

Specification Analysis of Continuous Time Models in Finance ¹

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

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

First Draft: October 1995

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

Abstract

The paper describes the use of the Gallant-Tauchen efficient method of moments (EMM) technique for diagnostic checking of stochastic differential equations (SDEs) estimated from financial market data. The EMM technique is a simulation-based method that uses the score function of an auxiliary model, called the score generator, to define a generalized method of moments (GMM) objective function. The technique can handle multivariate SDEs where the state vector is not completely observed. The optimized GMM objective function is distributed as chi-square and may be used to test model adequacy. Elements of the score function that have large values reflect features of data that a rejected SDE specification does not describe well and may be used for diagnostic checking. The diagnostics are illustrated by estimating a Yield-Factor Model from weekly, 1962–1995, term structure data comprised of short (3 month), medium (12 month), and long (10 year) Treasury rates. The Yield-Factor Model is sharply rejected, although an extension that permits the local variance function to be a convex function of affine functions of the interest rates comes much closer to describing the data.

1 Introduction

This is an expository paper that describes how the efficient method of moments (EMM) estimator proposed by Gallant and Tauchen (1996) can be used for estimation and specification analysis of a system of stochastic differential equations (SDEs). The paper is a summary of previous work. Ideas are illustrated by a yield curve application, and sources of software are indicated.

The EMM estimator is a generalized method of moments (GMM) estimator in which the moment function that enters the GMM criterion is the expectation of the score function of an auxiliary model for the data. We refer to the auxiliary model as the score generator hereafter. The estimator follows the line work on dynamic simulation estimators initiated by Ingram and Lee (1991) and Duffie and Singleton (1993). These estimators use long simulations to compute expectations given a candidate value of the parameter vector. An important feature of simulation estimators is that they are applicable when the state vector is partially observed. The distinguishing characteristic of the EMM method is the recommendation of specific moment functions that guarantee efficiency and facilitate specification analysis. EMM is less computationally demanding than the simulation estimator proposed by Gourieroux, Monfort, and Renault (1993) and Broze, Scaillet, and Zakoian (1995) because a binding function does not need to be computed at each candidate parameter value nor does a Hessian need to be computed.

Alternative estimation strategies for scalar stochastic differential equations are due to Aït-Sahalia (1995a, 1995b) and Hansen and Scheinkman (1995). These strategies rely on moment functions computed directly from the data, rather than moment functions computed by simulation. They require the state to be fully observed in order to estimate all of the parameters. Lo (1988) discusses maximum likelihood methods. Maximum likelihood methods are difficult to implement when the state is partially observed and do not provide statistics for specification analysis.

As noted above, appropriate choice of the score generator in EMM estimation guarantees efficiency and protects against specification error (Gallant and Long, 1995; Tauchen, 1995). Specifically, the score generator must provide a complete statistical description of the ob-

served series. Here we use the seminonparametric (SNP) density proposed by Gallant and Tauchen (1989). The SNP density has three identifiable components: a VAR-type location function, an ARCH-type scale function, and a Hermite polynomial. The Hermite polynomial accommodates general features of the data beyond those embodied in the location and scale functions which justifies a nonparametric interpretation of SNP estimates. The value of the optimized EMM objective function is chi-square, and thus can be used for forming confidence intervals and testing system adequacy. If a fitted stochastic differential equation is rejected, then the studentized scores associated with the parameters of the three components of the SNP density suggest how the system can be modified to improve the fit.

The paper is organized as follows. In Section 2 we describe the SNP nonparametric time series estimator. In Section 3 we describe the EMM estimator. In Section 4 we describe application to SDEs and describe SDE simulation schemes. In Section 5 we illustrate by application to continuous time models of the term structure of interest rates based on the Yield Factor Model of Duffie and Kan (1993). Section 6 contains concluding remarks.

2 SNP Nonparametric Time Series Analysis

Let $\{y_t\}_{t=-\infty}^{\infty}$, $y_t \in \mathfrak{R}^M$, be a stationary, multiple, discrete time series that is adapted to a filtration $\{\mathcal{F}_t\}_{t=-\infty}^{\infty}$ (Karatzas and Shreve, 1991, p. 4) and that is Markovian in L lags in the sense that conditional expectation satisfies $\mathcal{E}(y_t|\mathcal{F}_t) = \mathcal{E}(y_t|x_{t-1})$ where $x_{t-1} = (y_{t-L}, \dots, y_{t-1})$.

The stationary distribution of the process is presumed to have a density $p(y_{-L}, \dots, y_0)$ which is defined over \mathfrak{R}^ℓ , $\ell = M(L + 1)$. Put $y = y_0$, $x = x_{-1} = (y_{-L}, \dots, y_{-1})$, and write the stationary, marginal, and conditional densities of the process $\{y_t\}_{t=-\infty}^{\infty}$ as $p(x, y) = p(y_{-L}, \dots, y_0)$, $p(x) = \int p(y_{-L}, \dots, y_0) dy_0$, $p(y|x) = p(x, y)/p(x)$ respectively. Let $\{\tilde{y}_t\}_{t=-L}^n$ denote realization from the process $\{y_t\}_{t=-\infty}^{\infty}$; that is, $\{\tilde{y}_t\}_{t=-L}^n$ denotes data and $\{y_t\}_{t=-\infty}^{\infty}$ denotes the random variables to which the data correspond. We require estimates of the conditional density $p(y|x)$ rather than the stationary density $p(x, y)$ and therefore propose to estimate $p(y|x)$ directly rather than indirectly via an estimate of $p(x, y)$.

We shall describe a class of conditional densities

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

proposed by Gallant and Tauchen (1989), which they termed SNP for seminonparametric, that has two properties: (1) The union $\mathcal{H} = \bigcup_{K=1}^{\infty} \mathcal{H}_K$ is quite rich and it is reasonable to assume that the true density $p(y|x)$ of stationary data from a financial market is contained in \mathcal{H} . (2) If θ is estimated by quasi-maximum likelihood, viz.

$$\tilde{\theta}_n = \arg \max_{\theta \in \mathfrak{R}^{p_K}} \frac{1}{n} \sum_{t=0}^n \log[f_K(\tilde{y}_t | \tilde{y}_{t-L}, \dots, \tilde{y}_{t-1}, \theta)],$$

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

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

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

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

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

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

The location function μ_x is affine in x

$$\mu_{x_{t-1}} = b_0 + Bx_{t-1} \tag{2.1}$$

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

The scale function R_x is affine in the absolute value of the difference between x and the location function μ_x

$$\text{vech}(R_{x_{t-1}}) = \rho_0 + P|x_{t-1} - \mu_{x_{t-2}}| \quad (2.2)$$

where $\text{vech}(R)$ denotes a vector of length $M(M+1)/2$ containing the elements of the upper triangle of R and $|x|$ denotes elementwise absolute value. It is presumed to depend on $L_R + L_\mu \leq L$ lags which is accomplished by putting leading columns of P to zero as required. For $M > 1$ we recommend setting the elements of P that correspond to off-diagonal elements of R_x to zero. Note that were we to eliminate the dependence of the innovation density on x by writing $h(z)$ instead of $h(z|x)$ then $\{y_t\}_{t=-\infty}^{\infty}$ would be an ARCH-type process akin to that proposed by Nelson (1991). Since the absolute value function is not differentiable, $|u|$ is approximated in the formula for R_x above by the twice continuously differentiable function

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

Let $z^\alpha = z_1^{\alpha_1} \cdots z_M^{\alpha_M}$, $|\alpha| = \sum_{k=1}^M \alpha_k$, similarly for x^β , and consider the density

$$h_K(z|x) = \frac{[P_K(z, x)]^2 \phi(z)}{\int [P_K(u, x)]^2 \phi(u) du} \quad (2.3)$$

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 $\phi(z) = (2\pi)^{-M/2} e^{-z'z/2}$. $P_K(z, x)$ is a polynomial of degree K_z in z whose coefficients are, in turn, polynomials of degree K_x in x . The product $[P_K(z, x)]^2 \phi(z)$ is a Hermite polynomial in z with positivity enforced whose coefficients depend on x . The shape of the innovation density $h_K(z_t|x_{t-1})$ varies with x_{t-1} which permits $h_K(z_t|x_{t-1})$ to exhibit general, conditional shape heterogeneity. By putting selected elements of the matrix $A = [a_{\beta\alpha}]$ to zero $P_K(z, x)$ can be made to depend on only $L_p \leq L$ lags from x . Also, when M is large, coefficients $a_{\beta\alpha}$ corresponding to monomials z^α that represent high order interactions can be set to zero with little effect on the adequacy of approximations. Let $I_z = 0$ indicate that no interaction coefficients are set to zero, $I_z = 1$ indicate that coefficients corresponding to interactions z^α of order larger than $K_z - 1$ are set to zero, and so on; similarly for x^β and

I_x . One may note that if K_z is put to zero, then the innovation density $h_K(z|x)$ is Gaussian. If $K_z > 0$ and $K_x = 0$, then the density can assume arbitrary shape but innovations are homogeneous.

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

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

completes the description of the SNP density. The vector θ contains the coefficients $A = [a_{\beta\alpha}]$ 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 the dimension p_K of θ .

