INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

ESSAYS ON MODELS OF THE TERM STRUCTURE OF INTEREST RATES AND ECONOMETRIC METHODS FOR CONTINUOUS TIME STOCHASTIC PROCESSES

by

Hao Zhou

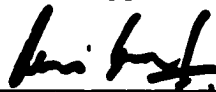Department of Economics
Duke University

Date: _May 1, 2000_
Approved:

_____
Ravi Bansal, Supervisor

_____
George Tauchen, Supervisor

_____
Tim Bollerslev

_____
Mark Coppejans

Dissertation submitted in partial fulfillment of the
requirements for the degree of Doctor of Philosophy
in the Department of Economics
in the Graduate School of
Duke University

2000

UMI Number: 9977694

Copyright 2000 by
Zhou, Hao

All rights reserved.

UMI®

.

ABSTRACT

(Economics—Finance)

ESSAYS ON MODELS OF THE TERM STRUCTURE OF INTEREST RATES AND ECONOMETRIC METHODS FOR CONTINUOUS TIME STOCHASTIC PROCESSES

by

Hao Zhou

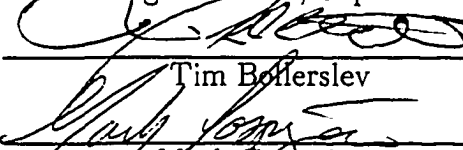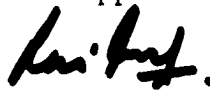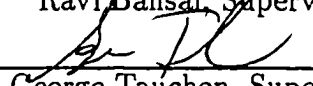Department of Economics
Duke University

Date: *May 1, 2000*

Approved:

_____
Ravi Bansal, Supervisor

_____
George Tauchen, Supervisor

_____
Tim Bollerslev

_____
Mark Coppejans

An abstract of a dissertation submitted in partial
fulfillment of the requirements for the degree
of Doctor of Philosophy in the Department of
Economics in the Graduate School of
Duke University

2000

Abstract

My dissertation is written in the fields of finance and econometrics. The research agenda starts with a Monte Carlo study, one which reveals structural limitations of the standard square-root model in fitting the term structure of interest rates and the finite sample inefficiency of simulation-based estimators in empirical finance. The discovery of these limitations naturally leads to the two major contributions of my thesis research. One is to develop new term structure models—ones that have discrete jumps or regime shifts—that are in line with the general equilibrium no-arbitrage pricing approach. The other significant contribution is to construct highly efficient econometric methods to estimate jump-diffusion and stochastic volatility processes, by providing closed form solutions to the conditional moments of the underlying point-in-time process or the integrated time series. The major findings of the research are these: the jump-augmented or regime-augmented yield curves are much more flexible in fitting the observed term structure, and the conditional-moment-based econometric estimators are statistically reliable and computationally efficient. What is gained from these new modeling and estimation strategies is that one may adequately capture the rich volatility pattern embedded in most financial time series data.

Acknowledgements

George Tauchen first brought me into the area of financial econometrics, and he has given me valuable advice and continuous support. My knowledge in financial economics, especially asset pricing, was developed with the enormous help of Ravi Bansal. Tim Bollerslev joined my committee two years ago and has challenged me to work in ground-breaking areas. During my dissertation research stage, Mark Coppejans, Pete Kyle, Ron Gallant, and Chien-Te Hsu were very generous in spending their time to help me with specific issues of their expertise. Since I came to Duke University, Dennis Yang, Kent Kimbrough, Mark Ann, and Greg Lawler also extended their assistance in a number of ways. The participants in the Duke financial econometrics lunch group have offered many valuable comments and discussions, and my experiences with them will have a long-lasting impact. Nine years ago, I was first involved in economic research with Justin Lin, a kind of career I have never been able to leave behind.

v

Contents

List of Tables

List of Figures

xiii

Chapter 1

Introduction

My dissertation consists of four papers written in the fields of finance and econometrics. These essays develop new term structure models with discrete jumps or regime shifts and highly efficient econometric methods to estimate jump-diffusion and stochastic volatility processes.

The following chapter performs a Monte Carlo study on Efficient Method of Moments (EMM) for a continuous time square-root model and then compares the results with maximum-likelihood estimators. A convenient Poisson-mixing-Gamma formula is implemented for Maximum Likelihood Estimation (MLE), and the exact solutions of the first two moments are used for Quasi-Maximum Likelihood Estimation (QMLE). The finite sample efficiency of EMM increases with the sample size, as the seminonparametric score generator (SNP) captures more of the distributional features. The overrejection bias of EMM decreases with the sample size, and the power of detecting misspecification is ultimately one. We also find that the standard square-root model is too restrictive in fitting the term structure volatility, and this finding suggests the jump-diffusion and regime switching methodologies subsequently being used in Chapters 3 and 4. Similarly, the slow convergence rate of the simulation-based estimator motivates the conditional moment-based estimation strategies developed in Chapters 3 and 5.

The third chapter develops a Multivariate Weighted Nonlinear Least Square estimator for a jump-diffusion interest rate model (MWNLS-JD), which also admits a closed-form solution for bond prices under an equilibrium no-arbitrage argument. The instantaneous interest rate is modeled as a mixture of a continuous square-root

1

diffusion and a discrete Poisson jump process, where the jump rate or amplitude can be constants, state-dependent functions, or independent random variables. Using Itô's formula, one can analytically derive the first four conditional moments, which form the basis of the MWNLS-JD estimator. The availability of closed-form solutions reduces the computational time to only a few minutes. A diagnostic conditional moment test and a classical Lagrange multiplier test can also be constructed from the fitted moment conditions. The time series estimation of the short rate suggests that the jump augmentation is statistically very significant and that the pure diffusion process is strongly rejected. The cross-sectional evidence indicates that the jump-diffusion yield curves are much more flexible and can significantly reduce the pricing errors.

In Chapter 4, I explore the possibility that changes in economic regimes have important effects on the term structure of interest rates. An equilibrium model of the term structure is developed with both the underlying short-term interest rate and the market price of risk being determined by latent regime-switching square-root processes. I provide analytical solutions for bond prices of different maturity and then estimate the model via Efficient Method of Moments (EMM), using short and long-term interest rate data from 1964-1995. Our empirical results show that a two-factor regime-switching model passes the EMM specification test, while standard term structure models (with up to three factors) are strongly rejected. The reprojection technique from EMM also indicates that only the preferred regime-switching model can mimic the observed conditional volatility and the correlation of the two interest rates. Further, we find that the key distinctions across regimes are the volatility of the short yield and the spread between the long and the short yields.

In Chapter 5, I extend the conditional moment estimator developed in Chapter 2 to the case of stochastic volatility diffusions. The analytical solutions of the first

2

two conditional moments are derived for the integrated volatility process, which is approximated by the quadratic variation constructed from the high frequency returns. This in turn allows for the construction of a standard GMM estimator in which we use both the primary and lag-augmented moment conditions. We successfully implement this method to a variety of ultra-high frequency data, including foreign exchange rates and stock market index. Our empirical results and the simulation evidence both indicate that this new method of estimating stochastic volatility diffusions is highly reliable and accurate. Furthermore, its computational speed is much faster when compared with the computational speeds of the other available methods, such as EMM and MCMC. One important extension of the method is to directly evaluate the option prices by exploiting the distributional features of the integrated volatility.

3

Chapter 2

A Monte Carlo Studay for A Square-Root Diffusion Process

[1]This chapter performs a Monte Carlo study on Efficient Method of Moments (EMM) for a continuous time square-root model and compares the result with maximum likelihood estimators. A convenient *Poisson*-mixing-*Gamma* formula is implemented for Maximum Likelihood Estimation (MLE), and the exact first two moments are used for Quasi-Maximum Likelihood Estimation (QMLE). The relative efficiency of EMM over MLE is increasing with the sample size, as the seminonparametric score generator (SNP) capturing more of the distribution feature. The contribution of this essay is to provide small sample evidence that the overrejection bias of EMM is decreasing with the sample size, and the power for detecting misspecification is ultimately one.

2.1 Introduction

When estimating a continuous time model in finance, one often faces the difficulty of partial observability. Usually the continuous time record is not available since the data is discretely sampled. A further complication is that the transitional density of the stochastic process does not always have a closed-form solution. Due to the lack of a tractable likelihood function, much of the interest has turned to nonlikelihood-based approaches. The Generalized Method of Moments (GMM) by Hansen (1982) reduces the reliance on distribution assumptions by matching the empirical moments with the theoretical ones. The Simulated Method of Moments (GMM) in time series

---

[1]The material of this chapter also appears in Zhou (1999a).

4

application (Ingram and Lee 1991, Duffie and Singleton 1993) minimizes the reliance on distribution assumptions by matching the empirical moments with the simulated ones. Both GMM and SMM are robust to the misspecification of likelihood functions, while retaining a parametric model to conduct simulation or projection. However, these methods of moments suffer from the *ad hoc* choice of moment conditions and must presume the existence of arbitrary population moments. The chi-square specification test of the overidentifying restrictions is subject to severe overrejection bias (Hansen, Heaton, and Yaron 1996, Andersen and Sørenson 1996). The efficiency loss of parameter estimates is closely related to the high cost in estimating the weighting matrix, as the variance-covariance matrix is typically heteroskedastic and serially correlated (Andersen and Sørenson 1996). The Wald test is also found to exceed its asymptotic size due to the difficulty in estimating the residue spectral-density matrix (Burnside and Eichenbaum 1996).

The Efficient Method of Moments (EMM), introduced by Bansal, Gallant, Hussey, and Tauchen (1995) and Gallant and Tauchen (1996b), endogenously selects the moment conditions in the first step. A seminonparametric score generator (SNP) uses the Fourier-Hermite polynomial to approximate the underlying transitional density. As an orthogonal series estimator, the SNP has a fast uniform convergence, given the smoothness of the underlying distribution function. A suitable model selection criterion, e.g., the Schwartz's Bayesian Information Criterion (BIC), is used to choose the direction and complexity of the auxiliary model expansion. The second stage of EMM is simply an SMM type estimator, minimizing the quasi-maximum likelihood score functions that are chosen appropriately in the first stage. Since the moment conditions are orthogonal, the weighting matrix (i.e., the information matrix from the quasi-maximum likelihood) should be nearly serially uncorrelated. Hence the asymptotic variance estimator approaches the minimum bound, and the parameter

5

estimates are asymptotically as efficient as MLE. It is proven that for an ergodic stochastic system with partially observed data, the efficiency of EMM approaches that of MLE, as the number of moment conditions and the number of lags entering each moment increase with the sample size (Gallant and Long 1997). Another salient feature of EMM is the capability to detect a misspecified structural model, if the auxiliary model is rich enough such that Hermite polynomial scores approximate the true scores fairly well (Tauchen 1996). Under correct specification of the maintained model, the normalized objective function value converges in distribution to a chi-square distribution. Under misspecification, the unnormalized objection function converges almost surely to a constant. For particular choice of score generator, this constant may be zero and the chi-square test loses power against the alternative. If the data generating process is adequately captured by a more flexible nonparametric score generator, the constant is positive and rejection of misspecification is almost certain.

Recent Monte Carlo studies of EMM documented significant efficiency gains of EMM over GMM (Andersen, Chung, and Sørenson 1999c), but with similar overrejection problems in specification tests (Chumacero 1997). In more analytical fashion, Gallant and Tauchen (1998a) show that EMM outperforms the conventional method of moments (CMM) for a representative class of econometric models. However, there is no universal theory regarding the efficiency of EMM versus that of CMM, and the comparison must be made case-by-case (Gallant and Tauchen 1998a). Therefore choosing EMM over QMLE (e.g., in Dai and Singleton (2000)) should be accompanied by solid argument or Monte Carlo evidence. This chapter complements these comparative studies in several areas. First, the Monte Carlo setup is a continuous time model, like many recent applications of EMM, which have focused on the stochastic differential equations. Second, we consider the relative efficiency of EMM

6

with respect to asymptotically efficient MLE and computationally efficient QMLE. Third, both blindfold and educated choice of moment conditions are examined, so as to best imitate the realistic approaches taken by researchers using EMM. The main contribution of this essay is to provide Monte Carlo evidence which shows that the overrejection bias converges to zero and that the probability of detecting misspecification converges to one.

