J = 0, 1, 2, \dots$

The upshot is simply that Monte Carlo numerical integration works: For a function $e(s_{t-J}, s_{t-J+1}, \dots, s_t)$ such that the expectation exists, then

$$\frac{1}{N} \sum_{\tau=J+1}^N e(\hat{s}_{\tau-J}, \hat{s}_{\tau-J+1}, \dots, \hat{s}_\tau) \xrightarrow{as} \mathcal{E}_\rho[e(s_{t-J}, s_{t-J+1}, \dots, s_t)],$$

where

$$\begin{aligned} \mathcal{E}_\rho[e(s_{t-J}, s_{t-J+1}, \dots, s_t)] = \\ \int e(s_{t-J}, s_{t-J+1}, \dots, s_t) p(s_{t-J}, s_{t-J+1}, \dots, s_t|\rho) d(s_{t-J}, s_{t-J+1}, \dots, s_t), \end{aligned} \quad (2.2)$$

and $\{\hat{s}_\tau\}_{\tau=1}^N$ is a simulated realization. Duffie and Singleton (1993) give more rigorous conditions for the data generator $\{q(\cdot|\gamma), h(\cdot|\psi)\}_{\rho \in \mathcal{R}}$, $\mathcal{R} \subset \mathfrak{R}^{\ell_\rho}$ to be asymptotically stationary and ergodic.

In many applications, the elements of the process s_t are not completely observed. Let

$$y_t = \phi(s_t) \quad (2.3)$$

be an $M \times 1$ function of s_t that is observed by the econometrician. If any elements of the forcing process enter the observed vector y_t , then the mapping from these elements of w_t to y_t is just the identity map.

Since y_t is a function of s_t , y_t is also stationary and ergodic, though not necessarily Markovian. Let

$$p_J(y_t|y_{t-J}, \dots, y_{t-1}, \rho) = \frac{p_J(y_{t-J}, \dots, y_t|\rho)}{p_{J-1}(y_{t-J}, \dots, y_{t-1}|\rho)} \quad (2.4)$$

denote the implied transition densities, $J = 1, 2, \dots$. A common presumption in much of time series econometrics is that the observed process is either Markovian, or nearly so, in the sense that $p_J(y_t|y_{t-J}, \dots, y_{t-1}) \approx p_L(y_t|y_{t-L}, \dots, y_{t-1})$, $J \geq L$, for some sufficiently large L .

Throughout, it is presumed the user has access to a good method to compute the solution function (2.1). Little attention is paid to that particular issue here, though it bears emphasizing that this is a very active research topic. Early work is discussed in Taylor and Uhlig (1990) and Tauchen and Hussey (1991); more recent work is reviewed in Judd (1994). The search for accurate and fast algorithms is well motivated, because experience suggests that most of the computer time in implementing simulation estimators is associated with generating the v_t , not with the calculations directly associated with the optimization.

3 Estimators

The notational convention in what follows is this: y_t is a subvector of a strictly stationary Markov process s_t generated by a stochastic model $\{q(\cdot|\gamma), h(\cdot|\psi)\}_{\rho \in \mathcal{R}}$, $\rho = (\gamma' \psi)'$, as considered in Section 2. $p_J(y_{t-J}, \dots, y_t|\rho)$ denotes the implied joint density given ρ of a stretch of length $J + 1$, $J = 0, 1, 2, \dots$, and $\{\tilde{y}_t\}_{t=1}^n$ denotes the realization observed by the econometrician. The stochastic model is assumed to be correctly specified, meaning that the joint density of $\{\tilde{y}_t\}_{t=1}^n$ is $p_{n-1}(y_1, \dots, y_n|\rho_0)$ for all $n \geq 1$, where $\rho_0 \in \mathcal{R}$ is the true value. Below, we often refer to this stochastic model as simply the p -model.

The econometrician's task is to estimate ρ_0 . The sample likelihood function $\mathcal{L}(\rho) = p_{n-1}(\tilde{y}_1, \dots, \tilde{y}_n|\rho)$ is presumed intractable, rendering maximum likelihood estimation infeasible and motivating consideration of minimum chi-square estimation. The econometrician generates simulated realizations on $\{y_t\}$ given $\rho \in \mathcal{R}$. For a candidate value ρ , $\{\hat{y}_\tau(\rho)\}_{\tau=1}^N$ denotes a simulated realization of length N , i.e, a random draw from $p_{N-1}(y_1, \dots, y_N|\rho)$.

3.1 The Auxiliary Model

Suppose the conditional density $f(y_t|y_{t-L}, \dots, y_{t-1}, \theta)$ $\theta \in \Theta \subset R^{l_\theta}$ is found to provide a good statistical description of the data. This conditional density is called the f -model. The f -model providing a good fit means that it is selected on the basis of some model selection criterion such as the Schwarz criterion (BIC) and that it does well on a battery of specification tests. The score function of the f -model

$$s_f(y_{t-L}, \dots, y_t, \theta) = \frac{\partial}{\partial \theta} \log[f(y_t|y_{t-L}, \dots, y_{t-1}, \theta)] \quad (3.1)$$

is of central importance in what follows.

The quasi-maximum likelihood estimator obtained by fitting the f -model to the realization $\{\tilde{y}_t\}_{t=1}^n$ is

$$\tilde{\theta} = \arg \max_{\theta} \mathcal{L}_n(\theta, \{\tilde{y}_t\}_{t=1}^n) \quad (3.2)$$

where

$$\mathcal{L}_n(\theta, \{y_t\}_{t=1}^n) = \frac{1}{n} \sum_{t=L+1}^n \log[f(\tilde{y}_t|\tilde{y}_{t-L}, \dots, \tilde{y}_{t-1}, \theta)] \quad (3.3)$$

is the sample mean likelihood function. The first order condition is

$$\frac{\partial}{\partial \theta} \mathcal{L}_n(\tilde{\theta}, \{y_t\}_{t=1}^n) = 0, \quad (3.4)$$

or, equivalently,

$$\frac{1}{n} \sum_{t=L+1}^n s_f(\tilde{y}_{t-L}, \dots, \tilde{y}_t, \tilde{\theta}) = 0. \quad (3.5)$$

It is to be emphasized that, although the dimension of the f -model needs to be at least as large as the p -model, i.e., $l_\theta \geq l_\rho$, the f -model need not nest the p -model. Indeed, the internal structure of the f -model might have little to do directly with the internal structure of the p -model.

3.1.1 Theory of Misspecified Models

The f -model is possibly misspecified, in the sense that there is no value of $\theta \in \Theta$ such that $f(y_t|y_{t-L}, \dots, y_{t-1}, \theta) = p_L(y_t|y_{t-L}, \dots, y_{t-1}|\rho_0)$ for all y_{t-L}, \dots, y_{t-1} in the support of $p_L(y_{t-L}, \dots, y_t|\rho_0)$. It is well understood what happens if one fits an f -model when the truth

is the p -model. From White (1994) and the references therein, it is well known that under reasonable regularity conditions $\tilde{\theta} \xrightarrow{as} \bar{\theta}$ where $\bar{\theta}$ is the *pseudo-true* value given by

$$\bar{\theta} = \mathcal{T}(\rho_0),$$

and $\mathcal{T} : \mathcal{R} \rightarrow \Theta$ denotes the *binding function* (Gouriéroux, Monfort, and Renault, 1993) given by

$$\mathcal{T}(\rho) = \arg \max_{\theta} \int \log[f(y_t|y_{t-L}, \dots, y_{t-1}, \theta)] p_L(y_{t-L}, \dots, y_t|\rho) d(y_{t-L}, \dots, y_t). \quad (3.6)$$

The corresponding first order condition is

$$\int \frac{\partial}{\partial \theta} \log[f(y_t|y_{t-L}, \dots, y_{t-1}, \mathcal{T}(\rho))] p_L(y_{t-L}, \dots, y_t|\rho) d(y_{t-L}, \dots, y_t) = 0 \quad (3.7)$$

identically in $\rho \in \mathcal{R}$, and, in particular,

$$\int \frac{\partial}{\partial \theta} f(y_t|y_{t-L}, \dots, y_{t-1}, \bar{\theta}) p_L(y_{t-L}, \dots, y_t|\rho_0) d(y_{t-L}, \dots, y_t) = 0. \quad (3.8)$$

The asymptotic distribution of $\tilde{\theta}$ about $\bar{\theta}$ is given by the usual three-term expression

$$\sqrt{n}(\tilde{\theta} - \bar{\theta}) \xrightarrow{\mathcal{D}} N(0, \mathcal{H}^{-1} \mathcal{I} \mathcal{H}^{-1}),$$

where

$$\mathcal{I} = \lim_{n \rightarrow \infty} \text{Var} \left\{ \frac{1}{\sqrt{n}} \sum_{t=L+1}^n \frac{\partial}{\partial \theta} \log[f(y_t|y_{t-L}, \dots, y_{t-1}, \bar{\theta})] \right\} \quad (3.9)$$

is the *pseudo-information matrix*, and

$$\mathcal{H} = \int \frac{\partial^2}{\partial \theta \partial \theta'} \log[f(y_t|y_{t-L}, \dots, y_{t-1}, \bar{\theta})] p_L(y_{t-L}, \dots, y_t|\rho_0) d(y_{t-L}, \dots, y_t) \quad (3.10)$$

is the hessian matrix. We note the characterization

$$\sqrt{n}(\tilde{\theta} - \bar{\theta}) \stackrel{LD}{=} \mathcal{H}^{-1} v, \quad v \sim N(0, \mathcal{I}), \quad (3.11)$$

$$\stackrel{LD}{=} \mathcal{H}^{-1} \mathcal{I}^{1/2} z, \quad z \sim N(0, I), \quad \mathcal{I}^{1/2} \mathcal{I}^{1/2'} = \mathcal{I}, \quad (3.12)$$

where $\stackrel{LD}{=}$ means the left and right-hand sides have the same distribution asymptotically.

3.2 The Score-Based Criterion (EMM)

One strategy for minimum chi-square estimation is to mimic the first order condition (3.5) induced by QML estimation (3.2) of the f -model. This estimator is introduced in Bansal, Gallant, Hussey, and Tauchen (1993, 1995) and developed further in Gallant and Tauchen (1996).

The idea is this: Since

$$\frac{1}{n} \sum_{t=L+1}^n s_f(\tilde{y}_{t-L}, \dots, \tilde{y}_t, \tilde{\theta}) = 0,$$

where the left side is averaged over the observed realization $\{\tilde{y}_t\}_{t=1}^n$, then a good estimator for ρ would be one that makes

$$\frac{1}{N} \sum_{\tau=L+1}^N s_f(\hat{y}_{\tau-L}(\rho), \dots, \tilde{y}_\tau(\rho), \tilde{\theta}) \approx 0, \quad (3.13)$$

with the average taken over the simulated realization. If $\ell_\theta = \ell_\rho$, then (3.13) can hold with equality, but, if as is typical, $\ell_\theta > \ell_\rho$, then (3.13) cannot in general hold with equality, which motivates GMM of Hansen (1982). Put

$$\hat{m}_N(\rho, \theta) = \frac{1}{N} \sum_{\tau=L+1}^N s_f(\hat{y}_{\tau-L}(\rho), \dots, \hat{y}_\tau(\rho), \theta). \quad (3.14)$$

The estimator is

$$\hat{\rho} = \arg \min_{\rho} \left\{ \hat{m}_N(\rho, \tilde{\theta})' \tilde{\mathcal{I}}^{-1} \hat{m}_N(\rho, \tilde{\theta}) \right\}, \quad (3.15)$$

where $\tilde{\mathcal{I}}^{-1}$ is a weighting matrix and $\tilde{\mathcal{I}} \xrightarrow{a.s.} \mathcal{I}$ as $n \rightarrow \infty$.

In (3.14), the criterion $\hat{m}_N(\rho, \theta)$ is the mean score of the f -model averaged with respect to the simulation $\{\hat{y}_\tau(\rho)\}_{\tau=1}^N$. The estimator in (3.15) minimizes with respect to ρ the length of $\hat{m}_N(\rho, \tilde{\theta})$ relative to $\tilde{\mathcal{I}}^{-1}$, with θ left fixed at the QML estimator $\tilde{\theta}$ defined in (3.2) and (3.3) above.

With modern computing equipment one uses very long simulations, and so we consider letting simulated sample size N go to infinity. By assumption, the p -model defines a stationary data generator so, as discussed in Section 2, $\hat{m}_N(\rho, \theta) \xrightarrow{a.s.} m(\rho, \theta)$ as $N \rightarrow \infty$, where

$$m(\rho, \theta) = \int s_f(y_{t-L}, \dots, y_t, \theta) p_L(y_{t-L}, \dots, y_t | \rho) d(y_{t-L}, \dots, y_t). \quad (3.16)$$

The “almost surely” is with respect to the random number generator; the data set $\{\tilde{y}_t\}_{t=1}^n$ is fixed. Here $m(\rho, \theta)$ is the mean score of the f -model with respect to the p -model. Apart

from the small numerical error due to Monte Carlo integration, the estimator is the solution to the problem

$$\min_{\rho} m(\rho, \tilde{\theta})' \tilde{\mathcal{I}}^{-1} m(\rho, \tilde{\theta}). \quad (3.17)$$

A sketch of the asymptotics follows. Throughout, $\hat{\rho}$ denotes the solution to either (3.15) or (3.17), since in practice the Monte Carlo integration should be accurate enough to reduce the numerical error to the point where the two problems are essentially equivalent.