Some structural characteristics of $f_K(y_t|x_{t-1}, \theta)$ may be noted. If $K_z, K_x,$ and L_r are put to zero, then $f_K(y_t|x_{t-1}, \theta)$ is a Gaussian vector autoregression. If K_x and L_r are put to zero, then $f_K(y_t|x_{t-1}, \theta)$ is a non-Gaussian vector autoregression with homogeneous innovations. If K_z and K_x are put to zero, then $f_K(y_t|x_{t-1}, \theta)$ is a Gaussian ARCH model. If K_x is put to zero, then $f_K(y_t|x_{t-1}, \theta)$ is a non-Gaussian ARCH model with homogeneous innovations. If $K_z > 0, K_x > 0, L_p > 0, L_\mu > 0,$ and $L_r > 0$, then $f_K(y_t|x_{t-1}, \theta)$ is a 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. At present we recommend moving along an upward expansion path using the BIC criterion (Schwarz, 1978)

$$\begin{aligned} \text{BIC} &= s_n(\tilde{\theta}) + (1/2)(p_K/n) \ln(n), \\ s_n(\theta) &= -\frac{1}{n} \sum_{t=0}^n \log[f_K(\tilde{y}_t|\tilde{y}_{t-L}, \dots, \tilde{y}_{t-1}, \theta)], \end{aligned}$$

to guide the search, models with small values of BIC being preferred. An alternative to BIC is the Hannan and Quinn (Hannan, 1987) criterion

$$\text{HQ} = s_n(\hat{\theta}) + (p_\theta/n) \ln[\ln(n)].$$

The expansion path has a tree structure. Rather than examining the full tree, our strategy is to expand first in L_u with $L_r = L_p = K_z = K_x = 0$ until BIC turns upward. We then

expand in L_r with $L_p = K_z = K_x = 0$. Next we expand K_z with $K_x = 0$ and lastly L_p and K_x . We also expand 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. Frequently this process terminates with $K_x = 0$. As a precaution we compute the residuals

$$\tilde{\epsilon}_t = [\text{Var}(y_t|x_{t-1})]^{-1/2} [y_t - \mathcal{E}(y_t|x_{t-1})]$$

with $\mathcal{E}(\cdot|x)$ and $\text{Var}(\cdot|x)$ computed under $f_K(y|x, \tilde{\theta})$ and test for the significance of a regression of $\tilde{\epsilon}_t$ on a polynomial of degree three in x_{t-1} . This regression often detects a need to increase L_p and K_x and also suggests the appropriated value for L_p . The tuning parameter selection strategy just described has given reasonable results in much applied work. See Fenton and Gallant (1996b) and the references therein.

Fortran code, a User's Guide in PostScript (Gallant and Tauchen, 1995a), and PC executables for quasi-maximum likelihood estimation of the parameters of the SNP density are available by anonymous ftp at site ftp.econ.duke.edu in directory pub/arg/snp. The Fortran code and User's Guide are also available from the Carnegie-Mellon University e-mail server by sending the one-line e-mail message "send snp from general" to statlib@lib.stat.cmu.edu.

3 Efficient Method of Moments

Let $\{y_t\}_{t=-\infty}^{\infty}$, $y_t \in \mathbb{R}^M$, be a stationary, multiple, discrete time series. The stationary distribution of a stretch y_{t-L}, \dots, y_t from the process $\{y_t\}_{t=-\infty}^{\infty}$ is hypothesized to have density $p(y_{-L}, \dots, y_0|\rho)$ where ρ is a vector of unknown parameters that must be estimated. Put $y = y_0$, $x = x_{-1} = (y_{-L}, \dots, y_{-1})$, and write $p(x, y|\rho) = p(y_{-L}, \dots, y_0|\rho)$, $p(x|\rho) = \int p(y_{-L}, \dots, y_0|\rho) dy_0$, $p(y|x, \rho) = p(x, y|\rho)/p(x|\rho)$. Let $\{\tilde{y}_t\}_{t=-L}^n$ denote the data from which ρ is to be estimated.

We are interested in the situation where an analytic expression for the maintained model $p(y_{-L}, \dots, y_0|\rho)$ is not available yet expectations of the form

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

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 $p(y|x, \rho)$ and puts

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

with N large enough that Monte Carlo error is negligible. Examples of this situation include representative agent asset pricing models, general equilibrium rational expectations models, auction models, and, of particular interest here, systems of stochastic differential equations.

Our objective is threefold: (1) estimate ρ ; (2) test the hypothesis that $p(y_{-L}, \dots, y_0|\rho)$ is the stationary distribution the process $\{y_t\}_{t=-\infty}^\infty$; and, (3) provide diagnostics that indicate how a rejected model should be modified to better describe the distribution of the process.

Gallant and Tauchen (1996) proposed an estimator for estimating ρ in the situation above. Termed the efficient method of moments (EMM) estimator, it would be regarded as a minimum chi-square estimator in the statistics literature and as a generalized method of moments (GMM) estimator in the econometrics literature. Being minimum chi-square, the optimized chi-square criterion can be used to test model adequacy. Also, as seen below, the moments that enter the criterion provide diagnostics that indicate how the maintained model $p(y_{-L}, \dots, y_0|\rho)$ should be modified if it is rejected by the test of model adequacy.

The moment equations for the EMM estimator are obtained from the score vector $(\partial/\partial\theta) \ln f(y|x, \theta)$ of an auxiliary model $f(y|x, \theta)$ that is termed the score generator. Gallant and Tauchen (1996) show that if the score generator $f(y|x, \theta)$ encompasses the maintained model $p(y|x, \rho)$, then their estimator is as efficient as maximum likelihood. Tauchen (1995) sets forth formulae that would lead one to expect that 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, etc. Gallant and Long (1995) support this conjecture by showing that if the score generator is the SNP density $f_K(y|x, \theta)$ described in Section 2, then the efficiency of the EMM estimator can be made as close to that of maximum likelihood as desired by taking K large enough. Moreover, they show that if $\{y_t\}_{t=-\infty}^\infty$ is stationary but not Markovian, then the same is true for large L .

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

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

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

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

and the corresponding estimate of the information matrix

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

If the score generator $f(y|x, \theta)$ is not an adequate statistical approximation to the data, then one of the more complicated expressions for $\tilde{\mathcal{I}}_n$ set forth in Gallant and Tauchen (1996) must be used. If the SNP density $f_K(y|x, \theta)$ of Section 2 is used as the score generator, and the model selection protocol of Section 2 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 above to be appropriate (Gallant and Long, 1995).

Define

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

which is computed by averaging over a long simulation

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

The estimator is

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

The asymptotics of the estimator, which are derived in Gallant and Tauchen (1996), are as follows. If ρ° denotes the true value of ρ and θ° is an isolated solution of the moment equations $m(\rho^\circ, \theta) = 0$, then

$$\begin{aligned} \lim_{n \rightarrow \infty} \hat{\rho}_n &= \rho^\circ \quad \text{a.s.} \\ \sqrt{n}(\hat{\rho}_n - \rho^\circ) &\xrightarrow{\mathcal{L}} N\left\{0, [(M^\circ)'(\mathcal{I}^\circ)^{-1}(M^\circ)]^{-1}\right\} \\ \lim_{n \rightarrow \infty} \hat{M}_n &= M^\circ \quad \text{a.s.} \\ \lim_{n \rightarrow \infty} \tilde{\mathcal{I}}_n &= \mathcal{I}^\circ \quad \text{a.s.} \end{aligned}$$

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

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

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

$$L_0 = n m'(\hat{\rho}_n, \tilde{\theta}_n)(\tilde{\mathcal{I}}_n)^{-1} m(\hat{\rho}_n, \tilde{\theta}_n) \quad (3.1)$$

is asymptotic chi-square on $p_\theta - p_\rho$ degrees freedom. Under the null hypothesis that $h(\rho^\circ) = 0$, where h maps \mathfrak{R}^{p_ρ} into \mathfrak{R}^q ,

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

is asymptotic chi-square on q degrees freedom where

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

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 i -th diagonal element of $\hat{\Sigma} = (1/n)[(\hat{M}_n)'(\tilde{\mathcal{I}}_n)^{-1}(\hat{M}_n)]^{-1}$. These intervals, which are symmetric, are somewhat misleading because they do not reflect the rapid increase in the EMM objective function $s_n(\rho) = m'(\rho, \tilde{\theta}_n)(\tilde{\mathcal{I}}_n)^{-1} m(\rho, \tilde{\theta}_n)$ when ρ_i approaches a value for which simulations from $p(y_{-L}, \dots, y_0 | \rho)$ become explosive. Confidence intervals obtained by inverting the criterion difference test L_h do reflect this phenomenon and are therefore more useful. To invert the test one puts in the interval those ρ_i^* for which L_h for the hypothesis $\rho_i^\circ = \rho_i^*$ is less than the critical point of a chi-square on one degree freedom. To avoid re-optimization one may use the approximation

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

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

When $L_0 = n m'(\hat{\rho}_n, \tilde{\theta}_n)(\tilde{\mathcal{I}}_n)^{-1} m(\hat{\rho}_n, \tilde{\theta}_n)$ exceeds the chi-square critical point diagnostics that suggest improvements to the model are desirable. Because

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

inspection of the t -ratios

$$T_n = S_n^{-1} \sqrt{n} m(\hat{\rho}_n, \tilde{\theta}_n), \quad (3.2)$$

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

$$\hat{T}_n = [(\text{diag}\tilde{\mathcal{I}}_n)^{1/2}]^{-1} \sqrt{n} m(\hat{\rho}_n, \tilde{\theta}_n), \quad (3.3)$$

which are under-estimates, may be adequate and are cheaper to compute because they avoid numerical approximation to \hat{M}_n . We investigate the adequacy of quasi- t -ratios in Section 5.