A square-root diffusion process (Cox, Ingersoll, and Ross 1985a) is chosen as the vehicle for conducting the Monte Carlo study. On the one hand, the CIR model is simple enough to give closed-form solutions for both the transitional density and the asset pricing formula. On the other hand, it is rich enough to generate a highly persistent volatility and non-Gaussian error distribution. The square-root process seems to be a good starting point to model more complicated financial time series data. EMM estimation of the interest rate diffusions is reported by Gallant and Tauchen (1998b), and the square-root model is firmly rejected. With a closed-form transitional density, the dynamic maximum likelihood estimation was implemented for the two-factor CIR model (Pearson and Sun 1994, Duffie and Singleton 1997). Gibbons and Ramaswamy (1993) employed a GMM estimator, using the stochastic Euler equations to generate the moment conditions. Their results favor the square-root model. The most recent interest in affine term structure (Dai and Singleton 2000) can be viewed as an immediate extension of the multifactor square-root model. The CIR model also has an explicit marginal density in terms of the drift and volatility functions, which motivated nonparametric specification test (Aït-Sahalia 1996b). Conley, Hansen, Luttmer, and Scheinkman (1997a) implemented a GMM estimator for a subset of the parameters by exploiting the reversibility of stationary Markov chain. When the square-root model is rejected, a critical issue is to understand whether the rejection comes from the intrinsic modeling inadequacy or from the overrejection bias

7

in finite samples.

The remaining sections are organized in the following way: Section 2.2 discusses some properties of the square-root model and characterizes the likelihood function for MLE; Section 2.3 introduces the Efficient Method of Moments estimator; Section 2.4 designs the Monte Carlo experiment and suggests the benchmark model choice; Section 2.5 reports the major findings; and Section 2.6 concludes.

2.2   Maximum Likelihood Estimation

This section defines a maximum likelihood estimator for the square-root model, based on a *Poisson*-mixing-*Gamma* characterization of the likelihood function. A quasi-maximum likelihood estimator is also available with analytical solutions to the first two conditional moments.

2.2.1   Probabilistic Solution to Square-Root Model

It is a well-known result that the square-root model,

$$dr_t = (a_0 + a_1 r_t)dt + b_0 r_t^{1/2} dW_t, \tag{2.1}$$

satisfies the regularity conditions for both a strong solution (pathwise convergent) and a weak solution (convergent in probability) (Karatzas and Shreve 1997). Obviously a strong solution implies a weak one, but not vice versa. If (1) $a_0 > 0$, (2) $b_0 > 0$, (3) $a_1 < 0$ and (4) $b_0^2 \leq 2a_0$, then the square-root model has a unique fundamental solution (Feller 1951). The marginal density is a *Gamma* distribution, and the transitional density is a type I Bessel function distribution or a noncentral chi-square distribution with a fractional order (Cox et al. 1985a). Intuitively condition (3) gives mean reversion, and condition (4) ensures the stationarity.

8

The marginal *Gamma* distribution is

$$f(r_0|\nu,\omega) = \frac{\omega^\nu}{\Gamma(\nu)} r_0^{\nu-1} e^{-\omega r_0}, \tag{2.2}$$

where $\omega = -2a_1/b_0^2$, $\nu = 2a_0/b_0^2$, and $\Gamma(\cdot)$ is the *Gamma* function. The unconditional mean and variance are

$$E(r_0) = \frac{-a_0}{a_1} = \frac{\nu}{\omega}, \tag{2.3}$$

$$V(r_0) = \frac{b_0^2 a_0}{2a_1^2} = \frac{\nu}{\omega^2}. \tag{2.4}$$

Notice that the first two moments merely identify the marginal distribution. Higher order moments are simply nonlinear functions of the first two moments. The marginal density alone can not identify all three parameters in the diffusion process. Any GMM type estimator must add at least one lagged instrumental variable (Gibbons and Ramaswamy 1993). A rejection of the marginal distribution can reject the square-root model; however, a non-rejection does not provide enough information for judging a particular parameter setting (Aït-Sahalia 1996b). Transitional information must be exploited to fully identify the dynamic structure.

The conditional density is

$$f(r_1|r_0; a_0, a_1, b_0) = c e^{-u-v} \left(\frac{v}{u}\right)^{\frac{q}{2}} I_q(2(uv)^{\frac{1}{2}}), \tag{2.5}$$

$$q = \frac{2a_0}{b_0^2} - 1, \tag{2.6}$$

where $c = -2a_1/(b_0^2(1 - e^{a_1}))$, $u = cr_0 e^{a_1}$, and $v = cr_1$. $I_q(\cdot)$ is a modified Bessel function of the first kind with a fractional order q (Oliver 1972). The conditional mean and variance are

$$E(r_1|r_0) = r_0 e^{a_1} - \frac{a_0}{a_1}(1 - e^{a_1}), \tag{2.7}$$

9

$$V(r_1|r_0) \quad = \quad r_0(-\frac{b_0^2}{a_1})e^{a_1}(1 - e^{a_1}) + \frac{b_0^2 a_0}{2a_1^2}(1 - e^{a_1})^2. \tag{2.8}$$

By the stationary property, the limit of the transitional density as the time interval goes to zero, is exactly the marginal density. Therefore any estimation strategy or specification test exploiting the transitional density will naturally nest those relying on marginal density.

It is a common practice in the literature to call this distribution a "noncentral chi-square distribution". However, the "integer order noncentral chi-square distribution" does not naturally extend to the "fractional order noncentral chi-square distribution". The latter arises commonly from the solution to a diffusion process (Feller 1971), while the former arises from the sample standard deviation of independent, nonidentical, noncentered, normal random variables (Johnson and Kotz 1970).

2.2.2 Maximum Likelihood Estimations

In industry and academics alike, one popular method in estimating the square-root model for interest rates is the Discretized Maximum Likelihood Estimation (DMLE), i.e., a misspecified QMLE based on the time discretization of the conditional mean and variance

$$E(r_1|r_0) \quad \approx \quad a_0 + (1 + a_1)r_0, \tag{2.9}$$

$$V(r_1|r_0) \quad \approx \quad b_0^2 r_0. \tag{2.10}$$

As pointed out by Lo (1988), DMLE is generally not consistent. The parameter estimates are asymptotically biased, since both moments are misspecified.

The exact expressions for the conditional mean and variance (equations 2.7 and 2.8) suggest a quasi-maximum likelihood estimator for the square-root model. QMLE is shown to be root-n consistent (Bollerslev and Wooldridge 1992), although its efficiency is certainly less than that of MLE. The computational efficiency of QMLE

10

could be very attractive, because the analytical gradients are available in many cases. Monte Carlo evidence on the relative efficiency of QMLE would be very useful in situations when EMM (asymptotically efficient) and QMLE (computationally efficient) are available but MLE is not. An empirical example is the multifactor affine model for term structure, which has tractable forms of conditional mean and variance (Fisher and Gilles 1996) but usually is estimated by EMM (Dai and Singleton 2000). When sample size is large, EMM will certainly outperform QMLE due to the asymptotic efficiency argument, because the endogenous score generator will naturally pick more moment conditions and more of the lags entering these moments. However, in small econometric samples, QMLE as a special case of conventional method of moments could outperform EMM, since no theoretical efficiency argument is universally true (Gallant and Tauchen 1998a). A computationally efficient approach for estimating affine term structure is very useful in asset pricing practice.

When implementing MLE for the square-root model, the Bessel function representation of the likelihood function is not at all a convenient form (Pearson and Sun 1994, Duffie and Singleton 1997). An alternative *Poisson*-mixing-*Gamma* characterization can be inferred from the simulation strategy suggested by Devroye (1986). Within the admissible parameter region, one can substitute the Bessel function with an infinite series expansion (Oliver 1972). With appropriate transformations ($y = v$, $\lambda = q + 1$, and $\beta = \sqrt{2u}$), the alternative mixing formula falls out nicely,

$$
\begin{aligned}
f(y) &= \sum_{j=0}^{\infty} \frac{y^{j+\lambda-1}e^{-y}}{\Gamma(j+\lambda)} \cdot \frac{\left(\frac{\beta^2}{2}\right)^j e^{-\frac{\beta^2}{2}}}{j!} \\
&= \sum_{j=0}^{\infty} \text{Gamma}(y|j+\lambda, 1) \cdot \text{Poisson}(j|\frac{\beta^2}{2}).
\end{aligned}
\tag{2.11}
$$

One needs to be cautious that the *Poisson* weights are not constant, but conditioning on the previous realization r_0. This formula corresponds to the "*Poisson* driven

11

Gamma process" in Feller (1971). The only difference is that $\lambda = q + 1$ remains a fractional number, not an integer. The evaluation of the log-likelihood function in MLE is greatly simplified, when using the *Poisson*-mixing-*Gamma* formula. It is fairly easy to achieve the single precision 10^{-8} by truncating the *Poisson* distribution around 100. One can also avoid any complication of complex value or non-convergence in evaluating the Bessel function.

2.2.3 Simulation and Inferences

The above characterization also defines a composite simulation method for the square-root process (Devroye 1986). First, one draws a random number j from the distribution $Poisson(j|\beta^2/2)$. Then, one draws another random number y from the distribution $Gamma(y|j + \lambda, 1)$. Finally, one calculates the desired state variable r_1 by $r_1 = y/c$. Notice the realized r_1 is the conditioning value r_0 in the next draw. The initial value r_0, when starting a simulation run, can be set to the theoretical unconditional mean. To pass on the transient effect, the first 1000 realizations can be discarded.

Since we know the true parameter value, the likelihood ratio tests can determine how often the confidence region centered at the estimated parameter value contains the truth. The same principle may be extended to QMLE if the expected quasi-likelihood function is uniquely maximized at the truth (identification requirement). In addition to the correct specification of the mean and variance, the remaining innovation error must be centered at zero or have a symmetric distribution (Newey and Steigerwald 1997). QMLE for the square-root model meets these conditions.

12

2.3 Efficient Method of Moments Estimation

This section describes the EMM estimator and emphasizes its asymptotic efficiency. For formal discussions see Gallant and Tauchen (1996b), Tauchen (1996), and Gallant and Long (1997).

2.3.1 Approximating True Density with Auxiliary Model

Denote the invariant probability measure implied by the underlying data generating process as the p-model. It is assumed that the direct maximum likelihood estimation of the p-model is not available. However, any smooth density function can be approximated arbitrarily close by a Hermite polynomial expansion.

Consider a scalar case. Let y be the random variable, x be the lagged y, and θ be the parameter. The auxiliary f-model has a density function defined by a modified Hermite polynomial,

$$f(y|x,\theta) = C\{[\mathcal{P}(z,x)]^2 \phi(y|\mu_x, \sigma_x^2)\}, \tag{2.12}$$

where $\mathcal{P}$ is a polynomial with degree K_z in z and the square of $\mathcal{P}$ makes the density positive. The argument of the polynomial is z, which is the transformation $z = (y - \mu_x)/\sigma_x$. The coefficient of the polynomial is another polynomial of degree K_x in x. The constant in the polynomial is set to 1 for identification. C is a normalizing factor to make the density proper. $\phi(\cdot)$ is a normal density of y with conditional mean μ_x and conditional variance σ_x^2. The length of the auxiliary model parameter is determined by the lag in mean $L\mu$, lag in variance L_r, lag in polynomial coefficient L_p, polynomial degree K_z, and polynomial degree K_x. Let $\{\tilde{y}_t\}_{t=1}^n$ be the observed data and $\tilde{x}_{t-1}$ be the lagged observations. The sample mean log-likelihood function is defined by

$$\mathcal{L}_n(\theta, \{\tilde{y}_t\}_{t=1}^n) = \frac{1}{n}\sum_{t=1}^n \log[f(\tilde{y}_t|\tilde{x}_{t-1}, \theta)]. \tag{2.13}$$

13

A quasi-maximum-likelihood estimator is obtained by

$$\tilde{\theta} = \arg\max_{\theta} \mathcal{L}_n(\theta, \{\tilde{y}_t\}_{t=1}^n). \tag{2.14}$$

The dimension of the auxiliary f-model, the length of θ, is selected by Schwarz's Bayesian Information Criterion (BIC). There are different choices of the information criteria for optimal model selection. For a finite dimensional stationary process, BIC with a bigger penalty for model complexity proves to be consistent, while the Akaike's Information Criterion (AIC) will overfit the model. On the other hand, if the true dimension is infinity or increases to infinity with the sample size, AIC with a smaller penalty for model complexity is optimal (Zheng and Loh 1995). The dimensions of the f-model needs to be as large as the p-model to meet the identification condition.