Consistency

A standard result of \sqrt{n} -asymptotics is that almost surely a solution to the problem (3.17) gets arbitrarily close to the set of solutions of the limiting problem

$$\min_{\rho} m(\rho, \bar{\theta})' \mathcal{I}^{-1} m(\rho, \bar{\theta}), \quad (3.18)$$

since $\tilde{\mathcal{I}} \xrightarrow{as} \mathcal{I}$ and $\tilde{\theta} \xrightarrow{as} \bar{\theta}$. Now, $m(\rho_0, \bar{\theta}) = 0$ by the first order condition (3.8) of the optimization problem defining the binding function. Hence, ρ_0 is a solution to the problem (3.18):

$$0 = m(\rho_0, \bar{\theta})' \mathcal{I}^{-1} m(\rho_0, \bar{\theta}) = \min_{\rho} m(\rho, \bar{\theta})' \mathcal{I}^{-1} m(\rho, \bar{\theta}).$$

The only issue is uniqueness, which pertains to identification. We assume ρ_0 is the only solution of

$$m(\rho, \bar{\theta}) = \int s_f(y_{t-L}, \dots, y_t, \bar{\theta}) p_L(y_{t-L}, \dots, y_t | \rho) d(y_{t-L}, \dots, y_t) = 0. \quad (3.19)$$

Hence $\hat{\rho} \xrightarrow{as} \rho_0$.

Asymptotic Normality

Asymptotic normality starts with the first order condition of (3.17):

$$\frac{\partial m(\hat{\rho}, \tilde{\theta})'}{\partial \rho} \tilde{\mathcal{I}}^{-1} m(\hat{\rho}, \tilde{\theta}) = 0.$$

Taking a mean value expansion of $m(\hat{\rho}, \tilde{\theta})$ about $m(\rho_0, \bar{\theta})$, and noting that $m(\rho_0, \bar{\theta}) = 0$ by (3.8), gives

$$\frac{\partial m(\hat{\rho}, \tilde{\theta})'}{\partial \rho} \tilde{\mathcal{I}}^{-1} \left[\frac{\partial m(\hat{\rho}, \tilde{\theta})}{\partial \rho'} (\hat{\rho} - \rho_0) + \frac{\partial m(\hat{\rho}, \tilde{\theta})}{\partial \theta'} (\tilde{\theta} - \bar{\theta}) \right] = 0,$$

where $\dot{\rho}$ and $\dot{\theta}$ mean that the rows of the derivative matrices are evaluated at intermediate points. This last expression gives

$$\sqrt{n}(\hat{\rho} - \rho_0) = -\left[\frac{\partial m(\hat{\rho}, \tilde{\theta})'}{\partial \rho} \tilde{\mathcal{I}}^{-1} \frac{\partial m(\hat{\rho}, \tilde{\theta})}{\partial \rho'}\right]^{-1} \frac{\partial m(\hat{\rho}, \tilde{\theta})'}{\partial \rho} \tilde{\mathcal{I}}^{-1} \frac{\partial m(\hat{\rho}, \tilde{\theta})}{\partial \theta'} \sqrt{n}(\tilde{\theta} - \bar{\theta}),$$

which highlights the fact that all of the randomness of the estimator is due solely to the fluctuations in $\tilde{\theta}$ about $\bar{\theta}$. Since $\tilde{\mathcal{I}} \xrightarrow{aS} \mathcal{I}$, $\tilde{\theta} \xrightarrow{aS} \bar{\theta}$, and $\hat{\rho} \xrightarrow{aS} \rho_0$, then

$$\sqrt{n}(\hat{\rho} - \rho_0) \stackrel{LD}{\equiv} -(D'_\rho \mathcal{I}^{-1} D_\rho)^{-1} D'_\rho \tilde{\mathcal{I}}^{-1} D_\theta \sqrt{n}(\tilde{\theta} - \bar{\theta}), \quad (3.20)$$

where $D_\theta = \partial m(\rho_0, \bar{\theta})/\partial \theta'$ and $D_\rho = \partial m(\rho_0, \bar{\theta})/\partial \rho'$. A key fact is that $D_\theta = \mathcal{H}$ where the hessian \mathcal{H} is given in (3.10); the verification is

$$\begin{aligned} D_\theta &= \frac{\partial m(\rho_0, \bar{\theta})}{\partial \theta'} \\ &= \int \frac{\partial}{\partial \theta'} s_f(y_{t-L}, \dots, y_t, \bar{\theta}) p_L(y_{t-L}, \dots, y_t | \rho_0) d(y_{t-L}, \dots, y_t) \\ &= \int \frac{\partial^2}{\partial \theta \partial \theta'} \log[f(y_t | y_{t-L}, \dots, y_{t-1}, \bar{\theta})] p_L(y_{t-L}, \dots, y_t | \rho_0) d(y_{t-L}, \dots, y_t) \\ &= \mathcal{H}. \end{aligned}$$

Using this fact together with (3.20) and the characterization (3.11) yields

$$\sqrt{n}(\hat{\rho} - \rho_0) \xrightarrow{\mathcal{D}} N[0, (D'_\rho \mathcal{I}^{-1} D_\rho)^{-1}]. \quad (3.21)$$

In addition, the limiting value of the normalized objective function is asymptotically chi-square:

$$nm(\hat{\rho}, \tilde{\theta})' \tilde{\mathcal{I}}^{-1} m(\hat{\rho}, \tilde{\theta}) \xrightarrow{\mathcal{D}} \mathcal{X}^2(\ell_\theta - \ell_\rho). \quad (3.22)$$

To justify (3.22), a mean-value approximation gives

$$m(\hat{\rho}, \tilde{\theta}) = \frac{\partial m(\hat{\rho}, \tilde{\theta})}{\partial \rho'} (\hat{\rho} - \rho_0) + \frac{\partial m(\hat{\rho}, \tilde{\theta})}{\partial \theta'} (\tilde{\theta} - \bar{\theta}),$$

and so

$$\sqrt{nm}(\hat{\rho}, \tilde{\theta}) \stackrel{LD}{\equiv} D_\rho \sqrt{n}(\hat{\rho} - \rho_0) + D_\theta \sqrt{n}(\tilde{\theta} - \bar{\theta}).$$

Using (3.20) gives

$$\sqrt{nm}(\hat{\rho}, \tilde{\theta}) \stackrel{LD}{\equiv} [I - D_\rho (D'_\rho \mathcal{I}^{-1} D_\rho)^{-1} D'_\rho \mathcal{I}^{-1}] D_\theta \sqrt{n}(\tilde{\theta} - \bar{\theta}).$$

From (3.11) $\sqrt{n}(\tilde{\theta} - \theta_0) \stackrel{LD}{=} \mathcal{H}^{-1}\mathcal{I}^{1/2}z$, where $z \sim N(0, I)$. Thus, remembering that $D_\theta = \mathcal{H}$,

$$\sqrt{nm}(\hat{\rho}, \tilde{\theta}) \stackrel{LD}{=} -[I - D_\rho(D'_\rho\mathcal{I}^{-1}D_\rho)^{-1}D'_\rho\mathcal{I}^{-1}]\mathcal{I}^{1/2}z. \quad (3.23)$$

Putting everything together gives

$$nm(\hat{\rho}, \tilde{\theta})'\tilde{\mathcal{I}}^{-1}m(\hat{\rho}, \tilde{\theta}) \stackrel{LD}{=} nm(\hat{\rho}, \tilde{\theta})'\mathcal{I}^{-1}m(\hat{\rho}, \tilde{\theta}) \quad (3.24)$$

$$\stackrel{LD}{=} \{[I - D_\rho(D'_\rho\mathcal{I}^{-1}D_\rho)^{-1}\mathcal{I}^{-1}]\mathcal{I}^{1/2}z\}'\mathcal{I}^{-1} \quad (3.25)$$

$$\{[I - D_\rho(D'_\rho\mathcal{I}^{-1}D_\rho)^{-1}\mathcal{I}^{-1}]\mathcal{I}^{1/2}z\} \quad (3.26)$$

$$\stackrel{LD}{=} z'[I - H(H'H)^{-1}H']z, \quad (3.27)$$

where $H = \mathcal{I}^{-1/2}D_\rho$. Note that H is $\ell_\theta \times \ell_\rho$ so $\text{rank}[I - H(H'H)^{-1}H'] = \ell_\theta - \ell_\rho$, and hence $nm(\hat{\rho}, \tilde{\theta})'\tilde{\mathcal{I}}^{-1}m(\hat{\rho}, \tilde{\theta}) \stackrel{LD}{=} \chi^2(\ell_\theta - \ell_\rho)$, as to be shown.

Gallant and Tauchen (1996) suggest using the estimate of \mathcal{I} based on the mean outer-product-of-the-gradient

$$\tilde{\mathcal{I}} = \frac{1}{n} \sum_{t=L+1}^n s_f(\tilde{y}_{t-L}, \dots, \tilde{y}_t, \tilde{\theta})s_f(\tilde{y}_{t-L}, \dots, \tilde{y}_t, \tilde{\theta})', \quad (3.28)$$

whenever the f -model is chosen so that $f(y_t|y_{t-L}, \dots, y_{t-1}, \bar{\theta})$ is a very good representation of the data. Then, the scores of the f -model are nearly serially uncorrelated as they are essentially martingale differences. If the scores of the f -model are not serially uncorrelated, the estimator is still consistent and asymptotically normal, though a weighted covariance matrix estimate would be needed:

$$\tilde{\mathcal{I}} = \sum_{t=L+1}^n \sum_{r=L+1}^n \omega_n(t-r)s_f(\tilde{y}_{t-L}, \dots, \tilde{y}_t, \tilde{\theta})s_f(\tilde{y}_{r-L}, \dots, \tilde{y}_r, \tilde{\theta})',$$

where $\{\omega_n(k)\}_{k=-(n-L)+1}^{n-L-1}$ are weights ensuring that $\tilde{\mathcal{I}}$ is positive definite and consistently estimates \mathcal{I} . There is an extensive literature on HAC estimation including Gallant (1987, p. 446), Newey and West (1987), and Andrews (1991).

3.3 The Indirect Inference Estimator

Smith (1990, 1993) develops a different strategy, which is extended to to very rich estimation environments by Gouriéroux, Monfort, and Renault (1993). This strategy, now termed

'Indirect Inference', mimics directly the optimization underlying the QMLE (3.3) instead of the first order conditions. Define

$$\mathcal{T}_N(\rho) = \arg \max_{\theta} \mathcal{L}_N(\theta, \{\hat{y}_{\tau}(\rho)\}_{\tau=1}^N), \quad (3.29)$$

where

$$\mathcal{L}_N(\theta, \{\hat{y}_{\tau}(\rho)\}_{\tau=1}^N) = \frac{1}{N} \sum_{\tau=L+1}^N \log[f(\hat{y}_{\tau}(\rho)|\hat{y}_{\tau-L}(\rho), \dots, \hat{y}_{\tau-1}(\rho), \theta)]. \quad (3.30)$$

The Indirect Inference estimator is

$$\hat{\rho}_{II} = \arg \min_{\rho} [\mathcal{T}_N(\rho) - \tilde{\theta}]' [\tilde{\mathcal{H}}^{-1} \tilde{\mathcal{I}} \tilde{\mathcal{H}}^{-1}]^{-1} [\mathcal{T}_N(\rho) - \tilde{\theta}], \quad (3.31)$$

where $\tilde{\mathcal{H}} \xrightarrow{a.s.} \mathcal{H}$ and $\tilde{\mathcal{I}} \xrightarrow{a.s.} \mathcal{I}$. The considerations raised at the end of Subsection 3.2 in regard to the estimator $\tilde{\mathcal{I}}$ apply to Indirect Inference as well. The natural estimator of \mathcal{H} is

$$\tilde{\mathcal{H}} = \frac{1}{n} \sum_{t=L+1}^n \frac{\partial^2}{\partial \theta \partial \theta'} \log[f(\hat{y}_t|\hat{y}_{t-L}, \dots, \hat{y}_{t-1}, \tilde{\theta})]. \quad (3.32)$$

As before, we let $N \rightarrow \infty$ (use sufficiently long realizations), and observe that the Indirect Inference estimator solves

$$\arg \min_{\rho} [\mathcal{T}(\rho) - \tilde{\theta}]' [\tilde{\mathcal{H}}^{-1} \tilde{\mathcal{I}} \tilde{\mathcal{H}}^{-1}]^{-1} [\mathcal{T}(\rho) - \tilde{\theta}], \quad (3.33)$$

where \mathcal{T} is the binding function given in (3.6).