Fortran code and a User's Guide in PostScript (Gallant and Tauchen, 1995b) that implement the EMM estimator using $f_K(y_t|x_{t-1}, \theta)$ as described in Section 2 as the score generator are available by anonymous ftp at site ftp.econ.duke.edu in directory pub/get/emm.

4 Stochastic Differential Equations

Consider application of the EMM estimation and inference methods described in Section 3 to a system of stochastic differential equations

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

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

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

where U_0 is the initial condition at time $t = 0$, and $\int_0^t B_i(U_s, \rho) dW_{is}$ denotes the Ito stochastic integral (Karatzas and Schreve, 1991).

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

$$y_t = T(U_{t+L}, \rho) \quad t = -L, -L + 1, \dots$$

of the state are recorded, where y_t is an M -dimensional vector and $L > 0$ is the number of lagged variables that enter the formulas of Section 3.

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

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

$$\left. \begin{aligned} y_{1t} &= U_{2,t+L} \\ y_{2t} &= U_{3,t+L} \end{aligned} \right\} t = -L, L+1, \dots, n$$

U_{1t} represents an unobserved flow of new information to the market that influences the volatility of asset prices U_{2t} and U_{3t} by changing the instantaneous conditional variances of U_{2t} and U_{3t} . The observed data \tilde{y}_{1t} and \tilde{y}_{2t} are prices at the end of each trading day. To achieve identification in estimation, a normalization rule such as $\rho_9 = 1$ should be imposed.

We assume that U_t , and hence y_t , is stationary and ergodic. We further assume that the stationary distribution of y_t is absolutely continuous. Thus, for each setting of parameter ρ and lag length L , there exists a time invariant density $p(y_{-L}, \dots, y_0 | \rho)$ of the form described in Section 3 such that

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

where $\{\hat{y}_t\}_{t=-L}^N$ is realization of length $N + L + 1$ from the system. This assumes that g is integrable and that either U_0 is a sample from the stationary distribution of U_t or that a longer realization was observed and enough initial observations were discarded for transients to have dissipated. Therefore, all that is needed for implementation of the EMM methodology is a practical means for generating the simulation $\{\hat{y}_t\}_{t=-L}^N$.

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

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

may be obtained by first generating the sequence

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

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

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

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

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

A simulation $\{\hat{y}_t\}_{t=-L}^N$ for graphical display or for computation of statistics such as density estimates may be generated similarly using an explicit order 1 strong scheme, which is described below. For the results in Subsection 5, time t is in weeks and $n_0 = 10$, which implies that the simulation step-size is $\Delta = 1/10$.

The two schemes used here are below. They represent certain choices of tuning parameters on our part to schemes from Kloeden and Platen (1992, p. 347, p. 376, p. 486). The weak scheme is valid for autonomous systems $dU_t = A(U_t, \rho)dt + B(U_t, \rho)dW_t$ only. The strong scheme is valid for non-autonomous systems $dU_t = A(t, U_t, \rho)dt + B(t, U_t, \rho)dW_t$ as well. Fortran code implementing them are available by anonymous ftp at site ftp.econ.duke.edu in directory pub/arg/libf as files stng1.f and weak2.f.

4.1 Explicit Order 2 Weak Scheme

Recursion

$$\begin{aligned} \hat{U}_{t+\Delta} &= \hat{U}_t + \frac{1}{2} [A(\Upsilon, \rho) + A(\hat{U}_t, \rho)] \Delta \\ &+ \frac{1}{4} \sum_{j=1}^k \left\{ [B_j(R_j^+, \rho) + B_j(R_j^-, \rho) + 2B_j(\hat{U}_t, \rho)] \Delta W_j \right. \\ &\quad \left. + \sum_{\substack{r=1 \\ r \neq j}}^k [B_j(Y_r^+, \rho) + B_j(Y_r^-, \rho) - 2B_j(\hat{U}_t, \rho)] \Delta W_j \Delta^{-1/2} \right\} \\ &+ \frac{1}{2} \sum_{j=1}^k \left\{ [B_j(R_j^+, \rho) - B_j(R_j^-, \rho)] I_{jj} + \sum_{\substack{r=1 \\ r \neq j}}^k [B_j(Y_r^+, \rho) - B_j(Y_r^-, \rho)] I_{rj} \right\} \Delta^{-1/2} \end{aligned}$$

Supporting values

$$\Upsilon = \hat{U}_t + A(\hat{U}_t, \rho) \Delta + \sum_{j=1}^k B_j(\hat{U}_t, \rho) \Delta W_j$$

$$\begin{aligned}
R_j^\pm &= \hat{U}_t + A(\hat{U}_t, \rho)\Delta \pm B_j(\hat{U}_t, \rho)\Delta^{1/2} \\
Y_j^\pm &= \hat{U}_t \pm B_j(\hat{U}_t, \rho)\Delta^{1/2}
\end{aligned}$$

Integral approximation

$$\begin{aligned}
I_{rj} &= (1/2)[\Delta W_j \Delta W_r + V_{rj}] \\
V_{rj} &= -\Delta I_{(0, \frac{1}{2}]}(U_{rj}) + \Delta I_{(\frac{1}{2}, 1]}(U_{rj}) & r < j \\
V_{rj} &= -\Delta & r = j \\
V_{rj} &= -V_{jr} & r > j
\end{aligned}$$

Independent random variables

$$\begin{aligned}
\Delta W_j &\sim N(0, \Delta) & j = 1, \dots, k \\
U_{rj} &\sim U(0, 1] & r = 1, \dots, j-1, \quad j = 1, \dots, k
\end{aligned}$$

4.2 Explicit Order 1 Strong Scheme

Recursion

$$\begin{aligned}
\hat{U}_{t+\Delta} &= \hat{U}_t + A(t, \hat{U}_t, \rho)\Delta + \sum_{j=1}^k B_j(t, \hat{U}_t, \rho)\Delta W_j \\
&\quad + \frac{1}{\sqrt{\Delta}} \sum_{j=1}^k \sum_{r=1}^k [B_j(t, \Upsilon_r, \rho) - B_j(t, \hat{U}_t, \rho)] I_{rj}
\end{aligned}$$

Supporting values

$$\Upsilon_j = \hat{U}_t + A(t, \hat{U}_t, \rho)\Delta + B_j(t, \hat{U}_t, \rho)\sqrt{\Delta}$$

Integral approximation

$$\begin{aligned}
I_{rj} &= (1/2)[(\Delta W_j)^2 - \Delta] & r = j \\
I_{rj} &= (1/2)\Delta W_r \Delta W_j + (\Delta C)^{1/2}(\mu_r \Delta W_j - \mu_j \Delta W_r) & r \neq j \\
&\quad + \frac{\Delta}{2\pi} \sum_{\ell=1}^p \frac{1}{\ell} \left[\zeta_{r\ell} \left(\frac{\Delta W_j}{\sqrt{\Delta/2}} + \eta_{j\ell} \right) - \zeta_{j\ell} \left(\frac{\Delta W_r}{\sqrt{\Delta/2}} + \eta_{r\ell} \right) \right] \\
C &= \frac{1}{12} - \frac{1}{2\pi^2} \sum_{\ell=1}^p \frac{1}{\ell^2} & p = 50
\end{aligned}$$

Independent random variables

$$\begin{aligned} \Delta W_j &\sim N(0, \Delta) & j = 1, \dots, k \\ \mu_j &\sim N(0, 1) & j = 1, \dots, k \\ \eta_{j\ell} &\sim N(0, 1) & j = 1, \dots, k, \quad \ell = 1, \dots, p \\ \zeta_{j\ell} &\sim N(0, 1) & j = 1, \dots, k, \quad \ell = 1, \dots, p \end{aligned}$$

5 An Application to the Term Structure

5.1 Data and Score Generator

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

On this data set we estimated by quasi-maximum likelihood the parameters θ of the SNP density $f_K(y|x, \theta)$ under the restriction $K_x = 0$. Following the model selection protocol described in Section 2 we obtain $L_u = 3$, $L_r = 4$, $K_z = 4$, $I_z = 3$ for the tuning parameters. This model takes the form of an ARCH model with conditionally homogeneous non-Gaussian errors. The error density is a modified Hermite density as discussed in Section 2 above. The polynomial $P(z)$ is a quartic in $z \in \mathfrak{R}^3$ with all interactions suppressed. We call the score from this fit the *Semiparametric ARCH Score*, which is of length $p_\theta = 60$.