2.3.2 Matching Auxiliary Scores with Minimum Chi-Square

From the first-stage seminonparametric estimates, one obtains the fitted scores as the moment conditions,

$$m_n(\tilde{\theta}) = \frac{1}{n}\sum_{t=1}^n \frac{\partial}{\partial\theta} \log f(\tilde{y}_t|\tilde{x}_{t-1}, \tilde{\theta}). \tag{2.15}$$

In the second stage, a SMM-type estimator is implemented in the following way. Although the direct MLE for p-model is assumed impossible, the simulation from the structural model (e.g., stochastic differential equation) is readily available. Let $\{\hat{y}_t\}_{t=1}^N$ be a long simulation from a candidate value of ρ, the parameter of the maintained structural model. The auxiliary score functions can be reevaluated at the simulated data,

$$\hat{m}_N(\rho, \tilde{\theta}) = \frac{1}{N}\sum_{t=1}^N \frac{\partial}{\partial\theta} \log f(\hat{y}_t|\hat{x}_{t-1}, \tilde{\theta}), \tag{2.16}$$

and the minimum chi-square estimator is simply,

$$\hat{\rho} = \arg\min_{\rho}\{\hat{m}_N(\rho, \tilde{\theta})'\hat{\mathcal{I}}^{-1}\hat{m}_N(\rho, \tilde{\theta})\}, \tag{2.17}$$

14

where the weighting matrix $\tilde{\mathcal{I}}^{-1}$ is estimated by the mean-outer-product of scores from the auxiliary model

$$\tilde{\mathcal{I}} = \frac{1}{n}\sum_{t=1}^{n}[\frac{\partial}{\partial\theta}\log f(\tilde{y}_t|\tilde{x}_{t-1},\tilde{\theta})][\frac{\partial}{\partial\theta}\log f(\tilde{y}_t|\tilde{x}_{t-1},\tilde{\theta})]'. \tag{2.18}$$

Remember that the efficiency loss of the GMM-type estimators is largely attributed to the high cost and low accuracy in estimating the serially correlated weighting matrix. In EMM the moment conditions are chosen orthogonally by the Hermite polynomial expansion; hence the information matrix is diagonal or nearly diagonal (i.e., serially uncorrelated). The efficiency argument of EMM relies on three conditions: the dimension of the auxiliary model is sufficiently large ($K \rightarrow \infty$), the lag in the the auxiliary model is sufficiently long ($L \rightarrow \infty$), and the simulation from the maintained structural model is sufficiently long ($N \rightarrow \infty$). This encompasses both Markovian and non-Markovian cases (Gallant and Long 1997).

2.3.3 Overrejection and Misspecification

The normalized criterion function value in the EMM estimation

$$c_n^2 = n\hat{m}_N(\hat{\rho},\tilde{\theta})'\tilde{\mathcal{I}}^{-1}\hat{m}_N(\hat{\rho},\tilde{\theta}) \tag{2.19}$$

forms a specification test for the overidentifying restrictions. Under the correct specification of the maintained model (Tauchen 1996), we have $c_n^2 \xrightarrow{D} \mathcal{X}^2(l_\theta - l_p)$, where the degree of freedom equals the parameter length of the auxiliary model minus that of the structural model. However, if the maintained model is misspecified (Tauchen 1996), we have

$$\frac{c_n^2}{n} \xrightarrow{a.s.} \bar{c}^2 = \hat{m}_N(\bar{\rho},\bar{\theta})'\bar{\mathcal{I}}^{-1}\hat{m}_N(\bar{\rho},\bar{\theta}) > 0,$$

where $\bar{\theta}$, $\bar{\rho}$, and $\bar{\mathcal{I}}$ are the asymptotic pseudo-true values under the maintained misspecification. As long as the sample size n is large enough, the SNP score generator

15

will be rich enough such that the false passage $\bar{c}^2 = 0$ will not occur under misspecification.

Considering the fact that the EMM estimation has overrejection bias converging to zero as the sample size increases (Andersen et al. 1999c, Chumacero 1997), it might be convenient to synthesize the above two situations in the following manner:

$$c_n^2 \xrightarrow{\mathcal{D}} \mathcal{X}^2(l_\theta - l_p) + nd_n^2,$$

where d_n^2 is either the overrejection bias or the noncentrality parameter. When the model is correctly specified, $d_n = o(1/n^{1/2+\epsilon})$, i.e., the normalized bias nd_n^2 is dissipating. For the misspecified model, $d_n = o(1)$, i.e., the normalized noncentrality nd_n^2 is exploding.

2.4 Monte Carlo Design and Benchmark Choice

2.4.1 Experimental Design

All computations are performed on one SunOS5.5 Solaris/Sparc and two SunOS 5.6 Solaris/X86 servers. Programs for generating random samples, QMLE, and MLE of the CIR model are written in FORTRAN language. The FORTAN codes for SNP and EMM are modified from SNP Version 8.5 (Gallant and Tauchen 1997) and EMM Version 1.3 (Gallant and Tauchen 1996a), incorporating automatic SNP search by BIC in half of the EMM simulation runs. NPSOL (Gill, Murray, Saunders, and Wright 1991) is the optimization routine used for all of the programs. The contrasting sample sizes are chosen to be 500 and 1500. The pseudo-random sample of the CIR model is simulated from the exact distribution function, i.e., the mixing formula in Section 2.2. The EMM estimator uses a discretized weak-order 2 scheme to approximate the ergodic stochastic systems. When generating the pseudo-random samples, 1000 initial stretch is discarded to pass on the transient effect. A total

16

of 1000 Monte Carlo replications are generated. Each QMLE run take about 5-10 seconds; each MLE run takes about 3-20 minutes; and each EMM run takes about 2-6 hours. The total computational time for the current version is nearly 12 months. The second stage of EMM estimation is similar to SMM. The simulation size should be at least 30,000. It does stabilize from 50,000 to 75,000, and could have more Monte Carlo errors at 100,000. Thus 50,000 turns out to be a conservative but economical choice (Gallant and Tauchen 1998b). For the EMM stage, I first conducted 1000 simulations using an automatic score generator. Since BIC is suspected to be conservative, the automatic SNP search may pick too few moments, affecting the estimation efficiency and specification test (fully discussed in Section 5.6). Based on the 1000 automatic EMM simulation, I adopted one particular score for the 500 sample size and another for the 1500 sample size. These specifications are at relatively higher dimensions, with abundant occurrence but without severely rejections. Then I performed an additional 1000 simulations for EMM, using these "posterior" fixed score generators. This procedure mimics the realistic situation, when an empirical researcher not only uses BIC as an objective criterion in model selection but also incorporates prior information to set some thresholds levels (Andersen et al. 1999c, Gallant and Tauchen 1998b).

2.4.2 Benchmark Model

To select a suitable parameter setting, we start with the empirical result from Gallant and Tauchen (1998b), $dr_t = (0.02491 - 0.00285 r_t)dt + 0.0275 r_t^{1/2} dW_t$. Using equations 2.3, 2.5, and 2.6-2.8, one can calculate the unconditional mean and variance, the Bessel function order, and the conditional mean and variance. It is not difficult to see that this original specification, Scenario 2.1 in Table 2.1, features low mean-reversion ($E(r_{t+1}|r_t)$ is nearly unit-root) and low conditional volatility ($V(r_{t+1}|r_t)$ is close zero).

17

Also the unconditional variance is unusually small, representing an abnormally quiet process. The order of the Bessel function, twice of which corresponds to the degree of freedom for an integer order noncentral chi-square distribution, is so large that the conditional density looks almost Gaussian. This benchmark model is not a good setting for the purpose of Monte Carlo study, because the distribution is not rich enough to generate overidentified moment conditions; hence the overrejection problem can not be investigated in a realistic setting.

If one increases only the variance parameter b_0 from Scenario 1 to 2-4 in Table 2.1, the Bessel function order q decreases gradually from the Gaussian-like specification, but the conditional volatility is still negligible. Alternatively, one can increase both the mean parameters a_0 and a_1 by a factor of 100 and the variance parameter b_0 by a factor of 10 from Scenario 1 to 5 in Table 1.1, while holding the unconditional mean and variance constant. This change will increase the conditional volatility slightly, but the Bessel function order q is still quite large (resembling a Gaussian-like distribution). If one increases the variance parameter b_0 from Scenario 5 to 6-8 in Table 2.1, both high conditional volatility and small Bessel function order are achieved. Scenario 8 is rich enough for the Monte Carlo study, particularly for the purpose of overrejection test. However, the high persistence in mean is sacrificed somewhat. Therefore one should not arbitrarily relate this setting to the real interest rate, which requires more flexible modeling.

The desired Scenario 8 in Table 2.1, $dr_t = (2.491 - 0.285r_t)dt + 1.1r_t^{1/2}dW_t$, is termed an **HMR-HCV** specification (high-mean-reversion and high-conditional-volatility), and the original Scenario 1 is termed as **LMR-LCV** specification (low mean reversion and low conditional volatility). By choosing **HMR-HCV** instead of **LMR-LCV** as the benchmark model, EMM is put in a least favorable position for studying the asymptotic efficiency and overrejection test, when conventional method

18

of moments could easily outperform EMM (Gallant and Tauchen 1998a). In Scenario 1 the Gaussian-like distribution can be virtually nested by a richer SNP auxiliary model, but in Scenario 8 the genuine type I Bessel function distribution can only be approximated by a suitable score generator.

There is a fundamental concern on how flexible the square-root model can be. In order to do interesting comparative studies, one would like to hold the unconditional mean $(-a_0/a_1)$ constant without explosion $(b_0 \leq 4a_0^2)$, resembling a stationary interest rate process. To achieve high persistence in both conditional mean and variance, one has only three parameters to manipulate. It seems impossible because one has to satisfy four constraints simultaneously. This indicates that the square-root model may not be flexible enough to model the interest rate dynamics. To find a more suitable model requires a fourth degree of freedom, for example, a stochastic volatility component (Gallant and Tauchen 1998b).

2.4.3 Testing Misspecification

The goal is to find a true data generating process, of which an adequate auxiliary score will not accommodate a misspecified model. As discussed in Section 2.1, the square-root model is widely used in fitting the short rate process, but most serious studies have rejected this specification. So it is natural to adopt the square-root process as a misspecified model to and use a non-rejected model as the true data generating process (for the short interest rate). A recent study by Gallant and Tauchen (1998b) gave the most favorable evidence for the "CKLS0-SV-FB" model,

$$dr_t = (0.014 - 0.002r_t)dt + (0.043 - 0.018r_t)e^{u_t}dW_{1t} \qquad (2.20)$$

$$du_t = (-0.006r_t - 0.157u_t)dt + (0.593 - 0.052u_t)dW_{2t}. \qquad (2.21)$$

The short rate process r_t has a linear drift and a linear diffusion, with the diffusion multiplied by an unobserved (exponential) stochastic volatility term e^{u_t}. The short

19

rate r_t is only partially observable in discrete time. The latent stochastic volatility process u_t also has a linear drift and a linear diffusion, with the drift added by a short rate level feedback. This model adequately passed the specification test for various simulation sizes and greatly outperformed the competing models in terms of reprojecting the conditional density and the conditional volatility. It would be our ideal choice of a true data generating process.

When we fit a misspecified square-root model to simulated interest rate data from this CKLS0-SV-FB model,

$$dr_t = (a_0 + a_1 r_t)dt + b_0 r_t^{1/2} dW_t,$$

the drift is correctly specified as a linear function, and the misspecification comes into the diffusion. From Itô's formula we know that the conditional mean is also correctly specified as linear; therefore the drift parameters are consistent estimates (Aït-Sahalia 1996a). It is equivalent to the case where Ordinary Least Square is consistent, but unaccounted heteroskedastic and/or correlated error structure may cause very noisy and inefficient estimates. The misspecified diffusion generates inconsistent estimates of the conditional variance, which may seriously distort the pricing of interest rate sensitive derivatives. Therefore the detection of diffusion misspecification is a critical challenge to financial econometricians.

2.5 Monte Carlo Results

Tables 2.2 to 2.10 and Figures 2.1 to 2.6 summarize the major findings of this paper. The discussions are organized along topics, and extensive comparisons are made across QMLE, EMM, and MLE. The main focuses are finite sample efficiency, over-rejection bias under the null, and detecting maintained misspecification. Both the automatic score generator and the fixed score generator are used in EMM. The likelihood ratio tests in MLE and QMLE provide joint inferences.

20

2.5.1 Simulation Schemes

The *Poisson*-mixing-*Gamma* formula is a useful characterization for the square-root model. The simulation accuracy based on the distribution function can provide an independent check for the derivations in Section 2.2. It is not only the Monte Carlo study needs a solid basis; further applications of MLE with this formula also require some justification. The discretized approximation to stochastic differential equations is a cornerstone of the EMM estimator, as a simulation-based estimator. EMM uses a weak-order 2 scheme (Kloeden and Platen 1992). To assess these simulation approaches, some empirical statistics from long realizations (100,000) are compared to their theoretical counterparts.