To establish consistency, one observes that ρ_0 is a solution to the limiting problem

$$\arg \min_{\rho} [\mathcal{T}(\rho) - \bar{\theta}]' [\mathcal{H}^{-1} \mathcal{I} \mathcal{H}^{-1}]^{-1} [\mathcal{T}(\rho) - \bar{\theta}], \quad (3.34)$$

and the identification condition is that ρ_0 is the only solution to this optimization problem. The asymptotic distribution theory for this case, which is presented in Smith (1990,1993), proceeds much as in Subsection 3.2.

4 Selection of the Auxiliary Model

We now examine in some detail four major considerations for selection of the auxiliary model: efficiency, diagnostics, behavior under misspecification, and dynamic stability. The discussion follows the score-based approach described in Subsection 3.2, and so the auxiliary model is termed the score generator.

4.1 Efficiency

For any fixed score generator, the minimum chi-square estimator defined in (3.15) is not as efficient asymptotically as maximum likelihood. As will be seen, though, by following a suitable strategy for selecting the score generator, the efficiencies can be made arbitrarily close. This motivates the terminology Efficient Method of Moments, as method of moments becomes asymptotically as efficient as maximum likelihood.

In what follows, \mathcal{V}_f denotes the asymptotic variance from (3.21) for the estimator when the score generator is f , and \mathcal{V}_0 denotes the asymptotic variance of maximum likelihood. From standard efficiency theory, $\mathcal{V}_f \geq \mathcal{V}_0$, and we are interested in how close \mathcal{V}_f comes to the lower bound.

To see the main idea, we start with the *iid* case. Suppose for all J , $p_J(y_t|y_{t-J}, \dots, y_{t-1}, \rho_0) = p(y_t|\rho_0)$. Drop the time subscript and write the random variables $s_p = (\partial/\partial\rho) \log[f(Y|\rho_0)]$ and $s_f = (\partial/\partial\theta) \log[f(Y|\bar{\theta})]$, where $Y \sim p(y|\rho_0)$. As before, ρ_0 is the true value of ρ and $\bar{\theta}$ is the pseudo-true value of θ given by

$$\bar{\theta} = \arg \max_{\theta} \int \log[f(y|\theta)]p(y|\rho_0)dy.$$

From maximum likelihood theory, $\mathcal{V}_0 = [\mathcal{E}(s_p s_p')]^{-1}$, while, from (3.21), $\mathcal{V}_f = (D'_\rho \mathcal{I}^{-1} D_\rho)^{-1}$. In this case, $\mathcal{I} = \text{Var}(s_f) = \mathcal{E}(s_f s_f')$. Furthermore,

$$\begin{aligned} D_\rho &= \frac{\partial m(\rho_0, \bar{\theta})}{\partial \rho'} \\ &= \left[\frac{\partial}{\partial \rho'} \left(\int \frac{\partial}{\partial \theta} \log[f(y|\bar{\theta})]p(y|\rho)dy \right) \right]_{\rho=\rho_0} \\ &= \int \frac{\partial}{\partial \theta} \log[f(y|\bar{\theta})] \left\{ \frac{\partial}{\partial \rho'} \log[p(y|\rho_0)] \right\} p(y|\rho_0) dy \\ &= \mathcal{E}(s_f s_p'). \end{aligned}$$

Now consider the population regression of s_p onto s_f :

$$s_p = B s_f + u, \quad \text{Var}(u) = \Omega,$$

where $B = \mathcal{E}(s_p s_f') [\mathcal{E}(s_f s_f')]^{-1}$ is $\ell_\rho \times \ell_\theta$ and by construction $\text{Cov}(u, s_f) = 0$. Hence $\mathcal{E}(s_p s_p') = B \mathcal{E}(s_f s_f') B' + \Omega = D'_\rho \mathcal{I}^{-1} D_\rho + \Omega$, and so

$$\mathcal{V}_0 = (\mathcal{V}_f^{-1} + \Omega)^{-1}, \tag{4.1}$$

or, equivalently,

$$\mathcal{V}_f = (\mathcal{V}_0^{-1} - \Omega)^{-1}. \quad (4.2)$$

The asymptotic variance-covariance matrix of the EMM estimator exceeds that of maximum likelihood by an amount that depends on the variance-covariance matrix of the error from a linear projection of s_p onto s_f . The closer s_p comes to lying in the linear span of s_f , then the closer is the EMM estimator to achieving the efficiency of maximum likelihood.

Recent work by Gallant and Long (1995) indicates that full efficiency can be achieved and in the dependent case as well. We first illustrate the ideas when $\{y_t\}$ is Markovian. Suppose there is an L such that

$$p_J(y_t|y_{t-J}, \dots, y_{t-1}, \rho_0) = p_L(y_t|y_{t-L}, \dots, y_{t-1}, \rho_0)$$

for all $J \geq L$. The dynamic maximum likelihood estimator

$$\hat{\rho}_{MLE} = \arg \max_{\rho} \frac{1}{n} \sum_{t=L+1}^n \log[p_L(y_t|y_{t-L}, \dots, y_{t-1}, \rho)]$$

is asymptotically normal with variance-covariance matrix $\mathcal{V}_0 = [\mathcal{E}(s_{pt}s'_{pt})]^{-1}$, where

$$s_{pt} = \frac{\partial}{\partial \rho} \log[p_L(y_t|y_{t-L}, \dots, y_{t-1}, \rho)].$$

Now consider using an SNP model for a score generator. The SNP family of conditional densities is an increasing nested hierarchy of finite dimensional models,

$$\{f_K(y_t|y_{t-L}, \dots, y_{t-1}, \theta_K)\}_{K=0}^{\infty},$$

where $\theta_K \in \Theta_K$ and $\Theta_K \subset \Theta_{K+1}$ is a nested sequence of Euclidean spaces. The relevant feature is that the SNP family is dense under a Sobelov norm $\|\cdot\|$:

$$\lim_{K \rightarrow \infty} \inf_{\theta_K \in \Theta_K} \|f_K(\cdot|\cdot, \theta_K) - p_L(\cdot|\cdot, \rho_0)\| = 0.$$

The asymptotic variance of $\hat{\rho}$ with $f_K(y_t|y_{t-L}, \dots, y_{t-1}, \theta_K)$ used as the score generator is

$$\mathcal{V}_{f_K} = [\mathcal{E}(s_{pt}s'_{f_K,t})(\mathcal{I}_K)^{-1}\mathcal{E}(s_{f_K,t}s'_{pt})]^{-1},$$

where

$$s_{f_K,t} = \frac{\partial}{\partial \theta} \log[f_K(y_t|y_{t-L}, \dots, y_{t-1}, \bar{\theta}_K)],$$

with

$$\bar{\theta} = \arg \max_{\bar{\theta}_K} \mathcal{E} \left\{ \log [f_K(y_t | y_{t-L}, \dots, y_{t-1}, \bar{\theta}_K)] \right\},$$

and

$$\mathcal{I}_K = \sum_{i=-\infty}^{\infty} \mathcal{E}(s_{f_K,t} s'_{f_K,t+j}).$$

One can check that $D_\rho = \mathcal{E}(s_{f_K,t} s'_{\rho t})$ in the dynamic case. Here, $\mathcal{E}[\cdot]$ means an expectation computed with respect to the joint *pdf* $p_J(y_{t-J}, \dots, y_t | \rho_0)$.

Suppose that for some finite and possibly large K , the p -model lies in the SNP hierarchy, though possibly with a reparameterization. Then $s_{\rho t} = B s_{f_K,t}$ for some nonsingular matrix of constants B . Evidently, then $\mathcal{V}_{f_K} = \mathcal{V}_0$, and full efficiency obtains. Of course, in general, no such K exists, though Gallant and Long show that

$$\lim_{K \rightarrow \infty} \mathcal{V}_{f_K} = \mathcal{V}_0.$$

Interestingly, analogous results remain true even when $\{y_t\}$ is not Markovian, with another limit being taken as $L \rightarrow \infty$.

4.2 Diagnostics

The criterion

$$m(\rho, \theta) = \int s_f(y_{t-L}, \dots, y_t, \theta) p_L(y_{t-L}, \dots, y_t, \rho) d(y_{t-L}, \dots, y_t)$$

is the mean score of the f -model under the p -model. By the definition of $\bar{\theta}$, $m(\rho_0, \bar{\theta}) = 0$, and the EMM estimator $\hat{\rho}$ minimizes $m(\rho, \tilde{\theta})' \tilde{\mathcal{I}}^{-1} m(\rho, \tilde{\theta})$. The elements of $m(\hat{\rho}, \tilde{\theta})$ contain diagnostic information on how well the p -model accounts for the scores of the f -model. Smaller elements indicate those scores that the p -model explains relatively well, while large elements indicate those scores that it has trouble accounting for. Of course small and large need to be judged relative to the statistical variability of the scores. For this assessment, note that from (3.23)

$$\sqrt{nm}(\hat{\rho}, \tilde{\theta}) \xrightarrow{\mathcal{D}} N(0, [\mathcal{I} - D_\rho (D'_\rho \mathcal{I}^{-1} D_\rho)^{-1} D'_\rho]). \quad (4.3)$$

Thus, each element of the vector of t -statistics

$$T_n = \{\text{diag}[\tilde{\mathcal{I}} - \hat{D}_\rho (\hat{D}'_\rho \tilde{\mathcal{I}}^{-1} \hat{D}_\rho)^{-1} \hat{D}'_\rho]\}^{-1/2} \sqrt{nm}(\hat{\rho}, \tilde{\theta}) \quad (4.4)$$

is asymptotically $N(0, 1)$, where $\hat{D}_\rho = (\partial/\partial\rho')m(\hat{\rho}, \tilde{\theta})$ is a consistent estimate of D_ρ . The individual elements of T_n convey diagnostic information about how well the p -model fits the corresponding score of the f -model. If a closed form expression for $(\partial/\partial\rho)m(\hat{\rho}, \tilde{\theta})$ is not readily available, then a numerical estimate of D_ρ might be formed by taking difference quotients around the optimum.

For a quick and easy assessment not involving numerical derivatives, one can examine the quasi- t -ratios

$$\tilde{T}_n = \{\text{diag}[\tilde{\mathcal{I}}]\}^{-1/2} \sqrt{nm}(\hat{\rho}, \tilde{\theta}). \quad (4.5)$$

The quasi- t -ratios are readily computable and can provide useful diagnostic information. The quasi- t -ratios satisfy $\tilde{T}_n \xrightarrow{\mathcal{D}} N(0, \mathcal{V}^*)$, where

$$\begin{aligned} \mathcal{V}^* &= [\text{diag}(\mathcal{I})]^{-1/2} [\mathcal{I} - D_\rho (D_\rho \mathcal{I}^{-1} D_\rho)^{-1} D_\rho'] [\text{diag}(\mathcal{I})]^{-1/2} \\ \text{diag}(\mathcal{V}^*) &\leq I. \end{aligned}$$

For the same reasons as in Durbin's (1970) classic paper, the quasi- t -ratios are downward biased relative to 2.0. Large magnitudes of the quasi- t -ratios indicate how a p -model goes wrong, but small magnitudes are not necessarily confirmation of the validity of the model.

4.3 Misspecification

The treatment of misspecification of the p -model is along the lines of Newey (1985a, 1985b) and Tauchen (1985), who consider the effects on the asymptotic chi-square non-centrality parameter of small deviations in the maintained model from the true model.

Suppose the process $\{y_t\}$ is strictly stationary and ergodic, but the true joint density of the stretch (y_{t-J}, \dots, y_t) is $h_J(y_{t-L}, \dots, y_t)$, for $J = 0, 1, 2, \dots$. Suppose also that the maintained density under the model is Markovian of lag length L with conditional density $p_L(y_t|y_{t-L}, \dots, y_{t-1}, \rho)$ and joint density $p_L(y_{t-L}, \dots, y_t|\rho)$. For notational ease we put $\xi_t = (y'_{t-L}, \dots, y'_t)$ and drop the t and L subscripts where unneeded, so $h(\xi)$ represents the true density while $\{p(\xi, \rho)\}_{\rho \in \mathcal{R}}$, represents the maintained parametric family. The p -model is misspecified when there is no value of ρ in \mathcal{R} such that $p(\xi, \rho) = h(\xi)$ h -almost everywhere. $\mathcal{E}[g(y_{t-J}, \dots, y_t)]$ means the expectation computed under the true DGP,

$$\int g(y_{t-J}, \dots, y_t) h_J(y_{t-J}, \dots, y_t) d(y_{t-J}, \dots, y_t).$$

Suppose $f(y_t|y_{t-L}, \dots, y_{t-1}, \theta)$ is the score generator. Let

$$s_f(\xi_t, \theta) = \frac{\partial}{\partial \theta} \log[f(y_t|y_{t-L}, \dots, y_{t-1}, \theta)].$$

Then $\tilde{\theta} \xrightarrow{aS} \bar{\theta}$, where the pseudo-true value $\bar{\theta}$ satisfies

$$\int s_f(\xi, \bar{\theta})h(\xi)d\xi = 0.$$

Also, $\hat{\rho} \xrightarrow{aS} \bar{\rho}$ where the pseudo-true value $\bar{\rho}$ satisfies

$$\bar{\rho} = \arg \min_{\rho} m(\rho, \bar{\theta})' \bar{\mathcal{I}}^{-1} m(\rho, \bar{\theta}),$$

with

$$m(\rho, \bar{\theta}) = \int s_f(\xi, \bar{\theta})p(\xi, \rho)d\xi,$$

and

$$\bar{\mathcal{I}} = \sum_{i=-\infty}^{\infty} \mathcal{E}[s_f(\xi_t)s_f(\xi_{t-i})'].$$

By (3.22), under the maintained assumption of correct specification, the scaled value of the objective function, $c_n^2 = nm(\hat{\rho}, \tilde{\theta})'\tilde{\mathcal{I}}^{-1}m(\hat{\rho}, \tilde{\theta})$, satisfies $c_n^2 \xrightarrow{\mathcal{D}} \mathcal{X}^2(\ell_{\theta} - \ell_{\rho})$. Under misspecification, however,

$$\frac{c_n^2}{n} \xrightarrow{aS} \bar{c}^2 = m(\bar{\rho}, \bar{\theta})'\bar{\mathcal{I}}^{-1}m(\bar{\rho}, \bar{\theta}).$$

Put $\bar{m} = m(\bar{\rho}, \bar{\theta})$ so $\bar{c}^2 = \bar{m}'\bar{\mathcal{I}}^{-1}\bar{m}$. The value \bar{c}^2 is the approximate slope (Geweke, 1983) of the specification test. If, for a particular h , \bar{c}^2 is zero, then the test has no power against that alternative; if $\bar{c}^2 > 0$, then the test will ultimately reject the maintained model.