Below, we use this score generator for EMM estimation of continuous time models of the term structure. EMM applied to a score from an SNP density constrained so that $K_x = 0$ is consistent and asymptotically normal, so long as the underlying continuous time model is correctly specified (Gallant and Tauchen, 1996), albeit with a possible efficiency loss (Gallant and Long, 1995). Failure to fit this score can, of course, be construed as sharp evidence against the specification of the continuous time model. However, for reasons discussed in Tauchen (1995), successful fitting of this score does not necessarily signal correct specification

of the continuous time model, as this score need not incorporate all salient features of the data.

5.2 Generalized Power Specifications

Let $\{U_t\}$, $t \geq 0$, denote the continuous record on the three interest rates for which $\tilde{y}_{t-L} = U_t$, $t = 0, 1, 2, \dots$. We assume U_t is generated by the autonomous system

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

where $A(U, \rho)$ is a 3×1 vector-valued function and $B(U, \rho)$ is 3×3 matrix-valued function, with $U \in \mathfrak{R}^3$ and $\rho \in \mathfrak{R}^{\ell_\rho}$.

We examine empirical characteristics of a multivariate power specification in which the drift and diffusion functions take the form

$$A(U, \rho) = \alpha_0 + \alpha_1 U \tag{5.1}$$

$$B(U, \rho) = \beta_0 \text{diag}[(\beta_1 + \beta_2 U)^\gamma] \tag{5.2}$$

where α_0 is 3×1 , α_1 is 3×3 , β_0 is 3×3 , β_1 is 3×1 , β_2 is 3×3 , γ is 3×1 , ρ contains the free elements of α_0 , α_1 , β_0 , β_1 , β_2 , γ , and

$$\text{diag}(V^\gamma) = \begin{bmatrix} V_1^{\gamma_1} & 0 & 0 \\ 0 & V_2^{\gamma_2} & 0 \\ 0 & 0 & V_3^{\gamma_3} \end{bmatrix}$$

for $V = \beta_1 + \beta_2 U \in \mathfrak{R}^3$. For $v, \psi \in \mathfrak{R}_1$ we interpret $v^\psi = \text{sign}(v)|v|^\psi$ as the signed power function. Using the signed power function makes the diffusion function globally defined, which is tremendously convenient in empirical work, though it does mean some parameters are identified only up to sign.

The multivariate power specification (5.1)-(5.2) subsumes the Yield-Factor model of Duffie and Kan (1993), which is obtained by taking $\gamma = (0.50 \ 0.50 \ 0.50)$. The Yield-Factor model is the multivariate generalization of the square-root model of Cox, Ingersoll, and Ross (1985). It is analytically convenient, because the local variance function $B(U, \rho)B(U, \rho)'$ is affine in U , which generates closed form expressions for bond prices. If U_t itself is viewed

as the vector of fundamental factors, then the bond prices are log-linear functions of U_t . This simplicity is lost, though, if $\gamma \neq (0.50 \ 0.50 \ 0.50)$. In this case, the multivariate power specification is a natural multivariate generalization of the univariate constant elasticity of variance specification developed by Chan, Karolyi, Longstaff, and Sanders (1992).

5.3 Estimation

We employ the EMM method described in Section 3 to estimate the multivariate power specification. In computation, we take

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

and use an explicit order 2 weak scheme as described in Subsection 4.1 above. For this work, time t is scaled so the interval $[t, t+1]$ is one week and $n_0 = 10$, which implies the simulation step size is $\Delta = 0.10$. Based on Gallant and Tauchen (1995c) and Tauchen (1995) we use $N = 50,000$ throughout. With the elements of γ pinned ρ contains $p_\rho = 30$ parameters. Since $p_\theta = 60$, then the scaled objective function L_0 , defined in (3.1) above, follows a $\chi^2(30)$ under the null hypothesis of correct specification of the SDE.

Estimating the multivariate power specification (5.1)-(5.2) with $\gamma = (0.50 \ 0.50 \ 0.50)$ gives $L_0 = 133.563$, p -value $< 10^{-12}$. Evidently, an affine local variance function is sharply at odds with the data.

Figures 2–4 reveal the empirical shortcomings of the specification with $\gamma = (0.50 \ 0.50 \ 0.50)$. The figures show the t diagnostics presented in Section 3 for the three groups of parameters of the SNP density. These diagnostics are the studentized elements of the score vector $m(\hat{\rho}_n, \tilde{\theta}_n)$. The figures contain both forms of the diagnostics in order to assess robustness. The quasi- t -ratios, defined in (3.3) above and shown in the top panels of the figures, are very easy to compute but biased downward in magnitude relative to 2.0. The t -ratios, defined in (3.2) above and shown in the bottom panels, are asymptotically unbiased indicators of misspecification but more difficult to compute. Numerical first derivatives were used to compute the Jacobian matrix estimate \hat{M}_n .

Figure 2 shows the t -ratio diagnostics on the elements of the score vector for the parameters of the location $\mu_{x_{t-1}}$ defined in (2.1). The t -ratio on the intercept in the TBILL12M equa-

tion suggests this specification has trouble fitting the unconditional mean of the TBILL12M series, though the other t -ratios would be considered statistically insignificant.

Figure 3 shows the t -ratio diagnostics on the elements of the score vector corresponding to the scale function $R_{x_{t-1}}$ defined in (2.2). Nearly all t -ratios on the scale function scores are above 2.0 in magnitude, indicating that this specification has trouble tracking the conditional second moments of all three interest rates. The signs of the t -ratios are the same because of correlations across the scores.

Figure 4 shows the t -ratio diagnostics on the elements corresponding to the Hermite polynomial parameters. From this figure it is seen that the specification fits the scores on the odd powers but fails to fit those on the the even powers of the polynomial. From extensive previous experience with SNP, we know that the even powers of the Hermite tend to control the tail-thickness while odd powers control asymmetries, though the association between degree and these features of the density is not quite exact because the polynomial is squared in (2.3). The evidence in Figure 4 suggests that the predicted discrete-time innovation density from the fitted SDE is not sufficiently leptokurtic relative to the Gaussian density.

From findings reported in Conley, Hansen, and Scheinkman (1995), Aït-Sahalia (1995a), and Tauchen (1995), who estimate continuous time models on univariate short-rate data, one expects a specification with exponents above 0.50 to do better. A very detailed examination of the objective is beyond the scope of this paper, and some preliminary findings are reported in Gallant and Tauchen (1995c). Here we re-estimate the SDE specification (5.1)-(5.2) with $\gamma = (0.80 \ 0.80 \ 0.80)$, which gives $L_0 = 57.089$, p -value = 0.00169. Though this specification does not quite accommodate the Semiparametric ARCH Score, the improvement in fit indicates that a local variance function that is convex in linear combinations of the interest rates is needed to fit the data.

Figures 5–7 show the t -ratio diagnostics on this specification. Comparing these three figures to to Figures 2–4 shows how the increase in the common value of the exponent diminishes many of the problems with the fit. The scores for the parameters associated with $\mu_{x_{t-1}}$ are reduced to to insignificance as are many of those for the parameters of $R_{x_{t-1}}$. Perhaps the most noticeable improvement is in the Hermite scores (Figures 7 versus Figure 4) indicating that higher value of the exponent generates higher order non-Gaussianity in the

data.

Interestingly, the top versus the bottom panels of Figures 6 and Figure 7 show that the adjustments to the standard errors of the scores matter. There are several scores with quasi- t -ratios below 2.0 in magnitude in the top panels but would be judged statistically significant on the basis of the bottom panels. If, for reasons of computational limitations, the quasi- t -ratios are used, then caution is needed in evaluating quasi- t -ratios just below 2.0, and the overall chi-square goodness of test should be used to protect the inference.

6 Conclusion

Our purpose in this paper has to been to explore the use of EMM diagnostics to guide the determination of the specification of a continuous time model for financial data. In the empirical work, we show how the scores of the score generator provide specific guidance on which features of the data the fitted SDE accommodates well and which features it misses. This information can be quite useful in determining the directions along which to expand the model. There is one key caveat. We have shown how to adjust the specification of the SDE in order to accommodate better the score function of a particular score generator, which takes the form of a semiparametric ARCH model. Though this score generator is a reasonable data descriptor, it would not emerge as the best fitting model from a complete and exhaustive search. Such a search is well beyond beyond the scope of this paper, but preliminary evidence (not reported here) indicates that it would detect additional features of the data beyond those captured by semiparametric ARCH. As discussed at length in Tauchen (1995), it is imperative to ultimately confront the SDE specification to the score function of this larger model in order to avoid specification error. That effort would entail using the EMM diagnostics in much the same way as considered in this paper.