Table 2.2 lists the calculation of two moments and three quantiles. Clearly both schemes work reasonably well, with the relative error ranging from 0.12% to 1.8%. Since our benchmark model features strong mean reversion and is far away from unit root ($E(r_{t+1}|r_t) = 0.75r_t$) and is hence stationary, the estimate of the unconditional mean is very precise. It is not surprising that the probabilistic method is superior to the discretized method, although the difference is negligible.

2.5.2 Score Generator

An important feature of EMM is the endogenous moment selection by a seminonparametric score generator (SNP), which contrasts with the *ad hoc* choice of moment conditions in GMM or SMM. The optimal SNP search and the inexpensive weighting matrix estimate are key to the efficiency argument and hopefully also improve the overrejection test. It is worthwhile to check whether the SNP score captures the distribution features of different dependent structures before launching the full-scale Monte Carlo experiment. Tables 2.3 (500 sample size) and 2.4 (1500 sample size) report the SNP searches for the 8 scenarios in Table 2.1. For each setting, the

21

frequencies of all kinds of model choices among 100 replications are listed. Model dimension is represented by a five-digit number, which stands for, consecutively, lag in mean, lag in variance, lag in polynomial, degree of Hermite polynomial, and degree of Hermite coefficient polynomial.

Scenario 1 in Tables 2.3 and 2.4 is the **LMR-LCV** case (low-mean-reversion, low-conditional-volatility). Not surprisingly the Gaussian auto-regression specification 10100 dominates other choices. It is consistent with the fact that the true density is close to Gaussian under this parameter setting (see Table 2.1). Moving from Scenario 1 to 2, 3, and 4, the conditional volatility increases gradually, since the variance parameter b_0 is altered (see Table 2.1). The majority choice is still Gaussian, and the chances of ARCH and/or non-Gaussian specifications increase slightly. Moving toward Scenarios 5-8, both mean parameters a_0, a_1, and variance parameter b_0 are altered (see Table 2.1), and ultimately one reaches the **HMR-HCV** case (high-mean-reversion, high-conditional-volatility). It is clear that the SNP search favors the nonlinear, nonparametric AR-ARCH specification. This is consistent with the low Bessel function order and high conditional variance (Table 2.1). Largely due to this "distribution-dependent" or "data-dependent" score generator, the EMM estimator is claimed to be asymptotically efficient and hopefully more reliable in the specification test.

Also evident from Tables 2.3 and 2.4 is that larger sample sizes enable the SNP to pick up higher model dimensions. In fact, the asymptotic efficiency argument requires that the number of moment conditions and the lags entering each moment increase with the sample size (Gallant and Long 1997). Table 2.5 verifies that this requirement is satisfied; i.e., the auxiliary model is getting richer as the sample size increases.

A salient question is whether the structural model can be identified when the SNP

22

search does pick the Gaussian-AR(1) score. This corresponds to a quasi-maximum likelihood estimator based on

$$z_t = (r_t - \alpha_0 - \alpha_1 r_{t-1})/\alpha_2 \sim N(0,1). \tag{2.22}$$

Since the conditional mean is correctly specified, the QMLE of α_0 and α_1 is a consistent estimators of e^{a_1} and $-a_0/a_1(1 - e^{a_1})$. It is just an Ordinary Least Square with a heteroskedastic and serially correlated error term (Aït-Sahalia 1996a). The conditional variance is misspecified as the constant α_2. However, according to the theory of misspecified maximum likelihood estimation (White 1994), the estimator $\hat{\alpha}_2$ converges to the pseudo-true value $\bar{\alpha}_2$. The key argument is that the misspecified asymptotic variance $\bar{\alpha}_2$ must be a function of the true variance parameter b_0, since both conditional variance and unconditional variance are determined by b_0. These asymptotic relations, two explicit and one implicit, are indeed the *binding functions* in the language of Indirect Inference (Gourieroux, Monfort, and Renault 1993). Obviously the structural parameters a_0, a_1, and b_0 are exactly identified by the auxiliary parameters α_0, α_1, and α_2. EMM is a feasible first-order approximation to the Indirect Inference (Gallant and Long 1997).

2.5.3 Dispersion and Asymmetry

The asymptotic theory suggests that the sampling distribution of parameter estimates should be approximately normal and become more concentrated as the sample size increases. Table 2.6 summarizes the relevant quantiles and mean statistics, across three parameters, between 500 and 1500 sample sizes and among different estimators.

The variance parameter b_0 is almost symmetric (Mean $\approx$ Median) and is very concentrated about the median. The drift parameters a_0 and a_1 are less symmetric and less concentrated. This result coincides with a well-documented fact that the

23

volatility can be precisely estimated, even if the drift estimate is inaccurate (Aït-Sahalia 1996a). All estimators seem to be internally consistent since both asymmetry (difference between Mean and Median) and dispersion (distance between Median and other Quantiles) across three parameters uniformly reduce as the sample size increases. Apparently MLE is the least dispersed and the most symmetric for all three parameters at all sample sizes. QMLE is better than the automatic score EMM but worse than the fixed score EMM. It is worth noting that the improvement of EMM over different sample sizes is larger than that of QMLE or MLE. A possible explanation may be that the SNP search is able to capture more distribution features with larger sample sizes by choosing a richer score.

2.5.4 Bias, RMSE, and Relative Efficiency

Theoretically the information matrix of MLE reaches the Cramer-Rao lower bound and hence is fully efficient. The efficiency of EMM approaches that of MLE, as the root-mean-squared-error (RMSE) reduces faster than $\sqrt{n}$ in finite samples. QMLE is superior to the automatic score EMM but inferior to the fixed score EMM.

In all cases, the parameter estimates are NOT significantly biased (Table 2.7). Also the biases decrease as sample size increases. The signs of biases for EMM are persistent, with upward bias for a_0 and downward bias for a_1 and b_0. One possible cause may be the asymmetry of the EMM objective function. The signs of biases for QMLE and MLE are not persistent. Overall the magnitudes of biases are negligible. Theoretically, all these estimators are consistent with zero asymptotic bias.

It is clear from Table 2.7 that the RMSEs of QMLE and MLE are shrinking almost exactly at the rate $\sqrt{3}$, as the sample size increase from 500 to 1500. For EMM, the convergence rates are greater than $\sqrt{3}$. The drift parameter estimates are less efficient, with the RMSEs ranging from 5% to 30% of the parameter value. The

24

variance parameter estimates are very efficient, with the same ratios ranging from 3% to 6%. Such a disparity between the drift and volatility estimates is large for automatic score EMM and is very small for MLE. The RMSEs of QMLE and fixed score EMM are in between. Overall, these estimators seem quite efficient.

The relative efficiency of EMM is rising significantly from 500 sample to 1500 sample (Table 2.8). This finding must be correct since EMM is proven to be asymptotically as efficient as MLE. QMLE has a computational advantage with the analytical gradients, which is not available in EMM or MLE. The advantage explains why QMLE is more efficient than automatic score EMM and why the QMLE of variance parameter estimate can be more efficient than MLE. Taking this fact into account, the fixed score EMM apparently matches or surpassed QMLE when sample size increases to 1500. The main reason for the increasing relative efficiency of EMM is that the SNP score generator is able to approximate the underlying distribution arbitrarily close, with more moments and lags in larger samples. The Monte Carlo result shows that for automatic score EMM, the average number of overidentified moment conditions is 3 for $N = 500$ and 5 for $N = 1500$. The educated choice of fixed score EMM set the number of overidentified moment conditions at 5 for $N = 500$ and 7 for $N = 1500$. Whenever MLE is unavailable, EMM will be favored by longer data series. However in small econometric samples, the computational efficiency of QMLE may outweigh the asymptotic efficiency of EMM. The choice must be made case-by-case (Dai and Singleton 2000).

2.5.5 Statistical Inferences

With a knowledge of the true specification, one can perform a likelihood ratio test to see whether the confidence ball centered at the estimated value contains the true parameter value. According to the null hypothesis, ideally the test should contain the

25

true model at 100%. The bottom line must be that the rejection rate does not exceed the asymptotic size. Table 2.9 shows that the MLE approach defined by the mixture formula in Section 2 is fully reliable. The inference based on the likelihood ratio test is above the significance level and close to full containment. The improvement from $T = 500$ to $T = 1500$ is consistent with the asymptotic argument that as $n \to \infty$, the confidence ball should concentrate to a single point of truth. Not surprisingly the likelihood inferences for QMLE fall short of the asymptotic level and do not improve as the sample size increases. As a misspecified MLE, the likelihood ratio for QMLE is only valid in an asymptotic sense.

2.5.6 Overrejection Bias

Even with the improved estimates of weighting matrix, the small-sample chi-square test in EMM still exceeds its asymptotic size. The simulation results on automatic score EMM can reveal some connections between the number of overidentifying moments and the overrejection rate. There is a significant improvement for the fixed score EMM. The remaining small sample bias can be corrected by the sampling distribution of the test statistics.

The 5% gross overrejection rate in automatic score EMM is about 20% for $T = 500$ and about 25% for $T = 1500$. As pointed out in Section 3, BIC tends to underfit if the true model dimension is increasing with sample size. By choosing Scenario 8 in Table 2.1, the EMM estimator is in a least favorable position for the overrejection test, since the true model has a high conditional volatility and is a non-Gaussian innovation. Figures 2.1 and 2.2 plot the rejection rates with the number of overidentified moment conditions, which is automatically chosen by BIC. The asymptotic size of the specification test is fixed at 5%. The occurrence rates show the frequencies of different numbers of moment conditions in 1000 replications. Some important

26

features need to be mentioned. First, the rejection curve does not uniformly shoot up when more moment conditions are included, since these moments are optimally selected by the SNP score generator. Second, the rejection rates are more stable at $T = 1500$ than $T = 500$, as more moments and lags are included. Third, the rejection rate could be remarkably small for certain low dimensions as well as for some high dimensions. Since BIC tends to underfit the auxiliary model in small samples, the higher level unrejected score is more likely capturing the true distribution. If the lower level unrejected score did pick the true specification, the rejection rate should uniformly shoot up beyond that level. The implication for empirical work is that an SNP search should go beyond the first optimal choice by BIC.

The educated choices of model dimensions in fixed-score EMM are 10111 with 5 overidentified moments at $T = 500$ and 10121 with 7 overidentified moments at $T = 1500$. At these two scores (Figures 2.1 and 2.2), not only the rejection rate should be small ($\leq 20\%$), but also the overall occurrence of that choice of that model dimension should be high ($\geq 10\%$). The simulation results for fixed0score EMM are more encouraging, some of which are discussed in previous subsections. As for the rejection rates, the 1%, 5%, and 10% size levels are respectively 8.7%, 17.4%, and 23.5% for $T = 500$, and 5.0%, 10.3%, and 16.3% for $T = 1500$ (see Figure 2.3). The improvement over the automatic score EMM is remarkable, and the improvement from $T = 500$ to $T = 1500$ is more than 50%.

There are at least four sources of overrejection bias in GMM-type estimators: inaccurate and costly estimates of the weighting matrix, *ad hoc* selection of the moment conditions, an inadequate number of moments to capture the distribution feature, and simply a small sample bias. A generic EMM approach overcomes the first two problems by adopting a serially uncorrelated information matrix and an optimal SNP score generator. The conservative BIC in EMM may choose too few moments, but

27

one can rectify this by using additional information and extending the SNP search beyond the BIC choice. The remaining small sample bias can be remedied by enlarging the sample size or by adopting a model-dependent optimal information criterion. If sample size is fixed and no better information criterion is available, an *ex post* correction of the test statistics may be applicable (Conley, Hansen, and Liu 1997b).

Table 2.10 gives the original and adjusted p-values for this particular square-root model. As expected, the correction is large for $T = 500$ and small for $T = 1500$. To calculate the adjustment, one first finds the critical value of a q'th quantile from the exact distribution of test statistics, then finds the p'th percentile corresponding to this critical value from the sampling distribution of the test statistics. In last row in Table 2.10, if the theoretical p-value is 10^{-6}, the adjusted p-value is 0.01 for $T = 500$ and 0.003 for $T = 1500$. If the corrected p-value is less than 0.01, then the rejection is likely to be final.