Write $p(\xi, \bar{\rho}) = [1 + \eta(\xi)]h(\xi)$, which defines $\eta(\xi)$, under the reasonable presumption that the supports of p and h are the same. The function η reflects the deviance of the maintained p -model from the true model. The value of \bar{c}^2 depends upon the maintained model only through \bar{m} . Observe that

$$\bar{m} = \int s_f(\xi, \bar{\theta})[1 + \eta(\xi)]h(\xi)d\xi = \int s_f(\xi, \bar{\theta})\eta(\xi)h(\xi)d\xi.$$

Consequently $\bar{m} = \text{Cov}(s_f, \eta)$ under $h(\xi)$. With $\bar{c}^2 = \text{Cov}(\eta, s_f)\bar{\mathcal{I}}^{-1}\text{Cov}(s_f, \eta)$, then the test of the overidentifying restrictions ultimately detects any misspecification such that $\text{Cov}(s_f, \eta) \neq 0$, and it will fail to detect misspecification such that $\text{Cov}(s_f, \eta) = 0$.

We can thus have a situation where, for a particular score generator, f , and true DGP h , $\text{Cov}(f, h) = 0$, so the test of the overidentifying restrictions is expected to be passed, but the model is misspecified. The test of the overidentifying restrictions will be fooled whenever the truth lies in a direction such that $\text{Cov}(s_f, \eta) = 0$. Without very strong *a priori* knowledge, the only way to avoid this situation is to take a flexible, more nonparametric approach to specification of the score generator.

4.4 Dynamic Stability

Tauchen (1995) examines the issue of dynamic stability of the p - and f -models. The upshot is that one really need not worry about imposing dynamic stability on the structural model itself. Dynamic stability is self-enforcing. If the optimizer wanders into the region of the parameter space where the underlying structural model is unstable, then the data simulator generates a wildly explosive simulated realization that induces a large value of the objective function. The time series properties of this explosive realization are very much unlike the time series properties of the observed data set to which the auxiliary model has been fitted, so the objective function attains an exceedingly high value. The situation is actually a bit more subtle, because automatic stability is ensured only if the auxiliary model itself is dynamically stable. The use of a dynamically unstable auxiliary model can be expected to define a GMM objective function with very poor numerical properties in both the stable and unstable regions of the parameter space.

Stability is of practical importance. It plays a role in the selection of the auxiliary model in the application below. Likewise, Andersen and Lund (1995) carefully examine a class of generalized GARCH and E-GARCH auxiliary models for the short-term interest rate. They find the former typically unstable, and therefore unusable as auxiliary models, while the latter are stable.

5 Application: An Interest Rate Diffusion

5.1 The SDE

We now consider estimation of the continuous time stochastic differential equation (SDE)

$$dV = \mu(V, \alpha)dt + \sigma(V, \beta)dW, \quad (5.1)$$

where V is a short-term interest rate, and $\mu(V, \alpha)$ and $\sigma(V, \beta)$ are the local drift and diffusion functions with parameters α and β , respectively. We let $\{V_t\}_{t \in [0, T]}$ denote a realization from the SDE and $\{v_t\}_{t=1}^n$ denote a discrete time equispaced selection from the continuous realization, $v_t = V_{th}$, $t = 1, 2, \dots, n$, $nh \leq T$, for some $h > 0$. For a single-factor model of the term structure, with the factor being a short-term interest rate, one would need to estimate α and β in order to compute the bond prices and interest rates at various horizons.

To implement the SDE as a stochastic data generator, we employ the Platen Strong Order 1 Scheme (Kloeden and Platen, 1992, pp. 374-75). This scheme modifies the elementary Euler method to include a second-order correction to take account of curvature in the diffusion function. The scheme generates discrete time data $\{v_t\}$ as

$$v_{t+1} = h(v_t, z_t^*, \alpha, \beta), \quad (5.2)$$

where $z^* = (z_{t1}, \dots, z_{tS})$ is a vector of S iid $N(0, 1)$ deviates and where S is the number of subintervals that $[t, t+h)$ is divided into. The scheme generates a strong solution, and hence a weak solution. As $S \rightarrow \infty$, the joint density of N iterations on (5.2) starting from v_0 converges to $p_{N-1}(v_1, \dots, v_N | v_0, \alpha, \beta)$ as implied by (5.1).

This scheme maps into the setup of Section 2 as

$$y_t = v_t, \quad \rho = (\alpha' \beta')',$$

and $w_t = z_t^*$ is the forcing process. By taking S very large and letting the iterations run for a long time to let transients die away, we can generate a simulated realization $\{\hat{y}_\tau\}_{\tau=1}^N$ from $p_{N-1}(y_1, \dots, y_N | \rho)$, the joint density of y_1, \dots, y_N implied by the SDE.

In special cases, the functional form of the transition probability density $p(y_{t+1} | y_t, \rho)$ is known. For instance, with a linear drift and constant diffusion function, this scheme defines

an AR(1) process for y_t . If the drift is linear and the diffusion function is proportional to $V^{1/2}$, i.e., the square root model of Cox, Ingersoll, and Ross (1985), $p(y_{t+1}|y_t, \rho)$ is a non-central chi-square. In this case, one can estimate ρ directly from an observed realization $\{\tilde{y}_t\}_{t=1}^n$ by dynamic maximum likelihood

$$\tilde{\rho}_{ML} = \arg \max_{\rho} \frac{1}{n} \sum_{t=2}^n \log[p(\tilde{y}_t|\tilde{y}_{t-1}, \rho)].$$

This approach is taken by Duffie and Singleton (1994) for the square-root model.

In general, however, the conditional density $p(y_{t+1}|y_t, \rho)$ does not have a tractable form, which motivates the use of alternative estimators. This point is well understood in the literature and highlighted in Melino's (1994) survey of continuous time methods in finance.

Ait-Sahalia (1995a) observes that the unconditional density $p(y|\rho)$ of the discrete time process $\{y_t\}$ implied by (5.1) has a closed form expression in terms of the drift function μ and diffusion function σ . He develops and implements the estimator

$$\tilde{\rho}_{NP} = \arg \min_{\rho} \sum_{t=1}^n [p(\tilde{y}_t|\rho) - \tilde{p}_{NP}(\tilde{y}_t)]^2$$

where $\tilde{p}_{NP}(\tilde{y}_t)$ is a kernel estimate of the unconditional density of y_t . This method thus uses the unconditional density of y_t as the standard of fit. Ait-Sahalia (1995a) develops a second method that uses a kernel estimate of the discrete-time transition density $p(y_{t+1}|y_t)$. The criterion of fit is developed from two equivalent expressions for the evolution of the continuous-time transition probability as implied by the forward and backward Kolmogorov equations under stationarity. Both of these two methods, as well as the others discussed further below, work directly from parametric specifications for μ and σ . In contrast, Ait-Sahalia (1995b) considers another approach that entails inverting the expression for unconditional density to generate a nonparametric estimate of σ from a nonparametric estimate of the unconditional density. When both μ and σ are parameterized this approach is unneeded, though a comparison of the parametric and nonparametric estimates of σ can be the basis of a specification test.

Conley *et al* (1995) adopt the methods of Hansen and Scheinkman (1995). They show how to compute tractable closed form expressions for

$$\mu_{\phi}(\rho) = \int \int \phi(y_{t-1}, y_t) p_1(y_{t-1}, y_t|\rho) dy_{t-1} dy_t$$

for test functions $\phi(y_{t-1}, y_t)$. Letting

$$m(y_{t-1}, y_t, \rho) = [\phi_1(y_{t-1}, y_t) - \mu_{\phi_1(\rho)}, \dots, \phi_J(y_{t-1}, y_t) - \mu_{\phi_J(\rho)}],$$

for a batch of J test functions, then a GMM estimator of ρ is given by

$$\hat{\rho}_{GMM} = \arg \min_{\rho} \left[\frac{1}{n} \sum_{t=2}^n m(y_{t-1}, y_t, \rho) \right]' \tilde{\mathcal{B}}^{-1} \left[\frac{1}{n} \sum_{t=2}^n m(y_{t-1}, y_t, \rho) \right],$$

where $\tilde{\mathcal{B}}$ is a weighted covariance matrix estimate of

$$\lim_{n \rightarrow \infty} \text{Var} \left[\frac{1}{\sqrt{n}} \sum_{t=2}^n m(y_{t-1}, y_t, \rho_0) \right].$$

Weighted covariance estimation is needed because $m(y_{t-1}, y_t, \rho_0)$ is not in general a serially uncorrelated vector process.

The EMM method provides an alternative to these procedures. A possible advantage of EMM in this context is that it provides a direct check on whether the time series properties of discrete time realizations $\{y_t\}$ generated via (5.1) are consistent with those implied by an independent estimate of the discrete time conditional density $f(y_t | y_{t-L}, \dots, y_{t-1})$. Another is that since the auxiliary model fits the model well, then the score should be nearly serially uncorrelated, thus eliminating the need for weighted covariance estimation. A disadvantage is that it entails the added complexity of a simulation-based method for a model with a single observed factor, whereas the other approaches do not need simulation. Plots of simulations are useful for assessing the predicted time series properties though, as in Figure 5.5 discussed below. All told, the procedures seem complementary. We now turn to implementation of EMM.

5.2 Data

The data are weekly (Friday) observations on the 30-day Eurodollar interest rate, January 1, 1975 – October 28, 1994, for 1035 observations. The top panel of Figure 5.1 is a plot of $\{\tilde{y}_t\}_{t=1}^{1035}$. Data on the seven-day Eurodollar rate were also collected but not used. As is typical of interest rates quoted for very short maturity intervals, the seven-day rate shows occasional extraordinarily large movements, presumably due to short-term liquidity or microstructure effects. My view is that such effects are not characteristic of the dynamics of the fundamental

factor for equilibrium asset pricing, though the EMM method could certainly be used to fit (5.1) to the seven-day rate.

5.3 Estimation of the Score Generator

The first step is to fit the auxiliary model for the conditional density of the interest rate process. Gallant and Tauchen (1989, 1992, 1995a) develop a very flexible approach, termed the SemiNonParametric (SNP) approach, to estimation of the conditional density of a discrete time stationary process. The SNP approach is based on a truncated Hermite series expansion. For the interest rate process y_t , the SNP specification is

$$f_K(y_t|y_{t-L}, \dots, y_{t-1}, \theta) = f_K(y_t|x_{t-1}, \theta) = \frac{[\mathcal{P}_t(z_t)]^2 n(z_t)}{\int [\mathcal{P}_t(u)]^2 n(u) du} \frac{1}{R_t}, \quad (5.3)$$

where $n(\cdot)$ is the standard normal density,

$$z_t = \frac{(y_t - \mu_t)}{R_t}, \quad (5.4)$$

$$x_{t-1} = (y_{t-L}, \dots, y_{t-1}), \quad (5.5)$$

$$\mu_t = \theta_{10} + \sum_{j=1}^{L_\mu} \theta_{1j} y_{t-j}, \quad (5.6)$$

$$R_t = \theta_{20} + \sum_{j=1}^{L_r} \theta_{2j} \text{abs}(z_{t-j}), \quad (5.7)$$

$$\mathcal{P}_t(z) = \sum_{\lambda \leq K_z} a_{\lambda,t} z^\lambda, \quad (5.8)$$

$$a_{\lambda,t} = \sum_{|\eta| \leq K_x} \theta_{3\lambda\eta} \prod_{j=1}^{L_p} y_{t-j}^{\eta_j}, \quad (5.9)$$

$\eta = \eta_1 + \dots + \eta_{L_p}$, $|\eta| = \sum_{j=1}^{L_p} \eta_j$, $K = (K_z \ K_x)$, $L = \max\{L_u, L_r, L_p\}$, $\theta_{300} = 1$ is a normalization to achieve identification, $\text{abs}(\cdot)$ is the absolute value function with the corner smoothed out, and by convention $L_p = 1$ when $K_x = 0$.