7 References

- Aït-Sahalia, Yacine (1995a), “Testing Continuous-Time Models of the Spot Interest Rate,” *The Review of Financial Studies*, forthcoming.
- Aït-Sahalia, Yacine (1995b), “Nonparametric Pricing of Interest Rate Derivatives”, *Econometrica*, forthcoming.
- Broze, Laurence, Olivier Scaillet, and Jean-Michel Zakoian (1995), “Quasi Indirect Inference for Diffusion Processes,” Working paper, C.O.R.E, Universite Catholique De Louvain.
- Chan, K.C., G. A. Karolyi, F. A. Longstaff, and A. B. Sanders (1992), “An Empirical Comparison of Alternative Models of the Short-Term Interest Rate,” *Journal of Finance* 47, 1209–1227.
- Clark, Peter K. (1973), “A Subordinated Stochastic Process Model with Finite Variance for Speculative Prices,” *Econometrica* 41, 135–156.
- Conley, Tim, Lars Peter Hansen, and Jose Scheinkman (1995), “Estimating Subordinated Diffusions From Discrete Time Data,” Working paper, University of Chicago.
- Cox, John C., Jonathan E. Ingersoll, Jr., and Stephen A. Ross (1985), “A Theory of the Term Structure of Interest Rates,” *Econometrica* 53, 385–407.
- Duffie, Darrell, and Rui Kan (1993), “A Yield-Factor Model of Interest Rates,” Working paper, Stanford University.
- Duffie, Darrell, and Kenneth J. Singleton (1993), “Simulated Moments Estimation of Markov Models of Asset Prices,” *Econometrica* 61, 929–952.
- Duffie, Darrell, and Kenneth J. Singleton (1994), “Econometric Modeling of Term Structures of Defaultable Bonds,” Working Paper, Stanford University.
- Engle, Robert F., and Gloria Gonzales-Rivera (1991), “Semiparametric Arch Models,” *Journal of Economic and Business Statistics* 9, 345–360.

- Fenton, Victor M., and A. Ronald Gallant (1996a), “Convergence Rates of SNP Density Estimators,” *Econometrica*, forthcoming.
- Fenton, Victor M., and A. Ronald Gallant (1996b), “Qualitative and Asymptotic Performance of SNP Density Estimators,” *Journal of Econometrics*, forthcoming.
- Gallant, A. Ronald, David A. Hsieh, and George Tauchen (1991), “On Fitting a Recalcitrant Series: The Pound/Dollar Exchange Rate, 1974–83,” in William A. Barnett, James Powell, and George Tauchen (eds.), *Nonparametric and Semiparametric Methods in Econometrics and Statistics, Proceedings of the Fifth International Symposium in Economic Theory and Econometrics*, Cambridge University Press, Cambridge, UK.
- Gallant, A. Ronald, David A. Hsieh, and George Tauchen (1994), “Estimation of Stochastic Volatility Models with Diagnostics,” Working paper, Duke University. Available by anonymous ftp at site ftp.econ.duke.edu in directory pub/arg/papers as file msv.psz.
- Gallant, A. Ronald, and Jonathan R. Long (1995), “Estimating Stochastic Differential Equations Efficiently by Minimum Chi-Square,” Working paper, University of North Carolina. Available by anonymous ftp at site ftp.econ.duke.edu in directory pub/arg/papers as file sde.psz.
- Gallant, A. Ronald, and Douglas W. Nychka (1987), “Seminonparametric Maximum Likelihood Estimation,” *Econometrica* 55, 363–390.
- Gallant, A. Ronald, Peter E. Rossi, and George Tauchen (1992), “Stock Prices and Volume,” *Review of Financial Studies* 5, 199–242.
- Gallant, A. Ronald, Peter E. Rossi, and George Tauchen (1993), “Nonlinear Dynamic Structures,” *Econometrica* 61, 871–907.
- Gallant, A. Ronald, and George Tauchen (1989), “Seminonparametric Estimation of Conditionally Constrained Heterogeneous Processes: Asset Pricing Applications,” *Econometrica* 57, 1091–1120.

- Gallant, A. Ronald, and George Tauchen (1995a), “SNP: A Program for Nonparametric Time Series Analysis, A User’s Guide,” Manuscript, University of North Carolina. The Guide and Fortran code are available by anonymous ftp at site ftp.econ.duke.edu in directory pub/arg/snp or by sending the one-line e-mail message “send snp from general” to statlib@lib.stat.cmu.edu.
- Gallant, A. Ronald, and George Tauchen (1995b). “EMM: A Program for Efficient Method of Moments Estimation, A User’s Guide,” Manuscript, Duke University. The Guide and Fortran code are available by anonymous ftp at site ftp.econ.duke.edu in directory pub/get/emm.
- Gallant, A. Ronald, and George Tauchen (1995c), “Estimation of Continuous Time Models for Stock Returns and Interest Rates,” Working paper, Duke University.
- Gallant, A. Ronald, and George Tauchen (1996), “Which Moments to Match?,” *Econometric Theory*, forthcoming.
- Gourieroux, C., A. Monfort, and E. Renault (1993), “Indirect Inference,” *Journal of Applied Econometrics* 8, S85–S118.
- Hannan, E. J. (1987), “Rational Transfer Function Approximation,” *Statistical Science* 2, 1029–1054.
- Hansen, Lars Peter, and Jose Alexandre Scheinkman (1995), “Back to the Future: Generating Moment Implications for Continuous Time Markov Process,” *Econometrica* 50, 1029–1054.
- Ingram, Beth F., and B. S. Lee (1991), “Simulation Estimation of Time Series Models,” *Journal of Econometrics* 47, 197–250.
- Karatzas, I., and S. E. Schreve (1991), *Brownian Motion and Stochastic Calculus*, Second Edition, Springer-Verlag, Berlin.
- Kloeden, Peter E., and Eckhard Platen (1992), *Numerical Solution of Stochastic Differential Equations*, Springer-Verlag, Berlin.

- Koedijk, Kees G., Francois G. J. A. Nissen, Peter C. Schotman and Christian C. P. Wolff (1995), "The Dynamics of Short-Term Interest Rate Volatility Reconsidered," Working paper, Limburg Institute of Financial Economics.
- Lo, Andrew (1988), "Maximum Likelihood Estimation of Generalized Ito Process with Discretely Sampled Data," *Econometric Theory* 4, 231–247.
- Nelson, D. (1991), "Conditional Heteroskedasticity in Asset Returns: A New Approach," *Econometrica* 59, 347–370.
- Schwarz, G. (1978), "Estimating the Dimension of a Model," *Annals of Statistics* 6, 461–464.
- Tauchen, George (1995), "New Minimum Chi-Square Methods in Empirical Finance," Working paper, Duke University. Available by anonymous ftp at site ftp.econ.duke.edu in directory pub/get/papers as file srvy.ps.Z.
- Tauchen, George, and Mark Pitts (1983), "The Price Variability-Volume Relationship on Speculative Markets," *Econometrica* 51, 485–505.

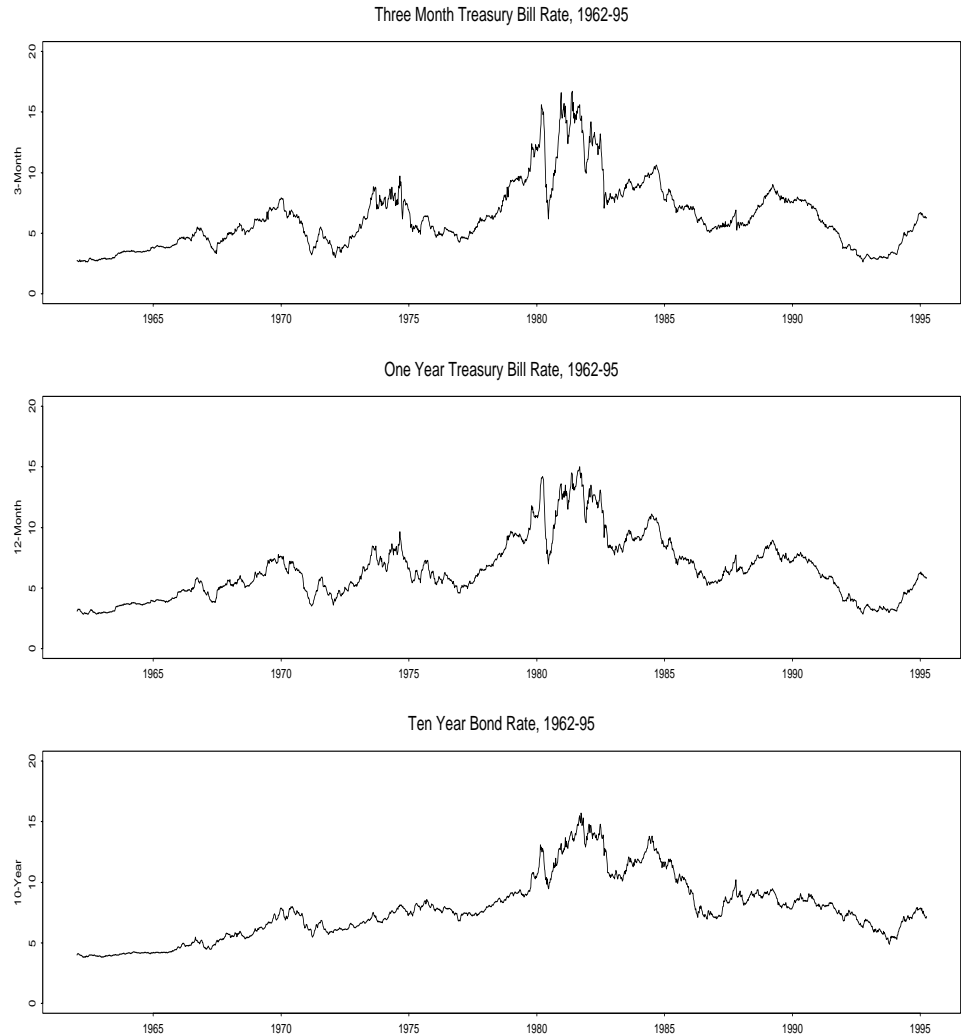
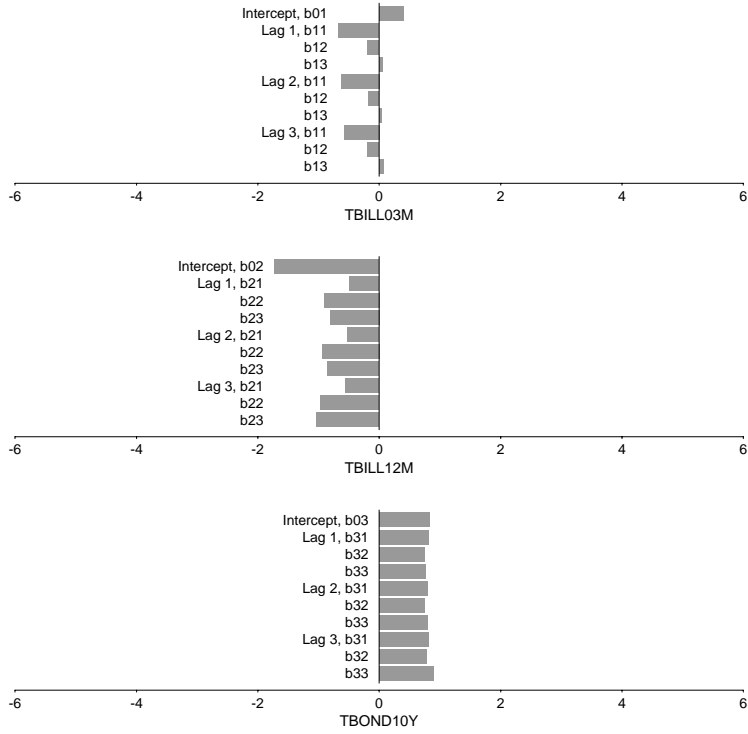