2.5.7 Detecting Misspecification

In the first stage, we use the benchmark stochastic volatility model (Section 4.3) to simulate twice 1000 replications of both 500 and 1500 sample sizes, and we fit a square-root diffusion process to the data. This time we let BIC automatically choose the best SNP score generator. Since the drift is linear, the conditional mean with lag one is correctly specified. For the 500 sample size, 91% trials select lag 1; and for the 1500 sample size, 93% select lag 1. The choice of conditional standard deviation is all over the place, due the nature of nonlinear stochastic volatility. For $T = 500$, the selection is scattered mainly from lag 1 to lag 4, and for $T = 1500$, it is mainly from lag 3 to lag 6. The choices of K_z and K_x are predominantly zero. Figure 2.4 and 2.5 plot the 5% rejection rates against the number of overidentified moments along with the occurrence ratio of these moment choices. The highlight is that the probability of

28

rejecting a misspecified model does converge to one very quickly. At $T = 500$, the 5% level rejection rate is around 80-90% for a range of overidentified moments between 1 to 6, and rejection rate almost 100% beyond that (Figure 2.4). At $T = 1500$, the rejection rate is always close to 100%, except in an exactly identified case (Figure 2.5).

In the second stage, similar to the studying of overrejection issue, we fix the SNP score generator and look at the rejection rate uniformly along the 1%-100% test level. The fix score generator for the 500 sample size is s1310000, which has 6 parameters. Since the misspecified square-root model has 3 parameters, the chi-square test is of degree 3. When $T = 1500$, the score is fixed at s1510000, with $8 - 3 = 5$ degree of freedom. Figure 2.6 gives the q-q plot from the sampling distribution of the test statistics. For the 500 sample size, the smallest rejection rate is 50% when the test level is 1%, and the rejection rate quickly reaches 90% at the 15% level. The rejection curve of $T = 1500$ starts out at 98% and quickly converges to 100%. The convergence of the rejection probabilities towards 100% is extremely fast from $T = 500$ to $T = 1500$. Since the misspecification rejection curves in Figure 6 are way above the overrejection curves in Figure 3, no *ad hoc* adjustment of p-value (such as those in previous subsection) can mistaken the intrinsic misspecification as an overrejection bias.

2.6 Conclusions

This paper investigates the finite sample properties of Efficient Method of Moments in conjunction with Maximum Likelihood Estimations for a square-root diffusion process. Major findings are:

By optimally choosing the moment conditions, the EMM estimator tends to be asymptotically as efficient as MLE. The overrejection bias of EMM converges to zero

29

when the sample size increases, and the rejection rate of misspecification is quickly converging to one with the sample size.

MLE for the square-root model is well defined by factorizing the transitional density into a *Poisson*-mixing-*Gamma* distribution. QMLE has attractive computational efficiency in small samples. The likelihood-ratio inference suggests that MLE is more reliable than QMLE.

2.7 Tables and Figures

Table 2.1: Benchmark Model Choice

The square-root model is $dr_t = (a_0 + a_1 r_t)dt + b_0 r_t^{1/2} dW_t$. Scenario 1 is taken down from Gallant and Tauchen (1998b). In Scenarios 2-4, the variance parameter b_0 is increased by a factor of 2, 3, and 4 respectively. From Scenario 1 to Scenario 5, the mean parameters a_0 and a_1 are multiplied by 100 and the variance parameter b_0 is multiplied by 10. From Scenario 5 to Scenarios 6-8, the variance parameter b_0 is increased by a factor of 2, 3, and 4 respectively. $E(r_t)$, $V(r_t)$, q, $E(r_{t+1}|r_t)$, and $V(r_{t+1}|r_t)$ are calculated using equations 2.3, 2.4, and 2.6-2.8.

$a_0 = 0.02491$ $a_1 = -0.00285$	Scenario 1 $b_0 = 0.0275$	Scenario 2 $b_0 = 0.055$	Scenario 3 $b_0 = 0.0825$	Scenario 4 $b_0 = 0.11$
$E(r_t)$	8.74	8.74	8.74	8.74
$V(r_t)$	1.16	4.64	10.44	18.55
Bessel q	64.88	15.47	6.32	3.12
$E(r_{t+1}|r_t)$	$0.997r_t + 0.025$	$0.997r_t + 0.025$	$0.997r_t + 0.025$	$0.997r_t + 0.025$
$V(r_{t+1}|r_t)$	$0.001r_t + 0.000$	$0.003r_t + 0.000$	$0.007r_t + 0.000$	$0.012r_t + 0.000$
$a_0 = 2.491$ $a_1 = -0.285$	Scenario 5 $b_0 = 0.275$	Scenario 6 $b_0 = 0.55$	Scenario 7 $b_0 = 0.825$	Scenario 8 $b_0 = 1.1$
$E(r_t)$	8.74	8.74	8.74	8.74
$V(r_t)$	1.16	4.64	10.44	18.55
Bessel q	64.88	15.47	6.32	3.12
$E(r_{t+1}|r_t)$	$0.75r_t + 2.17$	$0.75r_t + 2.17$	$0.75r_t + 2.17$	$0.75r_t + 2.17$
$V(r_{t+1}|r_t)$	$0.05r_t + 0.07$	$0.20r_t + 0.29$	$0.45r_t + 0.64$	$0.79r_t + 1.14$

Table 2.2: Simulation Schemes (100,000 Length)

For the square-root model $dr_t = (a_0 + a_1 r_t)dt + b_0 r_t^{1/2} dW_t$, the marginal distribution is a *Gamma* (equation 2.2), and the theoretical values are calculated accordingly. Simulation by distribution is based on the *Poisson*-mixing-*Gamma* formula (equation 2.11). Simulation by discretization is based on the weak-order 2 scheme (Gallant and Long 1997).

	Simulated by Distribution	Simulated by Discretization	Theoretical Value
Mean	8.75	8.70	8.74
Variance	18.34	18.22	18.55
5% Quantile	3.04	3.03	3.05
Median	8.08	8.02	8.05
95% Quantile	16.70	16.67	16.81

31

Table 2.3: SNP Search for 500 Sample Size

These results are from 100 replications of each scenario with 500 and 1500 sample sizes. The information criterion used in moment selection is Schwartz's BIC. Scenarios 1-8 are the same as those in Table 2.1. Each model specification is characterized by a 5-digit number. Consecutively each digit stands for lag in mean, lag in variance, lag in polynomial, degree of Hermite polynomial, and degree of Hermite coefficient polynomial.

Scenario	1	Scenario	2	Scenario	3	Scenario	4
Model	%	Model	%	Model	%	Model	%
10100	95	10100	96	10100	92	10100	88
11100	3	20100	3	10110	4	11100	7
20100	1	10110	1	11100	2	10110	1
21100	1			20100	2	10120	1
						20100	1
						21100	1
						11111	1

Scenario	5	Scenario	6	Scenario	7	Scenario	8
Model	%	Model	%	Model	%	Model	%
10100	90	10100	77	10100	46	10111	25
10110	5	10110	12	10110	25	10100	18
20100	3	11110	4	10111	8	10110	15
11100	1	11100	3	10121	7	10121	11
10120	1	10111	1	10120	6	11110	10
		10120	1	11110	3	11120	6
		12110	1	10130	1	10120	4
		20100	1	11100	1	12120	3
				11120	1	10131	2
				11130	1	11130	2
				12110	1	10112	1
						11111	1
						15110	1
						21120	1

Table 2.4: SNP Search for 1500 Sample Size

Scenario	1	Scenario	2	Scenario	3	Scenario	4
Model	%	Model	%	Model	%	Model	%
10100	96	10100	96	10100	85	10100	57
10110	2	11100	4	11100	11	11100	15
20100	1			20100	2	12100	9
11100	1			12100	1	13100	5
				13100	1	11111	2
						15100	2
						16100	2
						11121	1
						14100	1
						14110	1
						15110	1
						16110	1
						16111	1
						18100	1
						25100	1

Scenario	5	Scenario	6	Scenario	7	Scenario	8
Model	%	Model	%	Model	%	Model	%
10100	89	10111	43	10111	41	12120	12
10110	9	10100	30	10121	13	10131	11
10120	1	10110	10	11110	11	10121	10
11110	1	10120	6	11111	8	11110	8
		10121	4	11121	8	10121	7
		11110	2	10131	7	11140	7
		12120	2	10111	5	10111	5
		11111	1	12120	3	11130	5
		20111	1	10120	2	13120	5
		20120	1	10141	1	10122	4
				12110	1	12130	4
						11111	1
						11141	1
						11150	1
						11160	1
						12110	1
						14120	1
						21120	1
						21140	1
						22120	1

33

Table 2.5: Sample Size and SNP Search

For 500 and 1500 sample sizes, results are the top choices among 100 replications. For 5000, 10000, and 50000 sample sizes, results are single trials. Schwartz's BIC is used in the optimal model selection. The benchmark model is Scenario 8 in Table 2.1.

Sample Size	SNP Specification
500	10111
1500	12120
5000	12160
10000	12161
50000	16161

34

Table 2.6: Quantiles and Mean

The number of Monte Carlo replications is 1000.

True Value	5% Quantile	25% Quantile	50% Median	75% Quantile	95% Quantile	Mean
QMLE: 500 Sample Size						
a_0= 2.491	2.0472	2.3190	2.4932	2.7847	3.1932	2.5652
a_1=-0.285	-0.3682	-0.3211	-0.2921	-0.2653	-0.2315	-0.2950
b_0= 1.100	1.0313	1.0738	1.1015	1.1313	1.1737	1.1023
QMLE: 1500 Sample Size						
a_0= 2.491	2.1837	2.3581	2.4916	2.6591	2.8764	2.5134
a_1=-0.285	-0.3332	-0.3023	-0.2861	-0.2706	-0.2472	-0.2870
b_0= 1.100	1.0608	1.0842	1.0999	1.1174	1.1429	1.1003
EMM Automatic Score: 500 Sample Size						
a_0= 2.491	1.9677	2.2767	2.5039	2.7394	3.2655	2.5789
a_1=-0.285	-0.4007	-0.3215	-0.2899	-0.2588	-0.2193	-0.3014
b_0= 1.100	0.9883	1.0512	1.0858	1.1097	1.1577	1.0819
EMM Automatic Score: 1500 Sample Size						
a_0= 2.491	2.1244	2.3172	2.4591	2.5906	2.8751	2.4851
a_1=-0.285	-0.3401	-0.3014	-0.2846	-0.2678	-0.2409	-0.2872
b_0= 1.100	1.0356	1.0713	1.0898	1.1056	1.1440	1.0898
EMM Fixed Score: 500 Sample Size						
a_0= 2.491	2.1076	2.4066	2.5343	2.7463	3.3419	2.6233
a_1=-0.285	-0.4158	-0.3357	-0.3048	-0.2820	-0.2460	-0.3160
b_0= 1.100	0.9617	1.0540	1.0909	1.1072	1.1571	1.0782
EMM Fixed Score: 1500 Sample Size						
a_0= 2.491	2.1925	2.3529	2.4737	2.6013	2.8268	2.4843
a_1=-0.285	-0.3321	-0.3023	-0.2850	-0.2712	-0.2486	-0.2872
b_0= 1.100	1.0420	1.0715	1.0877	1.1040	1.1247	1.0863
MLE: 500 Sample Size						
a_0= 2.491	2.2021	2.3746	2.4231	2.4575	2.5106	2.4078
a_1=-0.285	-0.3109	-0.2920	-0.2821	-0.2591	-0.2377	-0.2765
b_0= 1.100	1.0239	1.0756	1.1060	1.1286	1.1704	1.1024
MLE: 1500 Sample Size						
a_0= 2.491	2.3086	2.3888	2.4386	2.4624	2.5008	2.4274
a_1=-0.285	-0.3026	-0.2879	-0.2840	-0.2693	-0.2520	-0.2792
b_0= 1.100	1.0555	1.0793	1.1000	1.1169	1.1399	1.0984

35

Table 2.7: Accuracy and Efficiency

The number of Monte Carlo replications is 1000.