The SNP model takes the form of a truncated Hermite series expansion modified to make the polynomial nonnegative and to integrate to unity. The leading term of the expansion, obtained by setting $K_z = K_x = 0$ is an $\text{AR}(L_u)$ - $\text{ARCH}(L_r)$ model: a linear autoregression of lag length L_u with an ARCH error structure of lag length L_r . Higher order terms accommodate deviations from this model. Letting $K_z > 0$, while $K_x = 0$, induces a time homogeneous

non-Gaussian error structure, while $K_z > 0$ and $K_x > 0$ permits non-Gaussianity and conditional heterogeneity beyond ARCH. Because of the extreme persistence of the interest rate, we do not use the logistic transformation of x_{t-1} described in Gallant and Tauchen (1995a).

We need to determine the degrees K_z , K_x , of the polynomial defined in (5.8) and (5.9) and the lag lengths L_u , L_r , L_p in (5.6), (5.7), (5.8), (5.9). A selection strategy that has worked well across several financial market applications is to use the BIC, i.e., the Schwarz (1978) criterion, to guide first determination of the appropriate lag lengths, L_u and L_r , for the leading term and then go on to determine the appropriate degrees of the polynomials along with L_p .

Table 5.1 shows the objective function surface for the SNP estimation. Throughout, the first 15 observations are reserved for forming lags, leaving 1020 observations, net. From the table, it is seen that $L_u = 1$ and $L_r = 5$ are appropriate under BIC. Expanding from this point to $K_z = 4$ does substantially better under BIC, which reflects a non-Gaussian error density typical of financial data. Fits with $K_z = 6$ and $K_z = 8$ did slightly better on BIC, though there were some numerical problems with convergence and with the parameter estimates locating in the nonstationary region of parameter space. (Nonstationarity is checked by examining plots of very long simulated realizations for evidence of explosive behavior, which is always readily apparent.) It was decided to leave $K_z = 4$, since it is imperative for the score generator to remain in the stationary region, so EMM is not trying to fit to an unstable model. Inspection of plots of error densities revealed only minor differences across the fits with $K_z = 4, 6$, or 8 ; the densities all assumed the typical shape with high concentration of mass near the origin, thick tails, and thin shoulders relative to the Gaussian. In previous experience with financial applications, $K_z = 4$ has typically been found adequate.

The SNP specification with $L_u = 1$, $L_r = 5$, $K_z = 4$, and $K_x = 0$ is a linear autoregression with an ARCH error structure and a homogeneous nonparametric error density. It is analogous to the semiparametric GARCH model of Engle and Gonzales-Rivera (1991). The score from this specification is termed the “Semiparametric ARCH Score.” From Table 5.1, however, it is seen that the further expansion to $K_x = 1$ is warranted, which is evidence that the Semiparametric ARCH model does not capture all features of the data. The SNP model with $L_u = 1$, $L_r = 5$, $K_z = 4$, $L_p = 1$, and $K_x = 1$ is the best under BIC of all models con-

sidered. The score from this preferred specification is termed the “Nonlinear Nonparametric Score.”

Evaluation of graphic output can help assess the nature of the adequacy of the fit. The middle and bottom panels of Figure 5.1 are time series plots of the one-step conditional means and conditional standard deviations from the Nonlinear Nonparametric score generator evaluated at each data point. This model is seen to do a good job of tracking the mean and volatility over the sample period.

This cut at the data generates two score generators for the SDE to confront. The Semi-parametric ARCH Score contains 12 indicators that the SDE must match to zero as closely as it can, while the Nonlinear Nonparametric Score contains 17 indicators.

5.4 Estimation of the SDE

We examine three basic specifications of the SDE (5.1). The first is

$$\mu(V, \alpha) = \alpha_0 + \alpha_1 V, \quad (5.10)$$

$$\sigma(V, \beta) = \beta_0 V^{1/2}, \quad (5.11)$$

which is the well-known square-root model of Cox, Ingersoll, and Ross (1985). In what follows, the specification (5.10)–(5.11) is termed Linear-Square-Root. The second specification is the more general constant elasticity of variance (CEV) setup of Chan, Karolyi, Longstaff, and Sanders (1992):

$$\mu(V, \alpha) = \alpha_0 + \alpha_1 V, \quad (5.12)$$

$$\sigma(V, \beta) = \beta_0 V^{\beta_1}. \quad (5.13)$$

(5.12) – (5.13) with $\beta_1 \geq 1/2$ is termed Linear-CEV specification. Motivated by Conley *et al* (1995) and Ait-Sahalia (1995a), we also consider the CEV diffusion with a nonlinear specification for the drift

$$\mu(V, \alpha) = \alpha_{-1} V^{-1} + \alpha_0 + \alpha_1 V + \alpha_2 V^2, \quad (5.14)$$

$$\sigma(V, \beta) = \beta_0 V^{\beta_1}. \quad (5.15)$$

(5.14) – (5.15) is termed the Nonlinear-CEV specification. One reason for considering nonlinear specifications for the drift function is the evidence from Pfann, Schotman, and Tschernig

(1995) that nonlinearities in the conditional mean play a role in resolving some of the key empirical puzzles of the term structure.

Conley *et al* (1995) present sufficient conditions for the existence of a stationary solution on $(0, \infty)$ when the drift function is a truncated two-sided series expansion and the diffusion function is of the CEV form. Interestingly, they find that, under suitable restrictions on the parameters of the drift function, a stationary solution exists even in what might be termed the high volatility cases where $\beta_1 \geq 1$ in (5.13) and $\beta_1 \geq 3/2$ in (5.15). In estimation, none of these conditions is imposed. As discussed in Subsection 4.4, the EMM estimation method incorporates a stability penalty to force the fitted parameters to generate stationary simulated data so long as the auxiliary model itself is stationary, as was checked in Subsection 5.3 above.

To implement the Platen Strong Order 1 Scheme, we scaled time so that one week equals 0.10 units, i.e., $h = 0.10$, and divided the interval $[t, t + 0.10)$ into 100 equispaced intervals. We generated realizations of length 51,000 and discarded the first 1,000 to attenuate the effects of transients, leaving $N = 50,000$. Experimentation showed the results were quite insensitive to various other choices for S , e.g., $S = 50$ or $S = 75$, and also insensitive to the choice of N so long as $N \geq 30,000$. Other experiments showed that the results hardly change if a Weak Order 2 Scheme is used. Labeling one week as 0.10 units of time reflects a particular choice of scaling of the parameters that maintains reasonable numerical stability for the nonlinear optimization. With this scaling, then $e^{0.10\alpha_1}$ is the discrete-time first order autocorrelation coefficient implied by model $dV = (\alpha_0 + \alpha_1 V)dt + \beta_0 dW$. For more general specifications, though, this interpretation does not hold because interaction effects alter the interpretation of the individual parameters.

Table 5.2 shows the results of fitting these three specifications of the SDE using the two score generators (Semiparametric ARCH and Nonlinear Nonparametric) identified in Subsection 5.3. (The computations were done using the general purpose EMM package of Gallant and Tauchen (1995b) which includes code, data, and a worked example.) The table shows parameter values $\hat{\rho} = (\hat{\alpha}' \hat{\beta}')$ and the objective function defined in (3.15) rescaled as in (3.22) to follow a chi-square distribution. The weighting matrix $\tilde{\mathcal{I}}^{-1}$ was formed using the outer-product-of-the-gradient as given in (3.28). As seen from the table, the Linear-Square-

Root specification fails to account for the Semiparametric ARCH Score; it is overwhelmingly rejected. Interestingly, though, by increasing β_1 from 0.50 to within the range 0.90–0.95, the Linear-CEV specification essentially can fit the Semiparametric ARCH Score. For $\beta_1 = 0.95$, the p -value is just over 1 percent. The top panel of Figure 5.2 shows the quasi- t -ratios, defined in (4.5) above, for the Linear-Square-Root specification; the middle panel shows the quasi- t -ratios for the Linear-CEV specification with $\beta_1 = 0.90$. As evident from the top panel, the Linear-Square-Root specification has a great deal of trouble accounting for both the scores of the ARCH parameters and the scores of the Hermite parameters, which govern deviations in the error density from normality. On the other hand, the Linear-CEV specification essentially fits the Semiparametric ARCH Score, a finding consistent with that of Koedijk, Nissen, Schotman, and Wolff (1995) for monthly data.

Fitting the Semiparametric ARCH Score, though, is not the same as fitting the data. The Nonlinear Nonparametric Score fully accounts for the richness of the data, as it emerges from a complete specification search. As seen from Table 5.2, the Linear-CEV specification fails to fit this score and is overwhelmingly rejected. There is evidence in favor of the added complexity of nonlinear mean specification (5.14), which confirms Conley *et al* (1995) and Ait-Sahalia (1995a). For $\beta_1 = 0.75$, a criterion difference test of $H_0 : \alpha_{-1} = \alpha_2 = 0$ gives $\mathcal{X}^2(2) = 8.484$, p -value=0.0143, while for $\beta_1 = 0.90$ the same test gives $\mathcal{X}^2(2) = 12.953$, p -value=0.0015.

The parameter β_1 is difficult to estimate precisely. A test of $H_0 : \beta_1 = 0.75$ versus $H_1 : \beta_1 = 0.90$ gives $\mathcal{X}^2(1) = 7.67$, p -value=.0056, so there is considerable evidence for $\beta_1 > 0.75$. The concentrated objective function becomes quite flat for values of β_1 much higher than 0.90, though, making it difficult to pin down β_1 precisely. Values of β_1 for which the objective function differs by less than 3.84 from the value at 0.90 would be included in a 95% confidence interval about 0.90. Values like $\beta_1 = 1.25$ and $\beta_1 = 1.50$ would be included in such an interval.

From Table 5.2, the Nonlinear-CEV specification with β_1 pinned at 0.90 comes very close but does not quite fit the data. The bottom panel of Figure 5.2 shows the quasi- t -ratios for this specification. The only remaining difficulty is the quasi- t -ratios on the quadratic and quartic terms of the Hermite polynomial, suggesting the model might be generating data

that is conditionally too Gaussian. The quasi- t -ratios are similar for values of β_1 such as 1.25 and 1.50. For the Nonlinear-CEV specification, the value $\beta_1 = 1.50$ is in the upper range of values consistent with the data. As a check on robustness, we explore various aspects of this fit.

Figure 5.3 provides an assessment of the unconditional density of the data. The top panel shows a kernel estimate of the unconditional density of the observed data while the bottom shows the fit to a long realization. The bottom panel is the model's prediction at the fitted parameter values, and it is seen that the model fits this dimension of the data quite well. Figure 5.4 shows the drift and diffusion functions (5.14)–(5.15) of the Nonlinear-CEV specification at the fitted parameter values. The estimated drift function in the upper panel shows extreme mean reversion at low and high values of the interest rate and little mean reversion in the middle, which is confirmation of the plots in Conley *et al* (1995) and Ait-Sahalia (1995a).

Figure 5.5 shows volatility scatter plots, which are plots of the first difference, Δy_t , against the lagged level y_{t-1} of the interest rate. The top panel is the volatility scatter of actual data. The middle is the volatility scatter of a long simulation from the Linear-Square-Root specification estimated using the Nonlinear Nonparametric Score. The bottom is the volatility scatter of a long simulation from the Nonlinear-CEV specification. With enough patience one could simulate and plot an indefinitely long simulation, so only the outer envelopes of the clouds of simulated points matter, not their densities within the ergodic sets. The middle panel shows that the ergodic set of the Linear-Square-Root specification is too small relative to the data, which explains why it performs so poorly. The Nonlinear-CEV scatter looks more like that of the data.

The top panel of Figure 5.6 shows an extremely long simulated realization (100 years worth) from the Nonlinear-Nonparametric Score generator while the bottom panel shows a simulation from the Nonlinear-CEV specification. The simulation from the Nonlinear-CEV specification is not unreasonable in appearance and, as to be expected, it is somewhat smoother than the simulation from the score generator to which it is fit.

In summary, there is considerable evidence for values of β_1 in the range 0.90–0.95 and possibly as high as 1.50, and for the nonlinear mean specification as well. The most general

Nonlinear-CEV specification comes close, but does not quite fit the data. The shape of the nonlinear drift function shown in Figure 5.4 confirms the shape reported in Conley *et al* (1995) and Ait-Sahalia (1995a), which was detected using substantially different (non-simulation) methods on different data sets.

6 Conclusion

One lesson from the application is the importance of using a flexible, fully nonparametric score generator. The Semiparametric ARCH Score reveals the inadequacy in the Linear-Square-Root specification and indicates values of β_1 in the range 0.90–0.95. However, using this score alone gives misleading findings, as it suggests the Linear-CEV model is adequate when in fact there is more structure to the data. The Nonlinear Nonparametric Score reveals the nonlinear structure of the data. From Table 5.1, this score generator is appropriate, as it is preferred under BIC to the Semiparametric ARCH Score.