Figure 1. Interest Rate Data. The top panel panel is the Three Month Treasury Bill Rate, the middle is the Twelve Month Treasury Bill Rate, and the bottom is the 10-Year Constant Maturity Treasury Bond Rate, weekly, January 5, 1962–March 31, 1995.

VAR Quasi T-Ratios, $\gamma=0.5$, Semiparametric ARCH Score



VAR T-Ratios, $\gamma=0.5$, Semiparametric ARCH Score

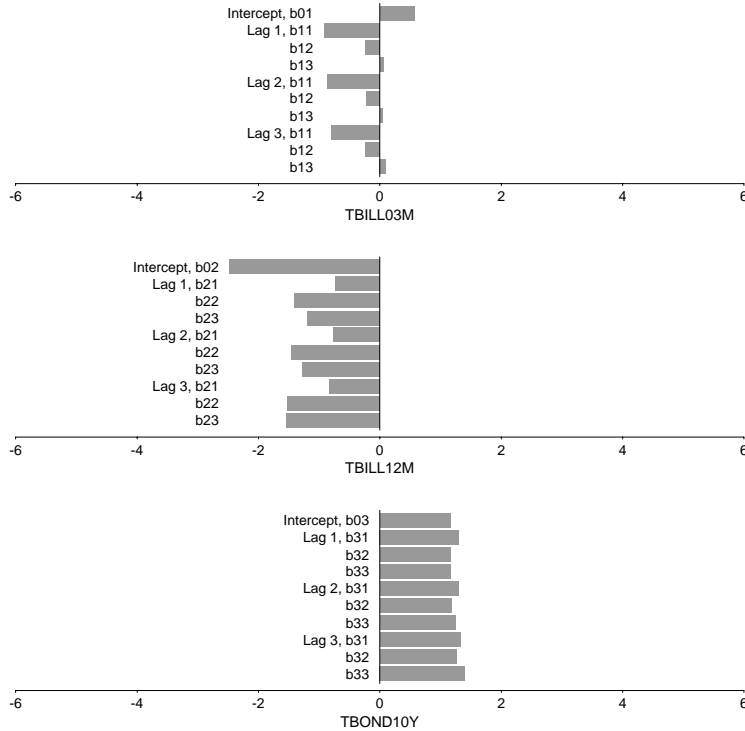


Figure 2. Location Function Diagnostics for Exponent=0.5. t -ratio diagnostics on SNP scores corresponding to the parameters of the location function defined in (2.1) for the estimate of the SDE defined in (5.1)–(5.2) with $\gamma = (0.50 \ 0.50 \ 0.50)$. Top: quasi- t -ratios defined in (3.3). Bottom: adjusted t -ratios defined in (3.2).

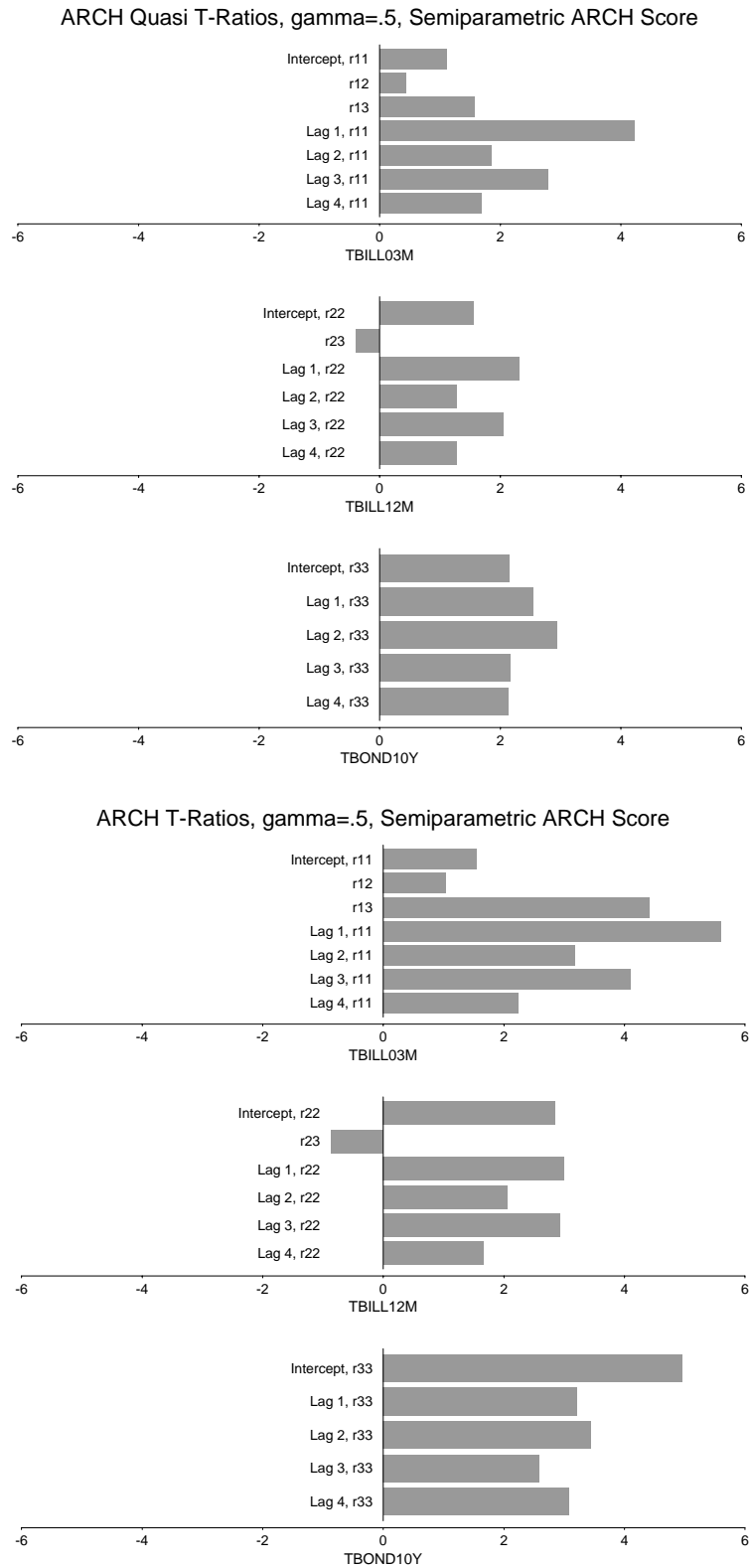
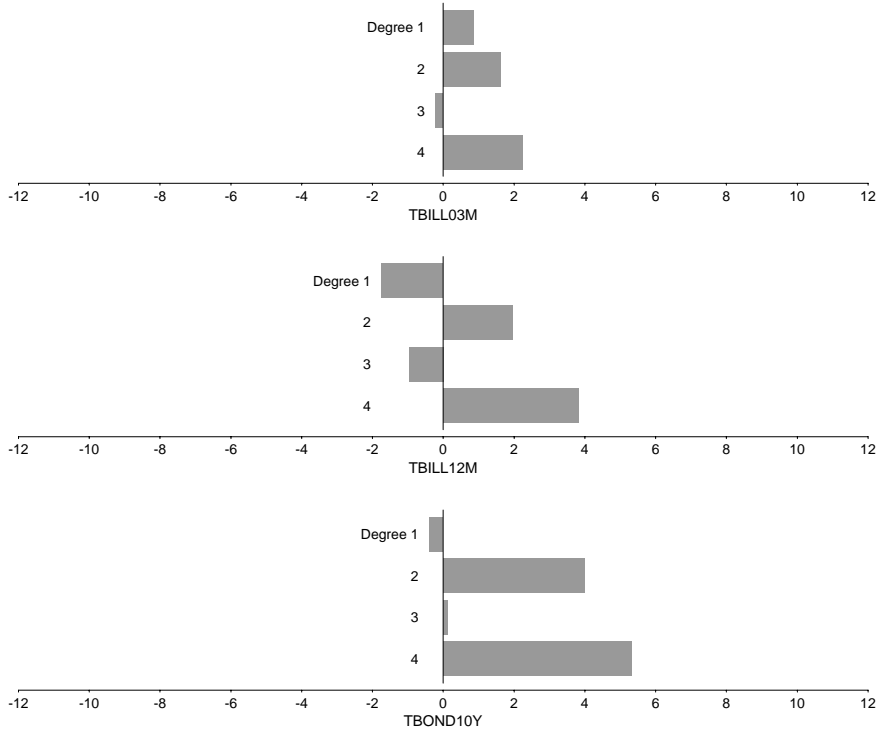