True Value	Mean Bias	Median Bias	RMSE
QMLE: 500 Sample Size			
a_0= 2.491	0.0742	0.0022	0.3613
a_1=-0.285	-0.0100	-0.0071	0.0448
b_0= 1.100	0.0023	0.0015	0.0430
QMLE: 1500 Sample Size			
a_0= 2.491	0.0224	0.0006	0.2111
a_1=-0.285	-0.0020	-0.0011	0.0258
b_0= 1.100	0.0003	-0.0001	0.0246
EMM Automatic Score: 500 Sample Size			
a_0= 2.491	0.0879	0.0129	0.6684
a_1=-0.285	-0.0164	-0.0049	0.1017
b_0= 1.100	-0.0181	-0.0142	0.0764
EMM Automatic Score: 1500 Sample Size			
a_0= 2.491	-0.0059	-0.0319	0.3305
a_1=-0.285	-0.0022	0.0004	0.0384
b_0= 1.100	-0.0102	-0.0102	0.0401
EMM Fixed Score: 500 Sample Size			
a_0= 2.491	0.1323	0.0433	0.4891
a_1=-0.285	-0.0310	-0.0199	0.0694
b_0= 1.100	-0.0218	-0.0091	0.0618
EMM Fixed Score: 1500 Sample Size			
a_0= 2.491	-0.0067	-0.0173	0.2000
a_1=-0.285	-0.0022	-0.0000	0.0257
b_0= 1.100	-0.0137	-0.0122	0.0296
MLE: 500 Sample Size			
a_0= 2.491	-0.0832	-0.0679	0.1337
a_1=-0.285	0.0085	0.0029	0.0251
b_0= 1.100	0.0024	0.0060	0.0432
MLE: 1500 Sample Size			
a_0= 2.491	-0.0663	-0.0524	0.0923
a_1=-0.285	0.0058	0.0010	0.0161
b_0= 1.100	-0.0016	0.0000	0.0263

36

Table 2.8: Relative Efficiency of QMLE and EMM

The theoretical relative efficiencies of QMLE and EMM are the asymptotic variance ratios of MLE over QMLE and EMM. In a finite sample Monte Carlo study, they are approximated by the ratios of RMSEs.

	QMLE		EMM Automatic Score		EMM Fixed Score	
	T = 500	T = 1500	T = 500	T = 1500	T = 500	T = 1500
a_0	37.01%	43.72%	20.11%	27.93%	27.34%	46.15%
a_1	56.03%	62.40%	24.68%	41.82%	36.17%	62.65%
b_0	100.47%	106.91%	56.54%	65.59%	69.90%	88.85%

Table 2.9: Likelihood Ratio Test for MLE and QMLE

The likelihood ratio test statistics is $2(\mathcal{L}_n(\theta^*) - \mathcal{L}_n(\theta^o))$, where $\mathcal{L}_n(\theta^*)$ is the unconstrained loglikelihood value with optimized parameters and $\mathcal{L}_n(\theta^o)$ is the constrained loglikelihood value with true parameters.

	Maximum Likelihood		Quasi-Maximum Likelihood	
Test Level	500 Sample	1500 Sample	500 Sample	1500 Sample
99%	100.0%	100.0%	97.6%	97.0%
95%	99.0%	99.7%	91.5%	91.4%
90%	98.1%	98.8%	86.7%	86.7%

37

Table 2.10: p-Value Adjustment for EMM Specification Test

The theoretical distribution of the test statistics is $\mathcal{X}^2(5)$ for $T = 500$ and $\mathcal{X}^2(7)$ for $T = 1500$. The adjustment is based on the sampling distribution of each 1000 simulations.

Theoretical p-Value	Adjusted p-Value $T = 500$	$T = 1500$
0.10	0.24	0.16
0.09	0.22	0.15
0.08	0.21	0.14
0.07	0.20	0.13
0.06	0.19	0.12
0.05	0.17	0.10
0.04	0.16	0.09
0.03	0.14	0.08
0.02	0.11	0.06
0.01	0.08	0.05
10^{-3}	0.034	0.019
10^{-4}	0.018	0.007
10^{-5}	0.013	0.004
10^{-6}	0.010	0.003

38
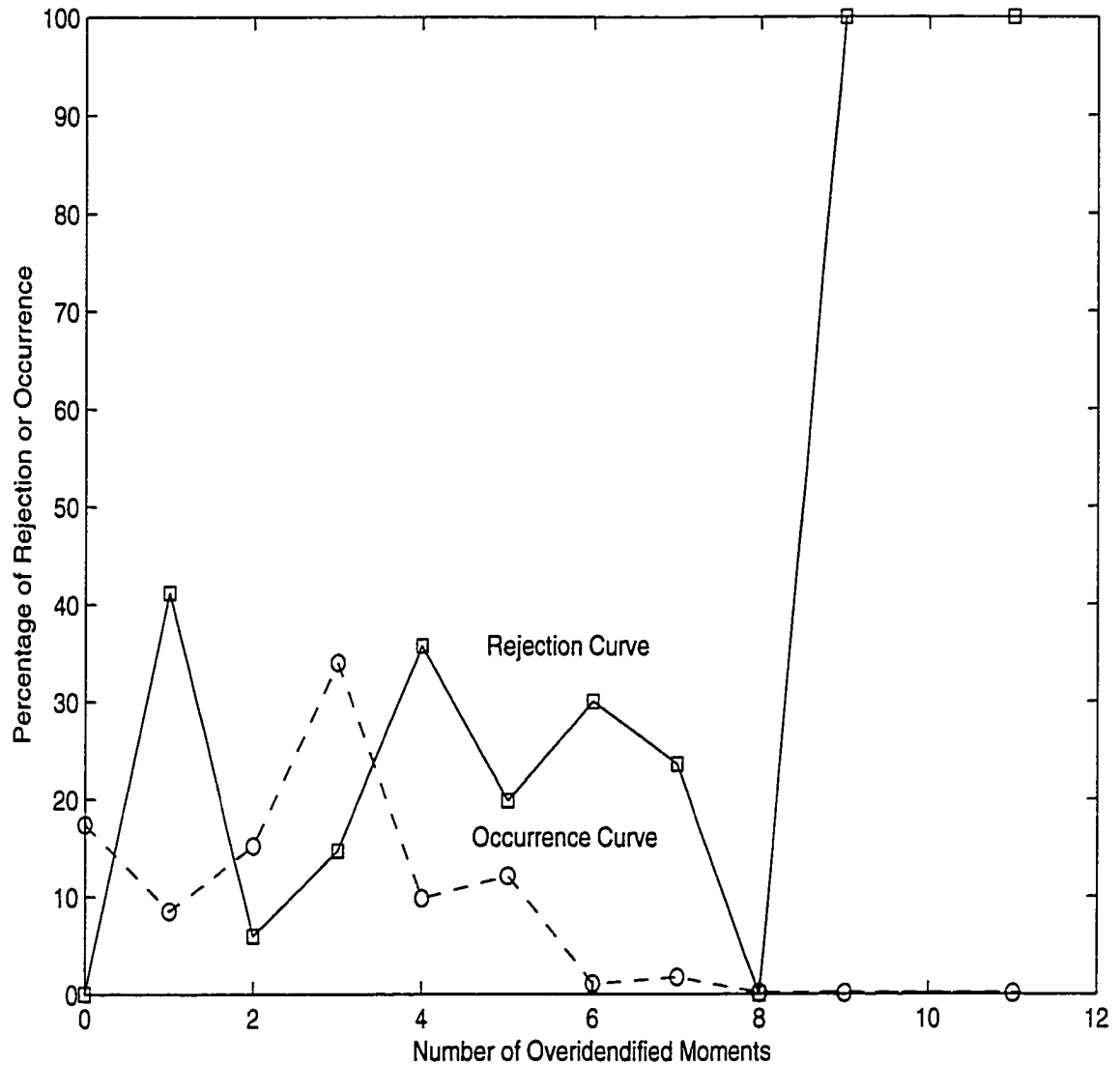
Figure 2.1: 5% Rejection Rate of EMM $T = 500$ with Automatic Score Generator. The occurrence rate is the frequency of the same moment choice divided by 1000. The rejection rate is the frequency of rejections divided by the number of occurrences.

39
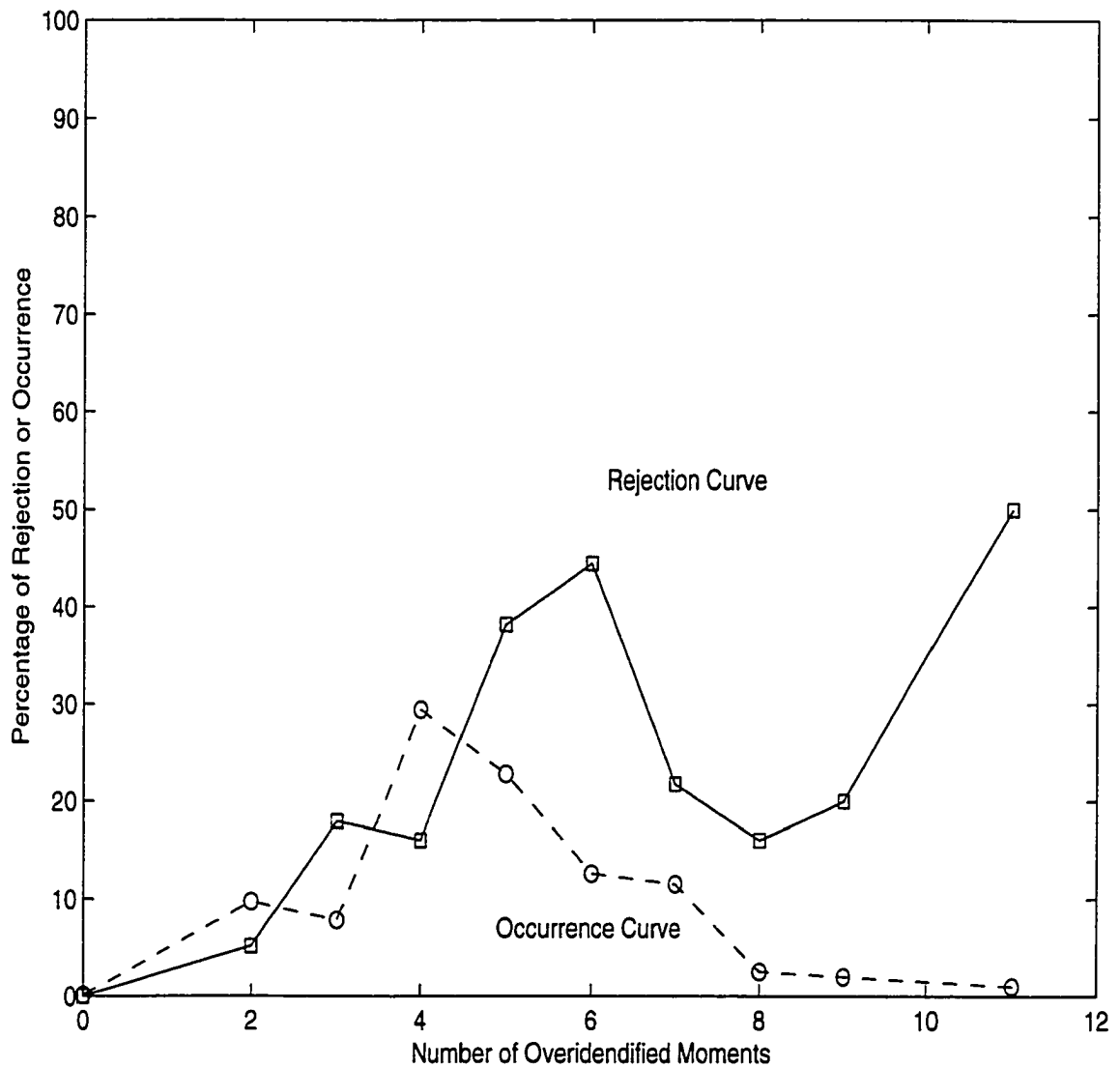
Figure 2.2: 5% Rejection Rate of EMM $T = 1500$ with Automatic Score Generator. The occurrence rate is the frequency of the same moment choice divided by 1000. The rejection rate is the frequency of rejections divided by the number of occurrences.