The remaining misspecification of the Nonlinear-CEV specification apparent from the bottom panel of Figure 5.2 can probably be handled by introduction of a second stochastic volatility factor where (5.1) is extended to a two-factor model:

$$dV = \mu_1(V, \alpha)dt + \sigma_1(V, U, \beta)dW_1 \quad (6.1)$$

$$dU = \mu_2(U, \gamma)dt + \sigma_2(U, \gamma)dW_2. \quad (6.2)$$

A version of this model is discussed in Koedijk, Nissen, Schotman, and Wolff (1995) and favorable empirical evidence based on a discrete time approximation is available in Brenner, Harjes, and Kroner (1994). The latent volatility factor U can generate more exaggerated deviations from Gaussianity in the marginal distribution of V than can (5.1), and thus presumably should fit the data better. This specification can be estimated via EMM. Andersen and Lund (1995) undertake such estimation using GARCH and generalized GARCH models as score generators. Their findings are that the volatility factor U is indeed needed to account adequately for all of the dynamics of the short interest rate.

References

- Ait-Sahalia, Yacine 1996a. 'Testing continuous-time models of the spot interest rate', Review of Financial Studies, 9:forthcoming.
- Ait-Sahalia, Yacine 1996b. 'Nonparametric pricing of interest rate derivatives', Econometrica, 64:forthcoming.
- Andersen, Torben G. and Lund, Jesper 1995. 'Estimating continuous-time stochastic volatility models of the short term interest rate', Manuscript, Kellogg Graduate School of Management, Northwestern University.
- Andrews, Donald W. K. 1991. Heteroskedasticity and autocorrelation consistent covariance matrix estimation', Econometrica 59: 307–346.
- Bansal, Ravi, Gallant, A. Ronald, Hussey, Robert and Tauchen, George 1993. 'Computational aspects of nonparametric simulation estimation', in: David A. Belsley, (ed.), Computational techniques for econometrics and economic analysis. Boston: Kluwer Academic Publishers, pp. 3–22.
- Bansal, Ravi, Gallant, A. Ronald, Hussey, Robert and Tauchen, George 1995. 'Nonparametric estimation of structural models for high-frequency currency market data', Journal of Econometrics 66:251–287.
- Bollerslev, Tim, Chou, Ray and Kroner, Kenneth 1992. 'ARCH modeling in finance: a selective review of the theory and empirical evidence, with suggestions for future research', Journal of Econometrics 52: 5–61.
- Brenner, Robin J., Harjes, Richard H. and Kroner, Kenneth F. 1994. 'Another look at alternative models of the short-term interest rate', Manuscript, University of Arizona.
- Buraschi, Andrea 1994. 'The nominal term structure, stock prices and consumption dynamics with nonneutral inflation', Manuscript, University of Chicago.

- 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.
- Conley, Tim, Hansen, Lars Peter, Luttmer, Erzo and Scheinkman, José 1995. 'Estimating subordinated diffusions from discrete time data', Manuscript, University of Chicago.
- Cox, John C., Ingersoll, Jonathan E. Jr. and Ross, Stephen A. 1985. 'A theory of the term structure of interest rates', Econometrica 53: 385–407.
- Duffie, Darrell and Kan, Rui 1993. 'A yield-factor model of interest rates', Manuscript, Stanford University.
- Duffie, Darrell and Singleton, Kenneth J. 1993. 'Simulated moments estimation of markov models of asset prices', Econometrica 61: 929–952.
- Duffie, Darrell and Singleton, Kenneth J. 1994. 'Econometric modeling of term structures of defaultable bonds', Manuscript, Stanford University.
- Durbin, J. 1970. 'Testing for serial correlation in least squares regression when some of the regressors are lagged dependent variables', Econometrica 38: 410–421.
- Engle, Robert F. 1982. 'Autoregressive conditional heteroskedasticity with estimates of the variance of united kingdom inflation', Econometrica 50: 987–1008.
- Engle, Robert F. and Gonzales-Rivera, Gloria 1991. 'Semiparametric ARCH models', Journal of Business and Economic Statistics 9: 345–360.
- Engle, Robert F. and Gary J. Lee, 1994. 'Estimating diffusion models of stochastic volatility', Manuscript, University of California at San Diego.
- Foster, Douglas and Viswanathan, S. 1995. 'Can speculative trading explain the volume-volatility relation?' Journal of Business and Economic Statistics 13: 379–398.
- Gallant, A. Ronald 1987. Nonlinear statistical models. New York, NY: John Wiley and Sons.

- Gallant, A. Ronald, Hsieh, David A. and Tauchen, George 1995. 'Estimation of stochastic volatility models with diagnostics', Manuscript, Duke University. (Available via anonymous ftp to ftp.econ.duke.edu, subdirectory home/get/papers, as PostScript file msv.ps.)
- Gallant, A. Ronald and Long, Jonathan R. 1995. 'Estimating stochastic differential equations efficiently by minimum chi-square', Manuscript, University of North Carolina at Chapel Hill.
- Gallant, A. Ronald and Nychka, D. W. 1987. 'Semi-nonparametric maximum likelihood estimation', Econometrica 55, 363–390.
- Gallant, A. Ronald and Tauchen, George 1989. 'Semi-nonparametric estimation of conditionally constrained heterogeneous processes: asset pricing applications', Econometrica 57: 1091–1120.
- Gallant, A. Ronald and Tauchen, George 1992. 'A nonparametric approach to nonlinear time series analysis: estimation and simulation', in E. Parzen, D. Brillinger, M. Rosenblatt, M. Taqqu, J. Geweke and P. Caines (eds.), New Dimensions in Time Series Analysis. New York: Springer-Verlag.
- Gallant, A. Ronald and Tauchen, George 1995a. 'SNP: a program for nonparametric time series analysis, a user's guide', Manuscript, University of North Carolina at Chapel Hill. (Available along with code via anonymous ftp to ftp.econ.duke.edu, subdirectory home/arg/snp.)
- Gallant, A. Ronald and Tauchen, George 1995b. 'EMM: a program for efficient method of moments estimation, A user's guide', Manuscript, Duke University. (Available along with code via anonymous ftp to ftp.econ.duke.edu, subdirectory home/get/emm.)
- Gallant, A. Ronald and Tauchen, George 1996. 'Which moments to match?', Econometric Theory, forthcoming. (Available via anonymous ftp to ftp.econ.duke.edu, subdirectory home/get/papers as PostScript file effgmm.ps.)

- Gennotte, G. and Marsh, T. A. 1993. 'Variations in economic uncertainty and risk premiums on capital assets'. European Economic Review 37: 1021–41.
- Geweke, John 1983. 'The approximate slope of econometric tests', Econometrica 49: 1427–1442.
- Ghysels, Eric and Jasiak, Joanna 1994. 'Stochastic volatility and time deformation: An application to trading volume and leverage effects', Manuscript, University of Montreal.
- Gouriéroux, C., Monfort, A. and Renault, E. 1993. 'Indirect inference', Journal of Applied Econometrics 8: S85–S118.
- Gouriéroux, C. and Monfort, A. 1994. 'Simulation based econometric methods', Manuscript, Center of Operations Research & Econometrics, University of Louvain.
- Hamilton, James D. 1989. 'A new approach to the economic analysis of nonstationary times series and the business cycle', Econometrica 57: 357–84.
- Hansen, Lars Peter 1982. 'Large sample properties of generalized method of moments estimators', Econometrica 50: 1029–1054.
- Hansen, Lars Peter and Scheinkman, José 1995. 'Back to the future: generating moment implications for continuous-time markov processes', Econometrica 63: 767–804.
- Hsu, Chien-Te and Kugler, Peter 1995. 'The term structure of interest rates: systematic monetary policy and nonlinear dynamics', Manuscript, University of Vienna.
- Ingram, Beth F. and Lee, B. S. 1991. 'Simulation estimation of time series models', Journal of Econometrics 47: 197–205.
- Judd, Kenneth L. 1994. 'Numerical methods in economics', Manuscript, Stanford University.
- Kloeden, Peter E. and Platen, Eckhard 1992. Numerical solution of stochastic differential equations. New York: Springer-Verlag.

- Koedijk, Kees G., Nissen, Francois G. J. A., Schotman, Peter C. and Wolff, Christian C. P. 1995. 'The dynamics of short-term interest rate volatility reconsidered', Manuscript, Limburg Institute of Financial Economics.
- McFadden, Daniel 1989. 'A method of simulated moments for estimation of discrete response models without numerical integration', Econometrica 57: 995–1026.
- Melino, Angelo 1994. 'Estimation of continuous time models in finance', in C. Sims (ed.), Advances in econometrics, sixth world congress, volume II. Cambridge University Press.
- Newey, Whitney K. 1985a. 'Generalized method of moments specification testing', Journal of Econometrics 29: 229–256.
- Newey, Whitney K. 1985b. 'Conditional moment specification testing', Econometrica 53: 1047–1071.
- Newey, Whitney K. and West, Kenneth D. 1987. 'A simple positive semi-definite heteroskedasticity and autocorrelation consistent covariance matrix estimator', Econometrica 55: 703–708.
- Pagan, A. R., Hall, A. D. and Martin, V. 1995. 'Modeling the term structure', Manuscript, The Australian National University.
- Pakes, Ariel and Pollard, David 1989. 'Simulation and the asymptotics of optimization estimators', Econometrica 57: 1027–1058.
- Pfann, Gerard A., Schotman, Peter C. and Tschernig, Rolf 1995. 'Nonlinear interest rate dynamics and implications for the term structure', Journal of Econometrics, forthcoming.
- Schwarz, Gideon 1978. 'Estimating the dimension of a model', The Annals of Statistics 6: 461–464.
- Smith, Anthony A. 1990. 'Three essays on the solution and estimation of dynamic macroeconomic models', Ph.D. Dissertation, Duke University, Durham, NC.

- Smith, Anthony A. 1993. 'Estimating nonlinear time series models using simulated vector autoregressions,' The Journal of Applied Econometrics 8: S63–S84.
- Tauchen, George 1985. 'Diagnostic testing and evaluation of maximum likelihood models', Journal of Econometrics 30: 415–443.
- Tauchen, George 1995. 'The objective function of simulation estimators near the unstable region of the parameter space', Manuscript, Duke University. (Available via anonymous ftp to ftp.econ.duke.edu, subdirectory home/get/papers as PostScript file so.ps.)
- Tauchen, George and Hussey, Robert 1991. 'Quadrature-based methods for obtaining approximate solutions to nonlinear asset pricing models', Econometrica 59: 371–396.
- Taylor, John B. and Uhlig, Harald 1990. 'Solving nonlinear stochastic growth models: a comparison of alternative solution methods', Journal of Business and Economic Statistics 8: 1–17.
- White, Halbert 1994. Estimation, inference, and specification analysis. Cambridge University Press.

List of Tables

Table 5.1. SNP estimation of the conditional density of the 30-Day Eurodollar interest rate, weekly, January 3, 1975 – October 28, 1994.

Table 5.2. Estimation of the stochastic differential equation by EMM to the 30-Day Eurodollar interest rate, weekly, January 3, 1975 – October 28, 1994.

List of Figures

Figure 5.1 The top panel is a time series plot of weekly observations $\{\tilde{y}_t\}_{t=1}^{1035}$ on the 30-day Eurodollar interest rate, January 3, 1975 – October 28, 1994. The middle and bottom panels are plots of the SNP one-step conditional mean, $\{\mathcal{E}(\tilde{y}_t|\tilde{y}_{t-5}, \dots, \tilde{y}_{t-1})\}_{t=6}^{1035}$, and conditional standard deviation $\{\sqrt{\text{Var}(\tilde{y}_t|\tilde{y}_{t-5}, \dots, \tilde{y}_{t-1})}\}_{t=6}^{1035}$ from the fitted SNP model termed the “Nonlinear Nonparametric,” where $L_m = 1$, $L_r = 5$, $K_z = 4$, $L_p = 1$, $K_x = 1$.

Figure 5.2 Barplots of quasi- t -ratios defined in (4.5).

Figure 5.3 The top panel is a kernel estimate of the unconditional density of weekly observations $\{\tilde{y}_t\}_{t=1}^{1035}$ on the 30-day Eurodollar interest rate, January 3, 1975 – October 28, 1994. The bottom panel is a kernel estimate from a long simulation of the unconditional density of the fitted stochastic differential equation (5.1), where the drift and diffusion functions are the Nonlinear-CEV specification given in (5.14) and (5.15) with $\beta_1 = 1.5$ and the other parameters fitted via EMM using the Nonlinear Nonparametric Score.

Figure 5.4 The two panels show the local mean and variance functions of the fitted stochastic differential equation (5.1), where the drift and diffusion functions are the Nonlinear-CEV specification given in (5.14) and (5.15) with $\beta_1 = 1.5$ and the other parameters fitted via EMM using the Nonlinear Nonparametric Score.