Figure 3. Scale Function Diagnostics for Exponent=.5. t -ratio diagnostics on SNP scores corresponding to the parameters of the scale function defined in (2.2) for the estimate of the SDE defined in (5.1)–(5.2) with $\gamma = (0.50 \ 0.50 \ 0.50)$. Top: quasi- t -ratios defined in (3.3). Bottom: adjusted t -ratios defined in (3.2).

Hermite Quasi T-Ratios, gamma=.5, Semiparametric ARCH Score



Hermite T-Ratios, gamma=.5, Semiparametric ARCH Score

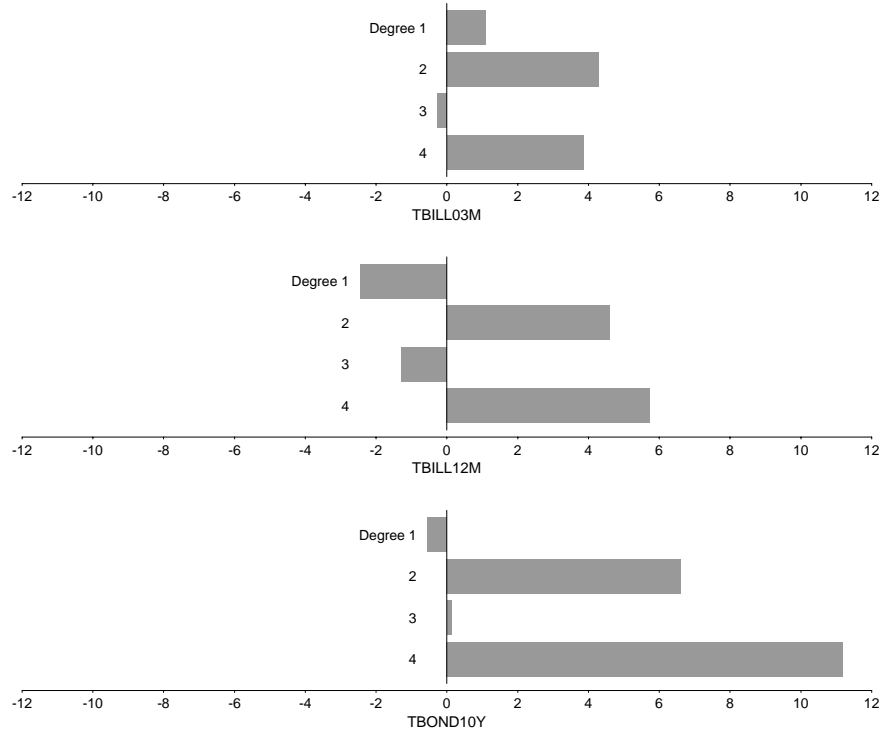
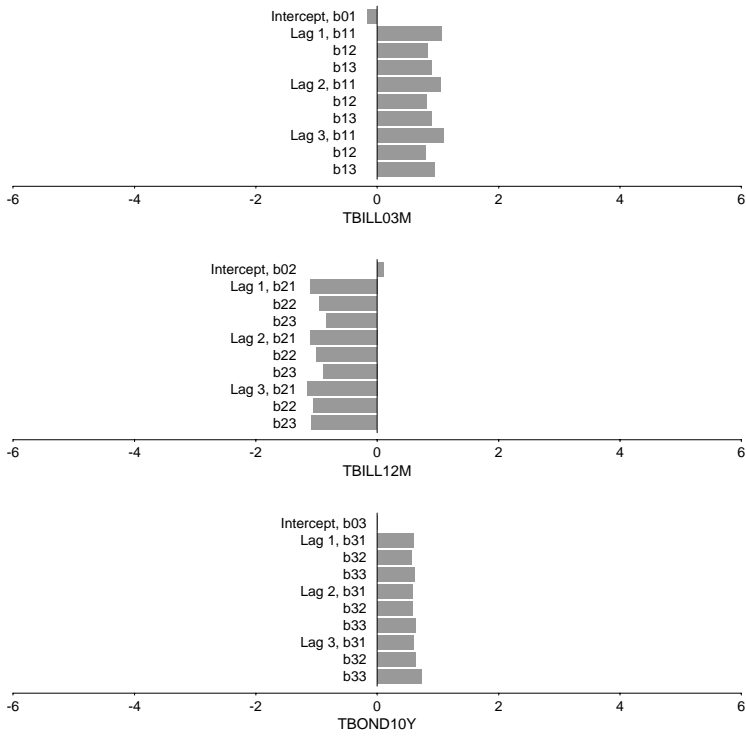


Figure 4. Hermite Polynomial Diagnostics for Exponent=.5. t -ratio diagnostics on SNP scores corresponding to the parameters of the Hermite polynomial defined in (2.3) for the estimate of the SDE defined in (5.1)–(5.2) with $\gamma = (0.50 \ 0.50 \ 0.50)$. Top: quasi- t -ratios defined in (3.3). Bottom: adjusted t -ratios defined in (3.2).

VAR Quasi T-Ratios, $\gamma=0.8$, Semiparametric ARCH Score



VAR T-Ratios, $\gamma=0.8$, Semiparametric ARCH Score

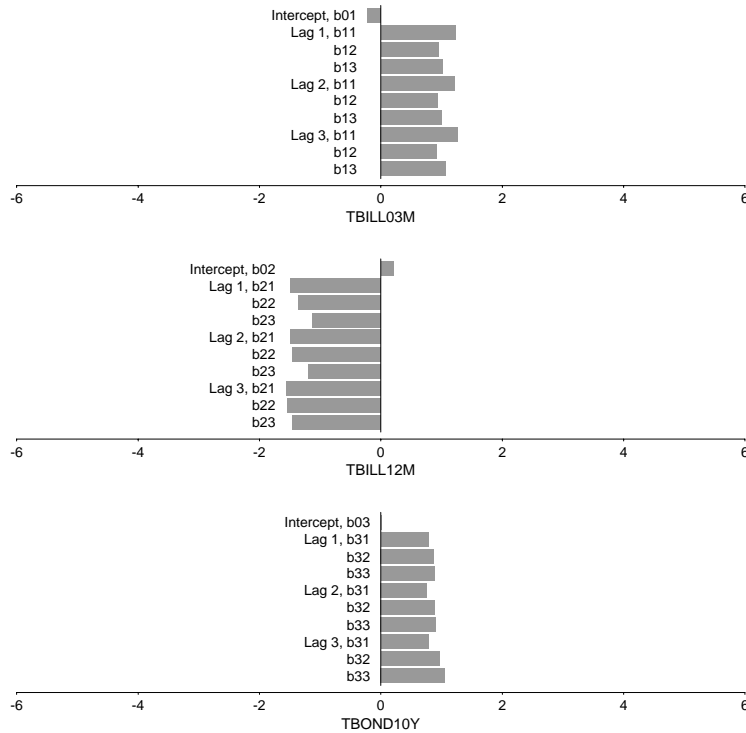


Figure 5. Location Function Diagnostics for Exponent=0.8. t -ratio diagnostics on SNP scores corresponding to the parameters of the location function defined in (2.1) for the estimate of the SDE defined in (5.1)–(5.2) with $\gamma = (0.80 \ 0.80 \ 0.80)$. Top: quasi- t -ratios defined in (3.3). Bottom: adjusted t -ratios defined in (3.2).