40
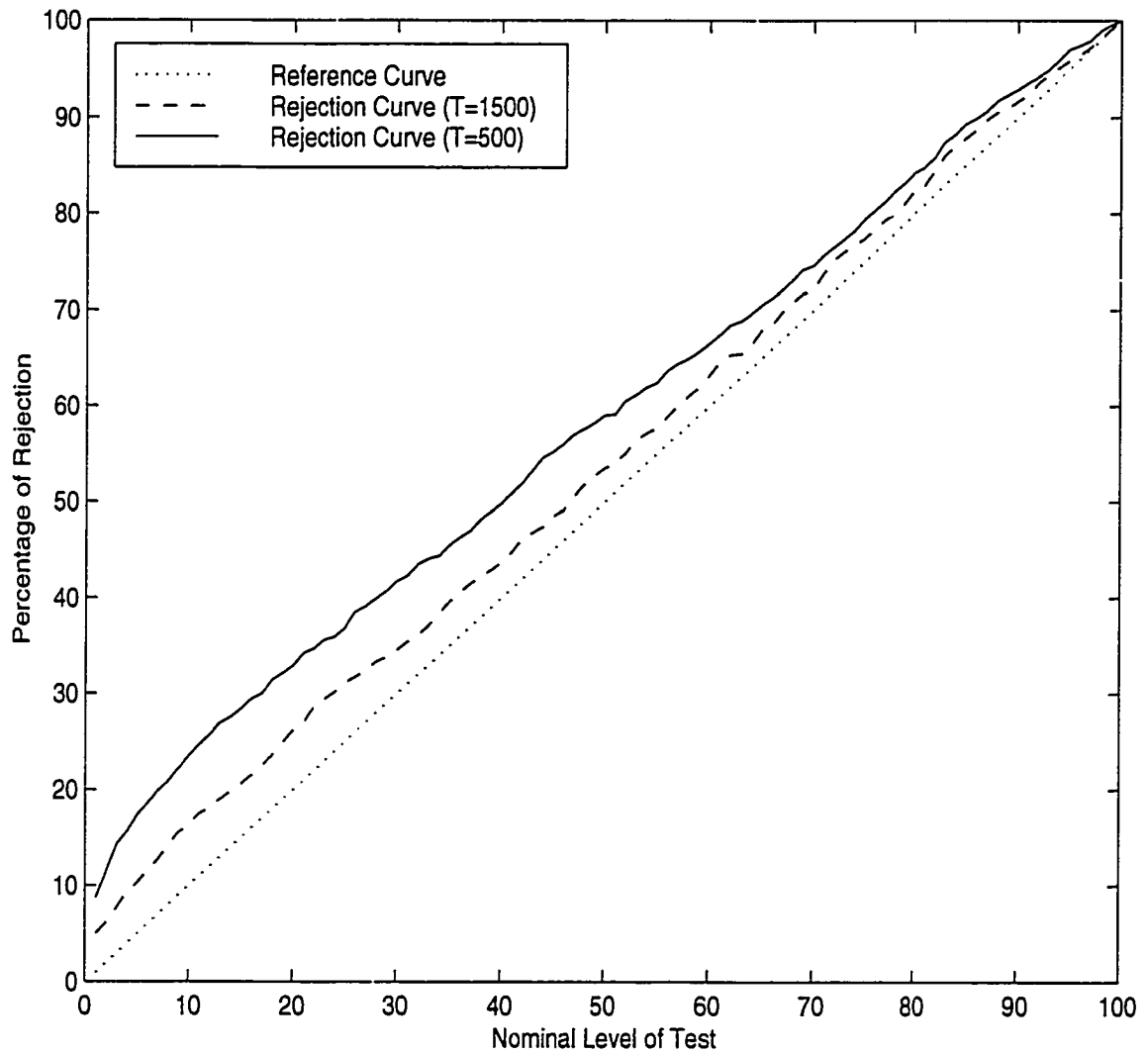
Figure 2.3: Overrejection Curve of EMM with Fixed Score Generator

41
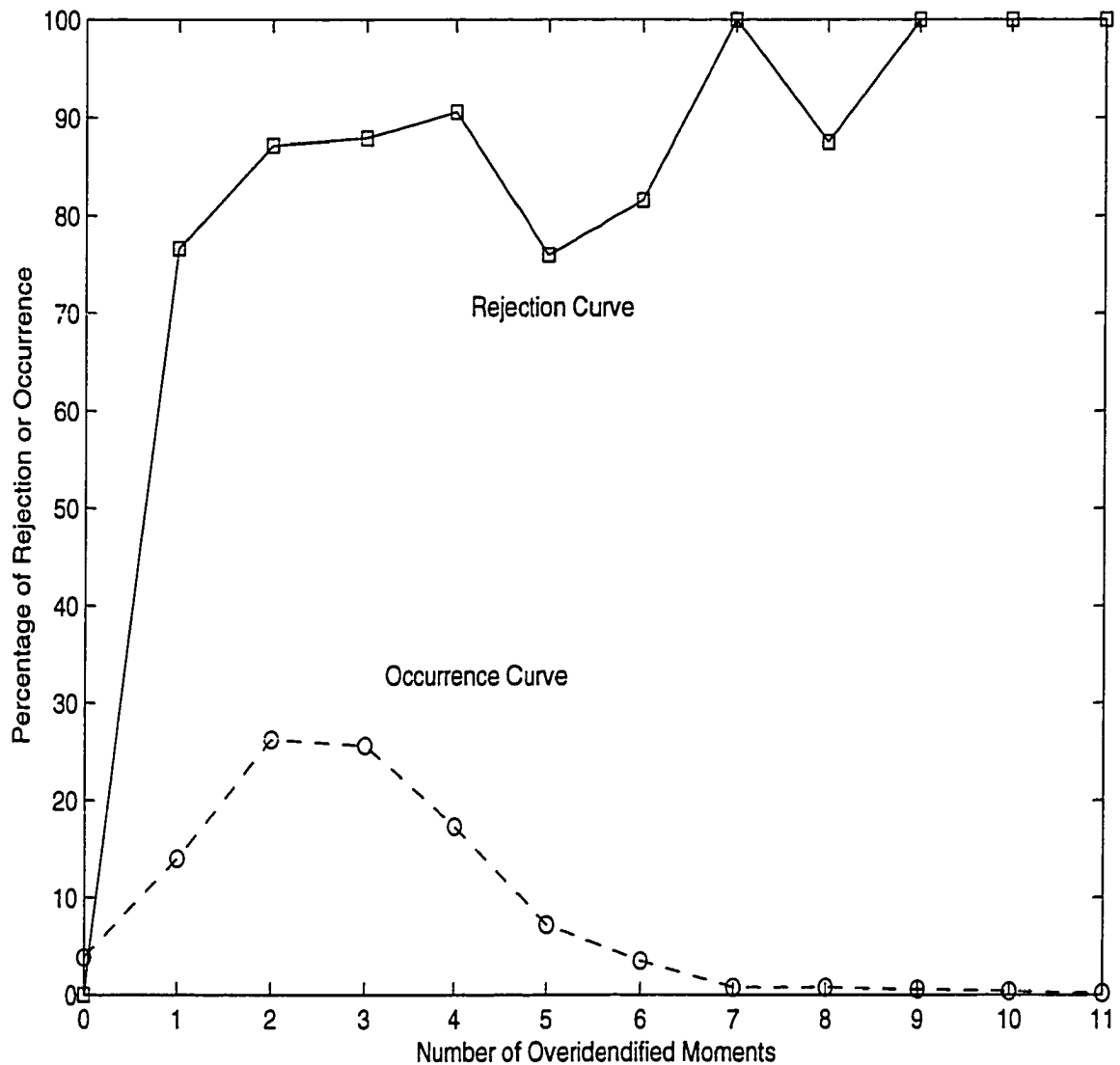
Figure 2.4: 5% Rejection of Misspecified Model $T = 500$
The occurrence rate is the frequency of the same moment choice divided by 1000.
The rejection rate is the frequency of rejections divided by the number of occurrences.

42
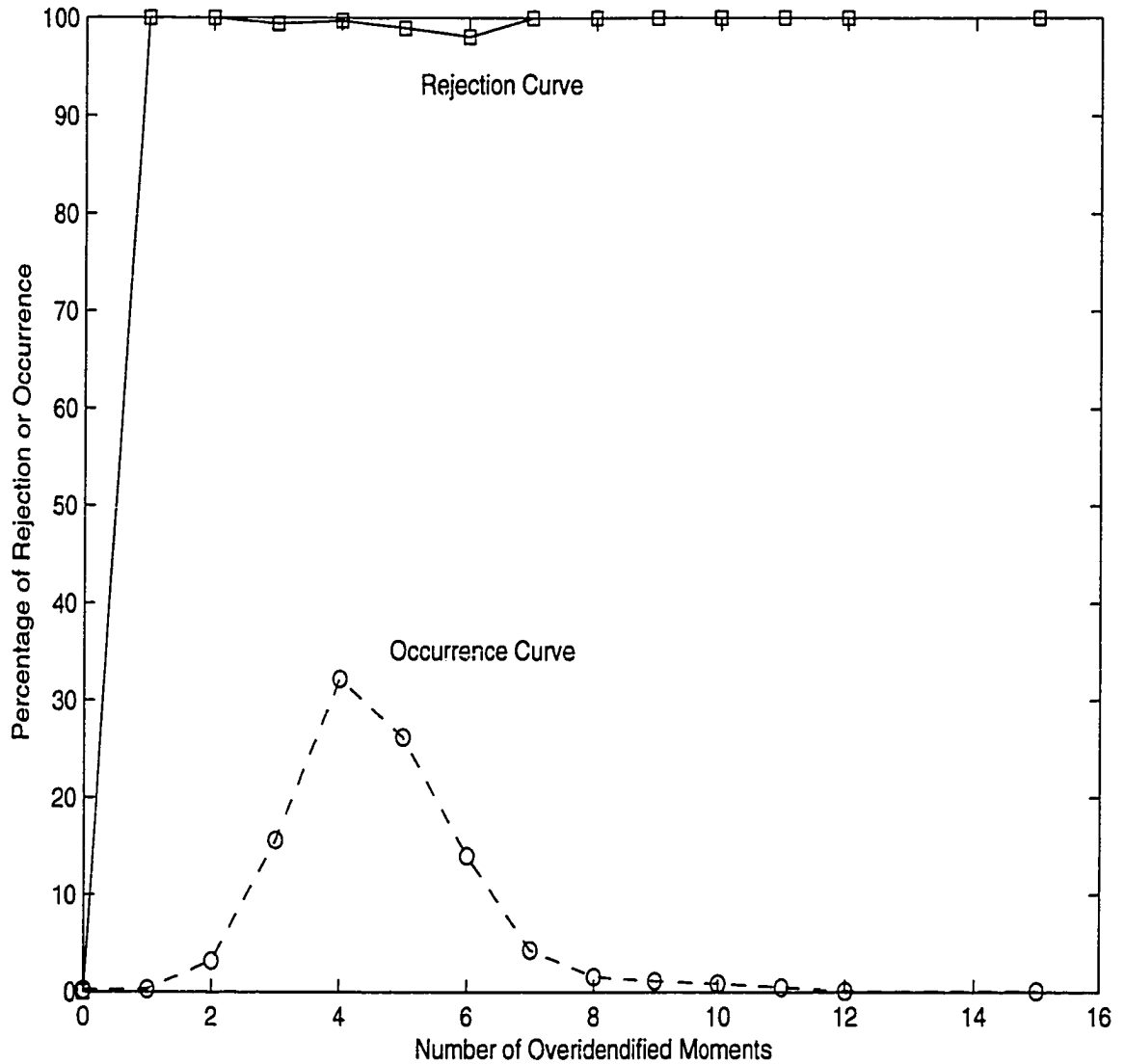
Figure 2.5: 5% Rejection of Misspecified Model $T = 1500$
The occurrence rate is the frequency of the same moment choice divided by 1000.
The rejection rate is the frequency of rejections divided by the number of occurrences.