Figure 5.5 The top panel is a plot of $\Delta\tilde{y}_t = y_t - y_{t-1}$ versus \tilde{y}_{t-1} based on weekly observations $\{\tilde{y}_t\}_{t=1}^{1035}$ on the 30-day Eurodollar interest rate, January 3, 1975 —October 28, 1994. The middle panel shows $\Delta\hat{y}_\tau$ against $\hat{y}_{\tau-1}$ for a long simulation $\{\hat{y}_\tau\}$ from the stochastic differential equation (5.1), where the drift and diffusion functions are given in the fitted Linear-Square-Root specification (5.10) and (5.11). The bottom panel shows $\Delta\hat{y}_\tau$ against $\hat{y}_{\tau-1}$ for a long simulation $\hat{y}_\tau = \hat{v}_\tau = V_\tau$ from the fitted stochastic differential equation (5.1) where the drift and diffusion functions are the Nonlinear-CEV specification given in (5.14) and (5.15) with $\beta_1 = 1.5$. Both the Linear-Square-Root and the Nonlinear-CEV specifications were estimated via EMM using the Nonlinear Nonparametric Score.

Figure 5.6 The top panel shows a long simulation $\{\hat{y}_\tau\}$ from the fitted SNP model labeled the “Nonlinear Nonparametric,” where $L_m = 1$, $L_r = 5$, $K_z = 4$, $L_p = 1$, $K_x = 1$. The bottom panel shows a long simulation $\{\hat{y}_\tau\}$ from the fitted stochastic differential equation (5.1), where the drift and diffusion functions are the Nonlinear-CEV specification given in (5.14) and (5.15) with $\beta_1 = 1.5$ and the other parameters fitted via EMM using the Nonlinear Nonparametric Score. Each simulation runs for 5,200 periods, i.e., 5,200 weeks, or 100 years.

Table 5.1.

| SNP Tuning Parameters | | | | | | Objective Function | |
|-----------------------|-------|-------|-------|-------|---------------|---------------------|-------------|
| L_u | L_r | L_p | K_z | K_x | ℓ_θ | $s_n(\hat{\theta})$ | Schwarz |
| 1 | 0 | 1 | 0 | 0 | 3 | -0.68717439 | -0.67711329 |
| 2 | 0 | 1 | 0 | 0 | 4 | -0.68910651 | -0.67569172 |
| 1 | 1 | 1 | 0 | 0 | 4 | -0.97324818 | -0.95983338 |
| 1 | 2 | 1 | 0 | 0 | 5 | -1.05383519 | -1.03706670 |
| 1 | 3 | 1 | 0 | 0 | 6 | -1.10456091 | -1.08443872 |
| 1 | 4 | 1 | 0 | 0 | 7 | -1.12721105 | -1.10373515 |
| 1 | 5 | 1 | 0 | 0 | 8 | -1.16777523 | -1.14094564 |
| 1 | 6 | 1 | 0 | 0 | 9 | -1.17026140 | -1.14007811 |
| 1 | 7 | 1 | 0 | 0 | 10 | -1.17372354 | -1.14018655 |
| 1 | 5 | 1 | 2 | 0 | 10 | -1.20305457 | -1.16951784 |
| → 1 | 5 | 1 | 4 | 0 | 12 | -1.25419099 | -1.21394660 |
| ⇒ 1 | 5 | 1 | 4 | 1 | 17 | -1.27719264 | -1.22017976 |
| 1 | 5 | 1 | 4 | 2 | 22 | -1.28353119 | -1.20974981 |
| 1 | 5 | 2 | 4 | 1 | 22 | -1.28060865 | -1.20682728 |

Table 5.2.

| Score Generator | α_{-1} | α_0 | α_1 | α_2 | β_0 | β_1 | $\chi^2(\hat{\rho})$ | df | p-value |
|-----------------|---------------|------------|------------|------------|-----------|-----------|----------------------|----|---------|
| SP-ARCH | | 0.085212 | -0.009535 | | 0.431449 | 0.500 | 42.082 | 9 | < .0001 |
| SP-ARCH | | 0.060660 | -0.005856 | | -0.198505 | 0.750 | 29.199 | 9 | 0.0006 |
| SP-ARCH | | 0.043520 | -0.003212 | | -0.124528 | 0.900 | 21.924 | 9 | 0.0091 |
| SP-ARCH | | 0.039678 | -0.002599 | | -0.107789 | 0.950 | 19.902 | 9 | 0.0185 |
| Nonlinear-NP | | 0.844394 | -0.110643 | | 0.217226 | 0.500 | 52.480 | 14 | < .0001 |
| Nonlinear-NP | | 0.252286 | -0.032167 | | 0.127801 | 0.750 | 44.886 | 14 | < .0001 |
| Nonlinear-NP | | 0.180280 | -0.017737 | | 0.105673 | 0.900 | 41.684 | 14 | 0.0001 |
| Nonlinear-NP | | 0.138474 | -0.009072 | | 0.106227 | 0.950 | 38.893 | 14 | 0.0004 |
| Nonlinear-NP | 83.891166 | -29.604147 | 3.353859 | -0.122553 | 0.205567 | 0.500 | 51.679 | 12 | < .0001 |
| Nonlinear-NP | 0.301075 | 0.104330 | -0.020774 | -0.000304 | 0.122584 | 0.750 | 36.402 | 12 | 0.0002 |
| Nonlinear-NP | 1.488935 | -0.395452 | 0.027440 | -0.000386 | 0.095634 | 0.900 | 28.732 | 12 | 0.0043 |
| Nonlinear-NP | 0.513064 | -0.086384 | 0.002571 | 0.000090 | 0.077414 | 1.000 | 32.590 | 12 | 0.0011 |
| Nonlinear-NP | 0.889781 | -0.300572 | 0.035812 | -0.001098 | 0.048673 | 1.250 | 33.041 | 12 | 0.0010 |
| Nonlinear-NP | 4.972764 | -2.083842 | 0.275028 | -0.010533 | 0.029594 | 1.500 | 32.589 | 12 | 0.0011 |
| Nonlinear-NP | 13.253822 | -5.662490 | 0.790175 | -0.034622 | 0.011334 | 2.000 | 32.594 | 12 | 0.0011 |
| Nonlinear-NP | 4.744214 | -3.192985 | 0.669735 | -0.041405 | 0.004157 | 2.500 | 31.771 | 12 | 0.0015 |

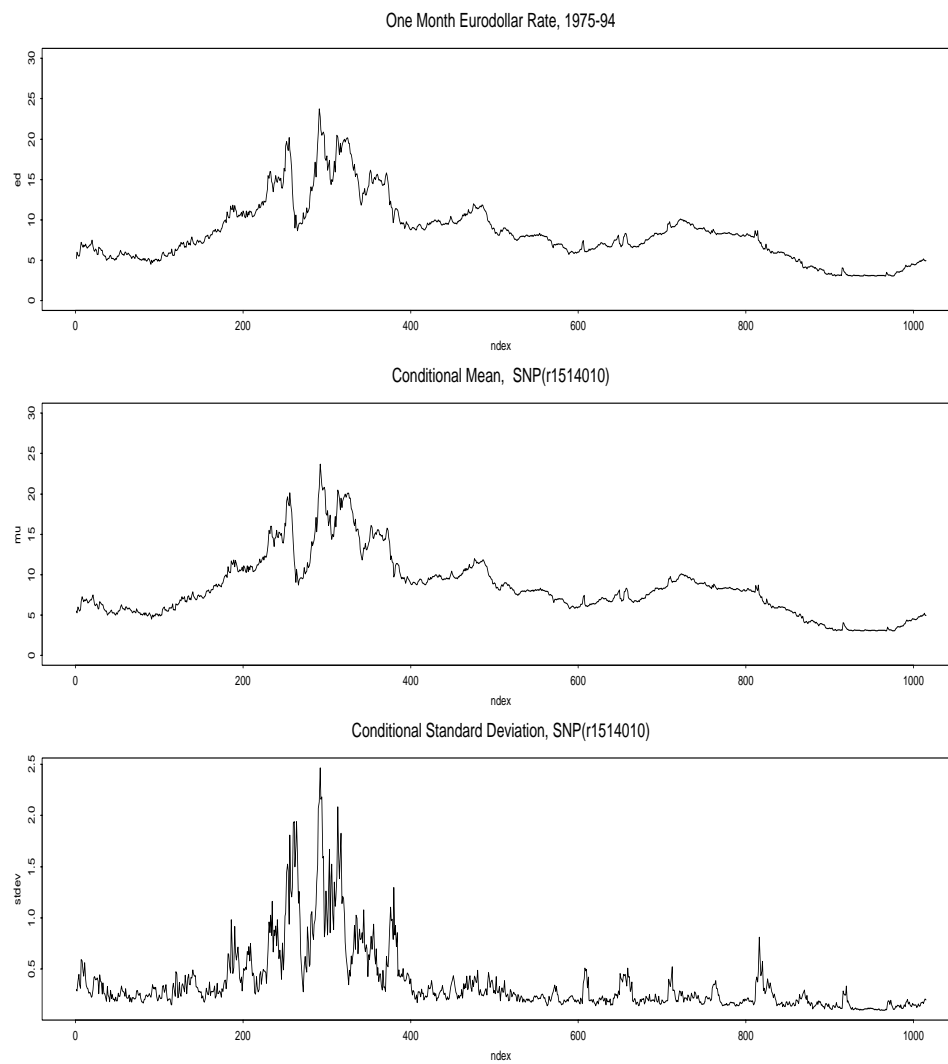
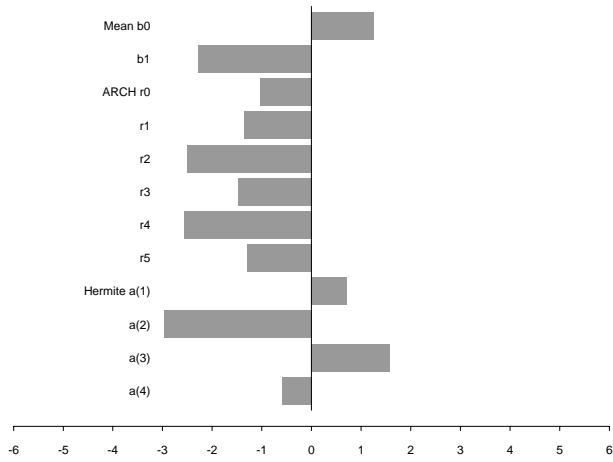
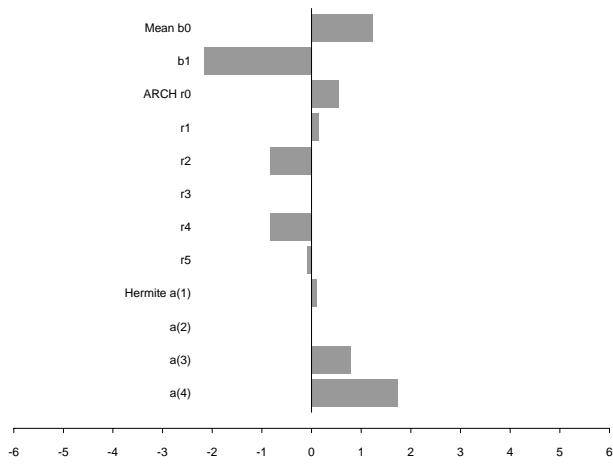


Figure 5.1

T-Ratios of Mean Score, Sqrt Specification
Score Generator: r1514000



T-Ratios of Mean Score, CEV beta1=0.90 Specification
Score Generator: r1514000



T-Ratios of Mean Score, Nonlinear SDE beta1=0.90
Score Generator: r1514010

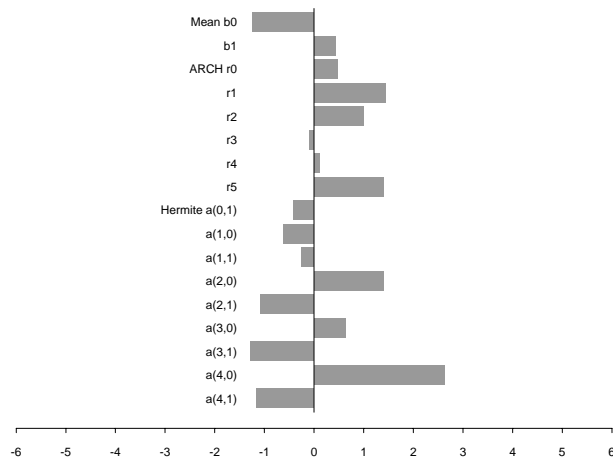
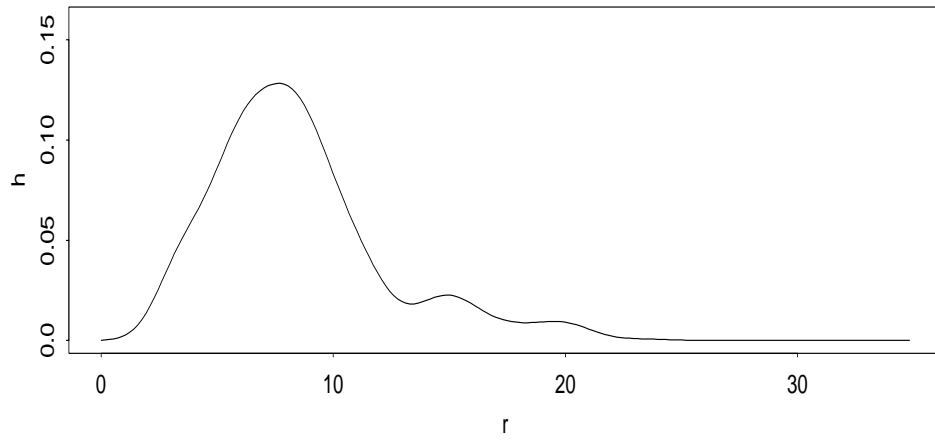