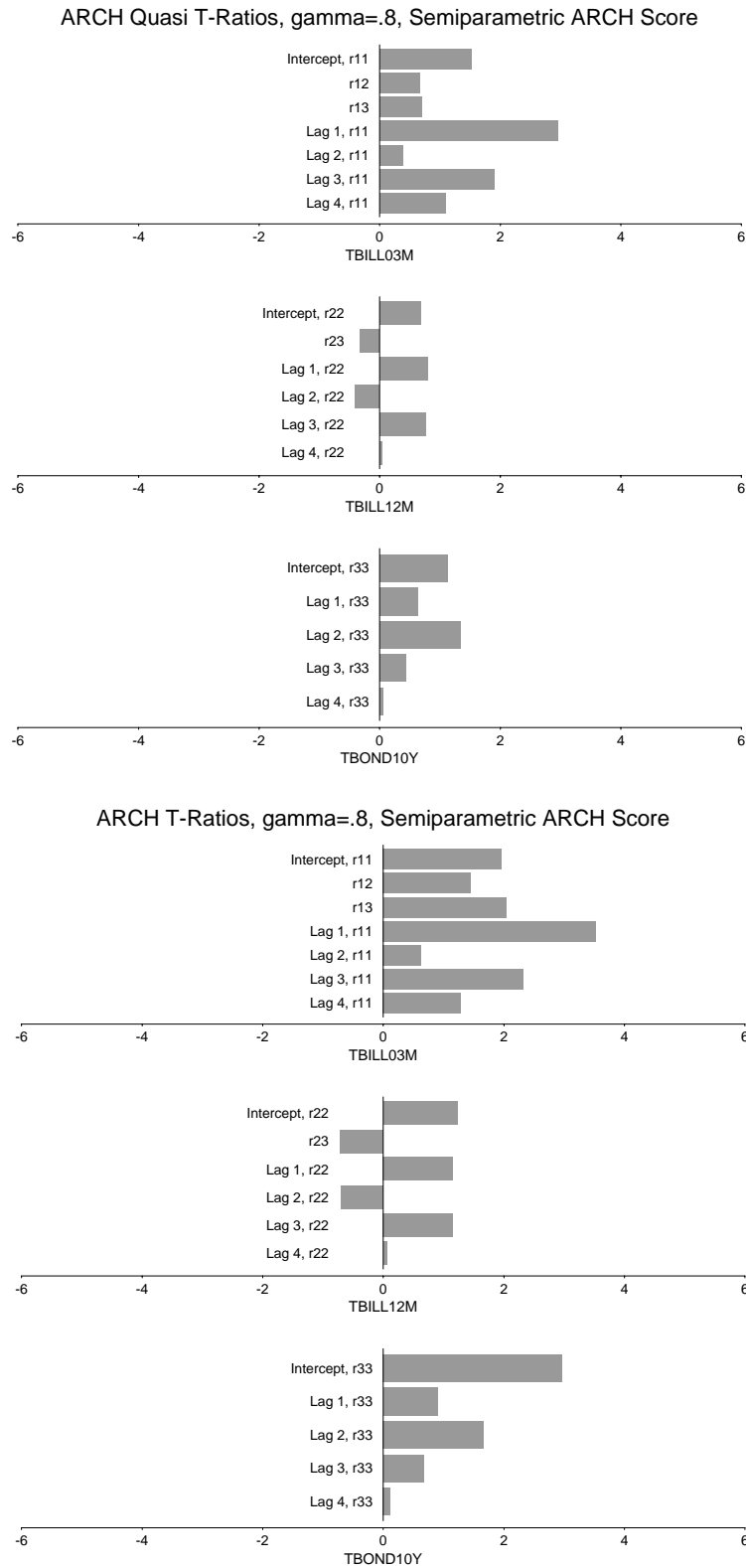
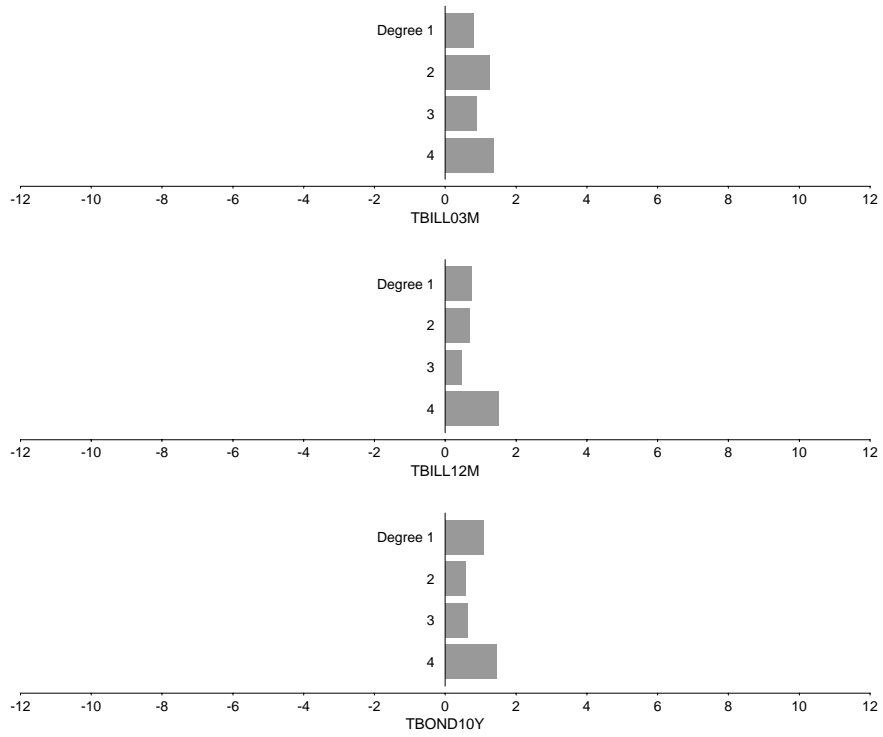


Figure 6. Scale Function Diagnostics for Exponent=.8. t -ratio diagnostics on SNP scores corresponding to the parameters of the scale function defined in (2.2) for the estimate of the SDE defined in (5.1)–(5.2) with $\gamma = (0.80 \ 0.80 \ 0.80)$. Top: quasi- t -ratios defined in (3.3). Bottom: adjusted t -ratios defined in (3.2).

Hermite Quasi T-Ratios, gamma=.8, Semiparametric ARCH Score



Hermite T-Ratios, gamma=.8, Semiparametric ARCH Score

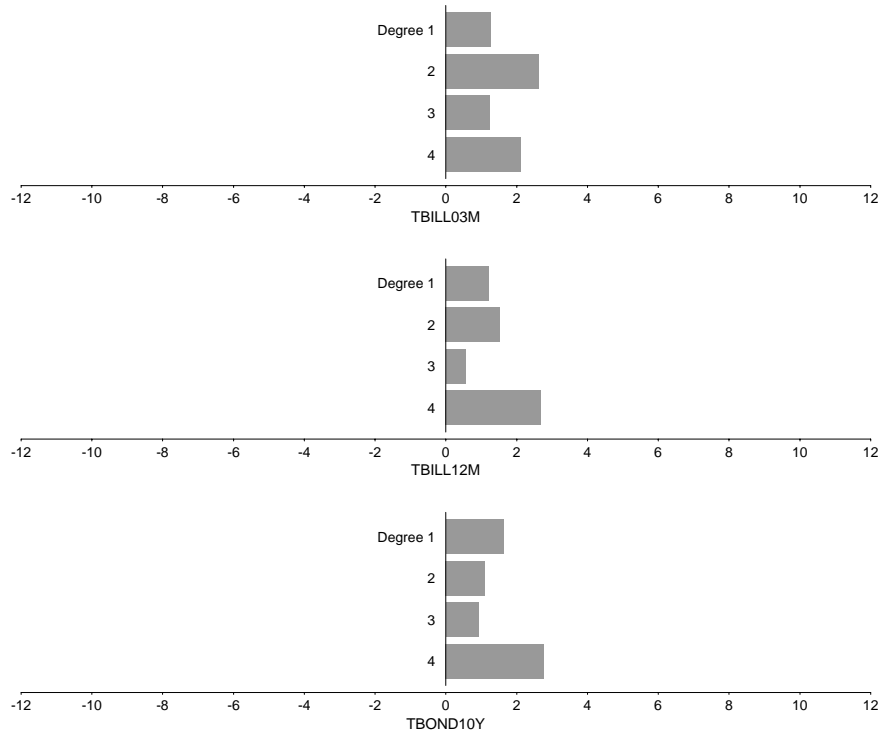


Figure 7. Hermite Polynomial Diagnostics for Exponent=.8. t -ratio diagnostics on SNP scores corresponding to the parameters of the Hermite polynomial defined in (2.3) for the estimate of the SDE defined in (5.1)–(5.2) with $\gamma = (0.80 \ 0.80 \ 0.80)$. Top: quasi- t -ratios defined in (3.3). Bottom: adjusted t -ratios defined in (3.2).