Figure 5.2

Unconditional Density of One Month Eurodollar Rate, 1975-94



Unconditional Density of Simulation From Nonlinear SDE, beta1=1.5

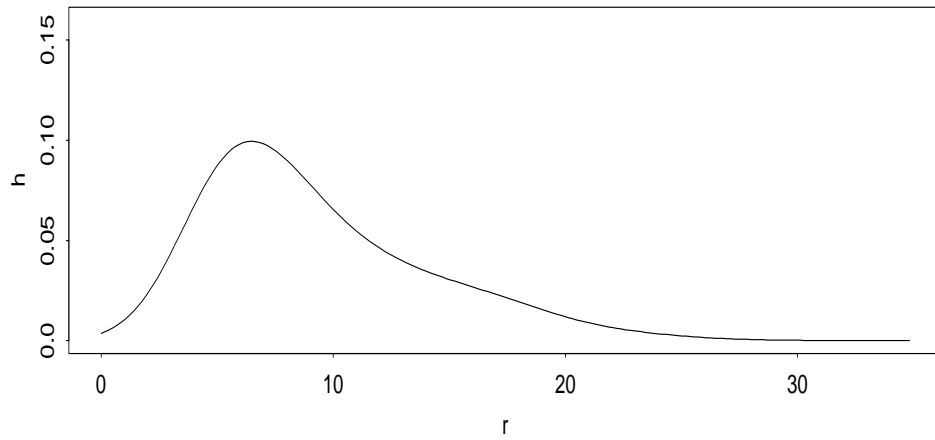
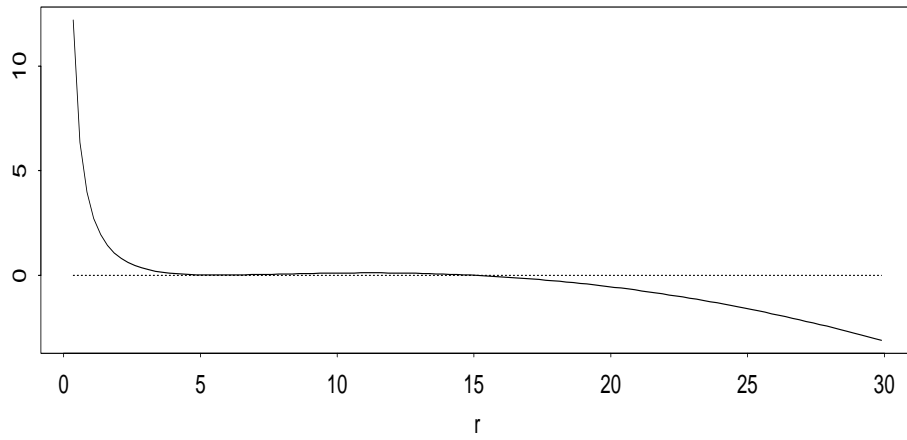


Figure 5.3

Local Mean, Nonlinear SDE with beta1=1.50 fit to (r1514010)



Local Variance, Nonlinear SDE with beta1=1.50 fit to (r1514010)

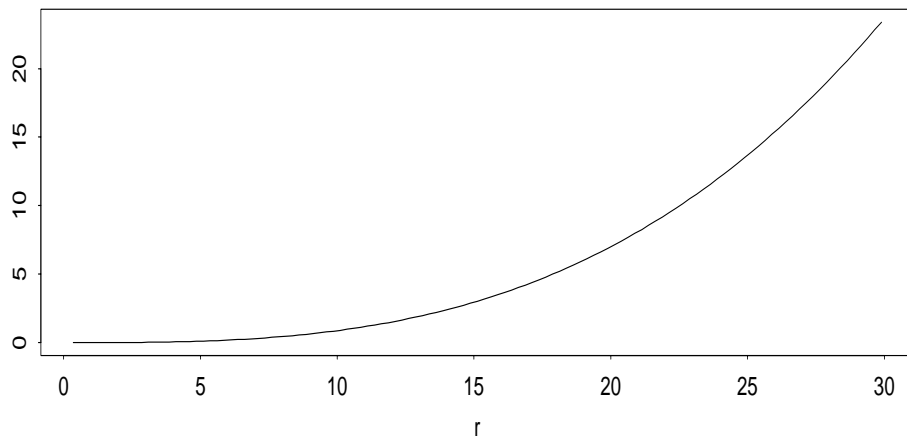


Figure 5.4

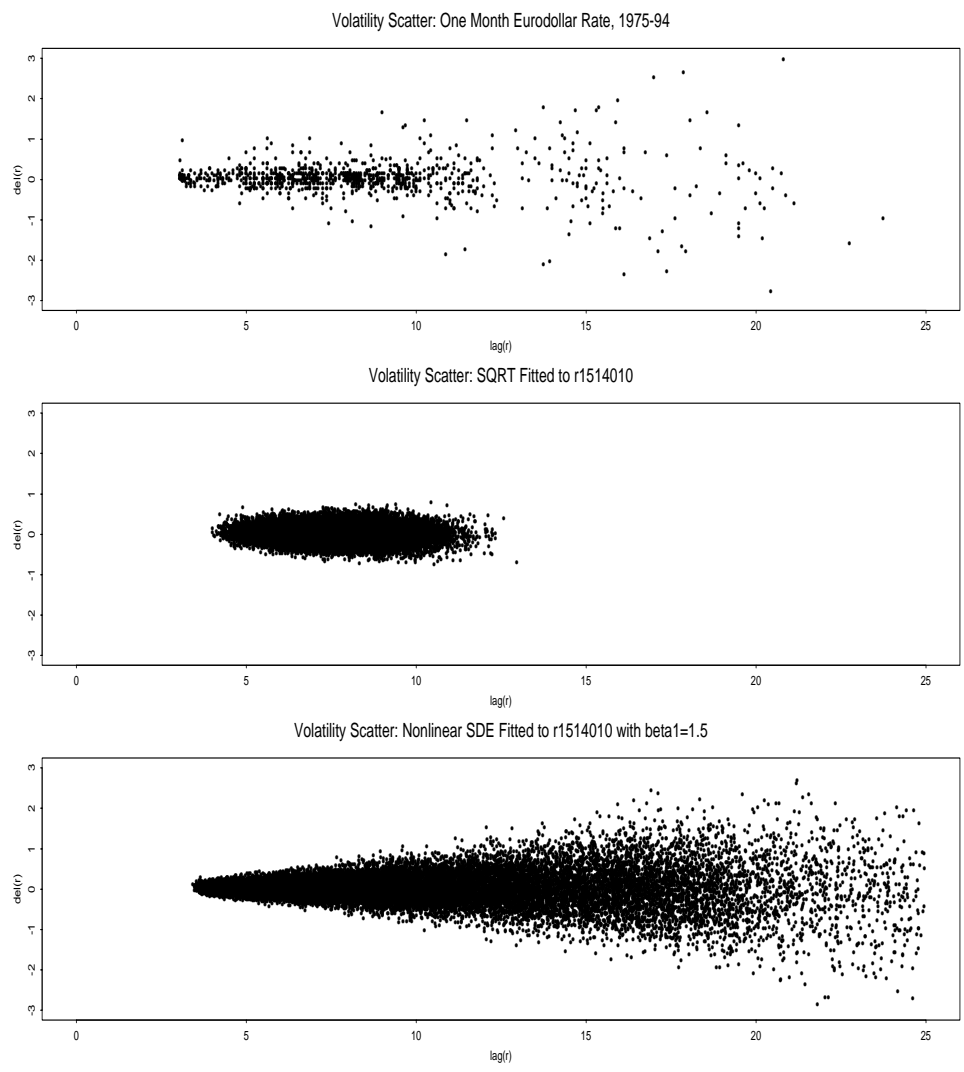
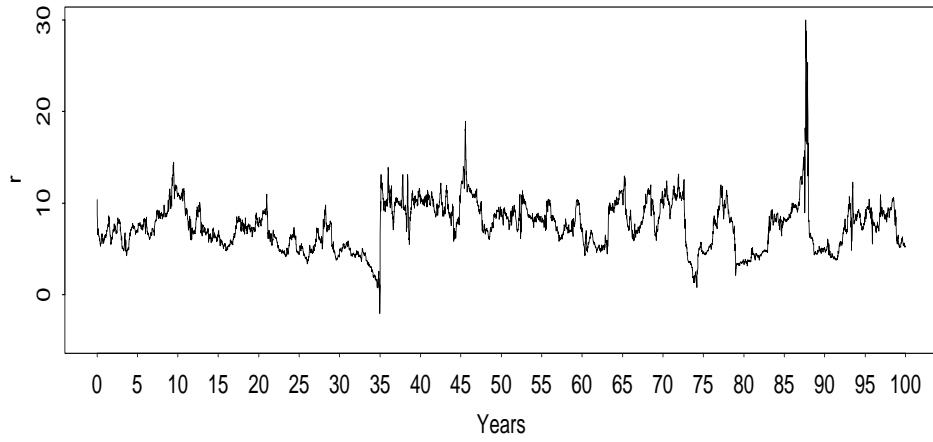


Figure 5.5

Long simulation from fitted SNP (r1514010)



Long simulation from fitted nonlinear SDE with beta1=1.5

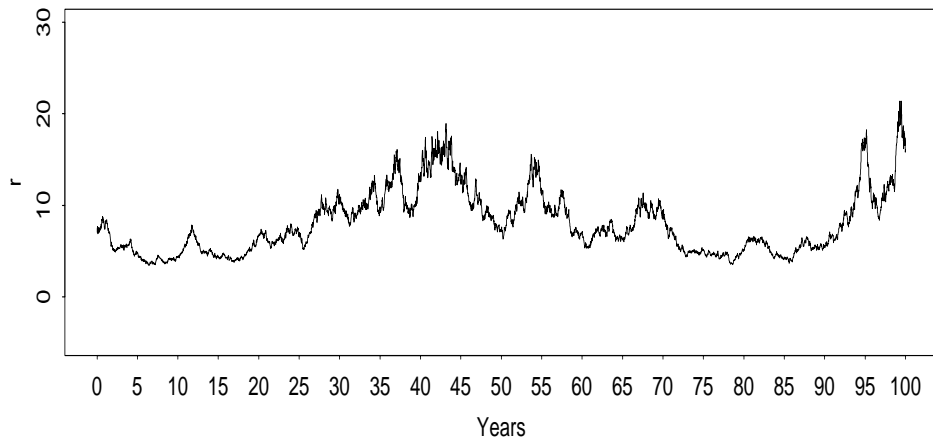


Figure 5.6

Acknowledgement Footnote

Prepared for the Seventh World Congress of the Econometric Society, Tokyo, Japan, August 22–29, 1995. Corresponding author: George Tauchen, Department of Economics, Duke University, Durham, NC 27708, (919) 660-1812. The most recent version of this paper is the PostScript file `srvy.ps` available by anonymous ftp to `ftp.econ.duke.edu`, subdirectory `home/get/papers`. The material is based upon work supported by the National Science Foundation. Thanks go to Yacine Ait-Sahalia, Torben Andersen, Lars Peter Hansen, and Peter Schotman for carefully reading and commenting on an earlier version and to Ravi Bansal, A. Ronald Gallant, Adrian Pagan, and Ken Singleton for useful discussions and remarks at various stages of this research. Much of this paper was written while I was a Visiting Fellow at the Australian National University.

Note to Table 5.1

Note:

L_u is the number of lags in the linear part of the SNP model; L_r is the number of lags in the ARCH part; L_p is the number of lags in the polynomial part, $P(z, x)$. The polynomial $P(z, x)$ is of degree K_z in z and K_x in x ; by convention, $L_p = 1$ if $K_x = 0$. ℓ_θ is the number of free parameters associated with the SNP model. The data set is $\{y_t\}_{t=21}^{1035}$, with 20 initial values reserved for lags. The symbol \rightarrow indicates Semiparametric ARCH; \Rightarrow indicates Nonlinear Nonparametric.

Note to Table 5.2

Note:

The first column lists the score generator model employed in the EMM estimation, where SP-ARCH is the Semiparametric ARCH model and Nonlinear-NP is the Nonlinear Nonparametric model. The remaining columns show the estimated values of the parameters α and β of the stochastic differential equation (5.1), where the drift and diffusion functions are defined in (5.10)–(5.15), along with the EMM objective function from estimation using the weekly Eurodollar interest rate data.