43
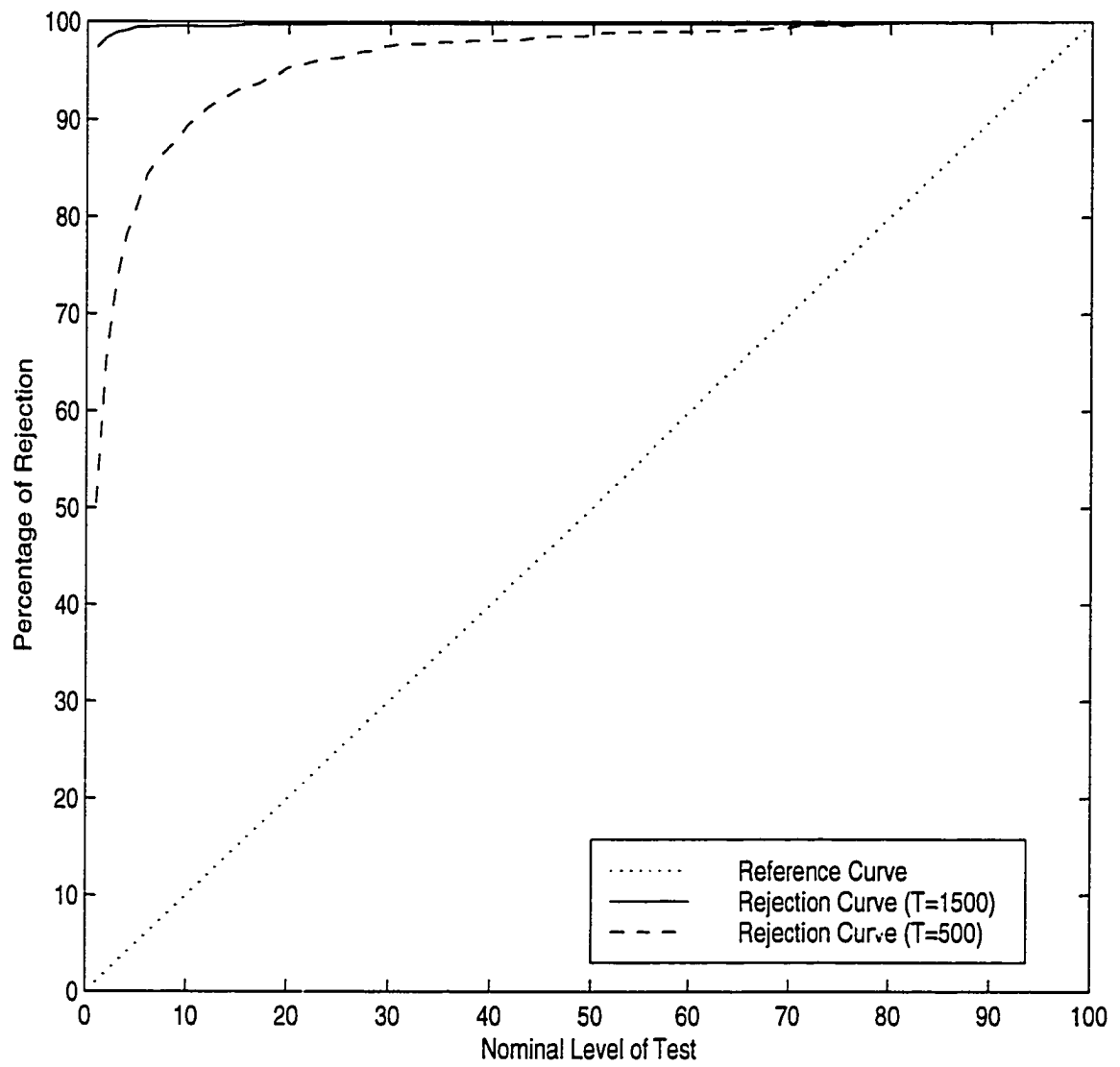
Figure 2.6: Rejection Curves of Misspecified Model with Fixed Score Generator

44

Chapter 3

Estimation of A Jump-Diffusion Term Structure Model

[1]This chapter develops a Multivariate Weighted Nonlinear Least Square estimator for a class of jump-diffusion interest rate processes (hereafter MWNLS-JD), which also admit closed-form solutions to bond prices under an equilibrium no-arbitrage argument. The instantaneous interest rate is modeled as a mixture of a continuous square-root diffusion and a discrete Poisson jump process, where the jump-rate and size can be constants, state-dependent functions, or independent random variables. We can derive analytically the first four conditional moments, which form the basis of our MWNLS-JD estimator. A diagnostic conditional moment test and a classical Lagrange multiplier test can also be constructed from the fitted moment conditions. The market prices of diffusion and jump risks are calibrated by minimizing the pricing errors between a model-implied yield curve and a target yield curve. The time series estimation of short rate suggests that the jump augmentation is highly significant and that the pure diffusion process is strongly rejected. The cross sectional evidence indicates that the jump-diffusion yield curves are much more flexible in reducing the pricing errors. Comparative statics are also consistent with economic intuitions, with the yield curves being more sensitive to factor risk and risk premium parameters.

3.1 Introduction

The famous square-root model by Cox, Ingersoll, and Ross (1985b) (CIR), although appealing in its general equilibrium nature and closed-form solution, is widely rejected

---

[1]The material of this paper also appears in Zhou (1999b).

in empirical applications. Brown and Dybvig (1986) first documented the difficulty in fitting the CIR model to US interest rates, especially for the volatile period 1978-1982. More recently, rigorous specification tests all rejected the square-root model, using the historical short rate data (Aït-Sahalia 1996b, Conley et al. 1997a, Gallant and Tauchen 1998b).[2] In time series perspective, underfitting the volatility parameter is the major cause for rejecting the short rate dynamics. In cross-sectional perspective, the pricing error of term structure is large when the understated volatility parameter reduces the flexibility of yield curves. As argued by Zhou (1999a) in a Monte Carlo study, square-root process can not model both the mean persistence and the variance persistence simultaneously, due to the stationarity condition and the given level of average interest rate.

Consequently, efforts to modify the square-root model largely concentrate on more flexible specifications of the volatility function. It is clear that the CIR model is just one special case of so-called linear CEV (constant elasticity of volatility) specification, where elasticity equals one half. Recent comparative studies (Chan, Karolyi, Longstaff, and Sanders 1992, Conley et al. 1997a, Tauchen 1996) found that elasticity close to one is marginally acceptable. Alternatively, one can free the specification of the diffusion function and estimate the volatility process nonparametrically (Aït-Sahalia 1996a, Stanton 1997). This approach exploits the long run invariant distribution of the short rate dynamics. Their empirical results suggest that the square-root model fits reasonably well for the medium range of interest rates; but the model entirely misses the nonlinear volatility feature at either the low end or the high end of short-rate levels. Another pertinent approach is to introduce an

---

[2] The bivariate extensions of CIR specification (Gibbons and Ramaswamy 1993, Chen and Scott 1993, Pearson and Sun 1994) also meet with poor empirical performance. Duffie and Singleton (1997) found favorable evidence for a two factor CIR model with serially-correlated error structure. Dai and Singleton (2000) estimated a three-factor affine model similar to Chen (1996) and passed the specification test. These two studies only use the swap yields data from 1987 to 1996. When fitting the same specifications to US Treasury yields from 1964 to 1995, they are strongly rejected (Bansal, Hsieh, and Shen 1998).

46

unobserved stochastic volatility factor into the diffusion function, and finds considerable improvement in the specification test (Andersen and Lund 1998, Andersen and Lund 1997). Since the dimension and functional form of the latent factor are not very restrictive, the stochastic volatility model is cable of capturing both the long memory and the asymmetric features of the short rate process. The more recent approach in literature is the jump-diffusion interest rate (Das 1998), with a constant diffusion function (Vasicek 1977) and an independent jump term. Under this setting, the unconditional volatility structure is enriched, while the conditional volatility persistence is still absent.

This essay follows the jump-diffusion approach in term structure literature (Das 1998) and allows for time-varying volatility persistence. The jump-diffusion approach in equity return literature (for example, see Merton (1976) and Bates (1996)) is not directly applicable to the interest rate dynamics. In those models, the timing of the jumps does not affect the evolution of the return sample path, since the observed asset return does not feed back into the local mean and variance functions. In other words, a jump may shift the asset price up and down, but it has no impact on the trend or fluctuation of the returns, which is the intrinsic feature of any lognormal asset pricing model (Ingersoll 1987). But for interest rates data, the time varying mean reversion and volatility persistence are strongly related to the short rate level. Also the CIR model is a better starting point to model the short rate, because it already incorporates the time varying volatility, it automatically satisfies the nonnegativity constraint, and it has inherent nonsymmetry and fatter right tail. The conditional volatility persistence can be enhanced by introducing jumps, without violating the stationarity assumption of the square-root diffusion component. Since the jump effects are directly observable through the state variable, a variety of jump-diffusion specifications can sustain closed-form or near closed form solutions to the term struc-

47

ture. This is a clear advantage compared to the other modeling strategies, requiring numerical simulation to calculate the derivative prices. Since the choices of jump-rate and jump-size function are flexible, the fitting of cross sectional term structure may produce smaller pricing error.

Conceptually, the jump-diffusion approach is more similar to the stochastic volatility strategy, but with better economic interpretation. The information arrivals on the financial markets can be either gradual, small perturbations or sudden, large shocks (Merton 1976). The examples of microeconomic information flow include the temporary imbalances of supply and demand, the change of economic outlook among a small number of market participants, or earning reports from several large companies within a week. On the other hand, the Federal Reserve Board may adjust the discount rate by a quarter percent, the OPEC oil agreement can produce a supply shock, the Asian or Russian financial crises may affect assets demand across the world. These macroeconomic information shocks may completely alter the market perception of the economic fundamentals.[3] By the very nature, they arrive only randomly at certain points of time, and their impacts on the market movement are in large, discrete sizes. Those discrete-size information shocks are reflected in financial market as data outliers. Many times the continuous sample path distributions (e.g., the CIR model) fail to explain the extraordinary volatility structure, since the empirical data is typically asymmetric and has a fat tail. However, the occasional volatility cluster is *the* feature of asset returns, and the equilibrium no-arbitrage pricing theory requires that the empirical methods be able to capture and explain both the smooth and the rough periods of the financial markets. A mixture model of a continuous Brownian motion and a discrete Poisson jump may be capable of capturing the real time evolution of most financial asset prices (Ingersoll 1987).

---

[3]Alternatively, one can model these discrete changes of interest rates dynamics as regime shifts and solve the entire term structure consistently by an equilibrium pricing technique (Bansal and Zhou 1999).

48

The innovation of this study is to generate the parametric moment conditions and to construct an efficient MWNLS-JD estimator with diagnostics. Maximum Likelihood Estimation is available only for a very restricted class of jump-diffusion models (Lo 1988). Our method differs with the infinitesimal generator of Hansen and Scheinkman (1995) in that it exploits the conditional information, does not relies on simulations as do Duffie and Singleton (1993), uses model-dependent moments instead of data-dependent moments (Gallant and Tauchen 1996b), generalizes to an arbitrary number of moments rather than only to conditional mean and variance (Fisher and Gilles 1996), and has faster solutions for both estimation and pricing in comparison with the nonparametric approach (Aït-Sahalia 1996a). As shown below, our method reduces a complicated task of solving a stochastic differential equation (SDE) to a simple matrix solution of an ordinary differential equation (ODE) system. The solution becomes a linear least square problem with nonlinear parameter constraints, or at most a multivariate nonlinear least square problem. The computational burden is reduced to only few minutes. In the literature, those studies that are closely related are the nonparametric regression method with stochastic Taylor series approximation (Stanton 1997) and the generalized eigenvalue-eigenfunction method with orthogonal series approximation (Conley et al. 1997b). The distinct feature of my essay is the maximum exploitation of the parametric information, which is contained in the drift, diffusion, and jump specifications.

The body of this paper is organized as follows: Section 3.2 applies the martingale pricing technique to the jump-diffusion term structure and derives closed-form solutions to bond prices for a special class of interest rate models; Section 3.3 characterizes the first four conditional moments and constructs an efficient estimator with diagnostics testing; Section 3.4 implements the jump-diffusion term structure empirically and studies the comparative statics of the fitted yield curves; and Section 3.5

49

concludes.

3.2 Term Structure in a Jump-Diffusion Economy

The classical martingale pricing result can be extended to the jump-diffusion environment by appropriately adjusting the non-central tendency of the jump process. Given an underlying process of the instantaneous interest rate, the crucial step is to correctly choose the market-price-of-risk process, which should be supported by the underlying equilibrium economy. Once this is done, the solution to discount bond prices falls out naturally.

3.2.1 Short Rate Process and No-Arbitrage Pricing

Suppose that the evolution of short interest rates is governed by a square-root jump-diffusion process

$$dr_t = \kappa(\theta - r_t)dt + \sigma\sqrt{r_t}dW_t + J(\cdot)dN(\rho(\cdot)t), \qquad (3.1)$$

where W_t is a standard Brownian motion, $N(\rho(\cdot)t)$ is a Poisson jump driving process with an intensity function $\rho(\cdot)$. For the square-root part, κ is the mean reversion parameter, θ is the long-run mean parameter, and σ is the local variance parameter. For the jump part, $J(\cdot)$ is the jump-size function. Both the jump-rate and jump-size can be constants, functions of state variables, or other independent random variables. The state-dependent drift and diffusion functions feature time-varying mean reversion and volatility persistence. If a jump occurs at a high interest rate, both the local variance and the mean reversion are discretely large, and the volatility clustering is enhanced. On the other hand, if a jump occurs at a lower interest rate level, the local variance is just slightly increased, while the mean reversion to the center is even reduced. The intrinsic asymmetric responses of the interest rate level and volatility

50

to the discrete jumps are indeed driving our empirical yield curves to fit the term structure.

Certain regularity assumptions need to be imposed. The restrictions on the square-root part is offered by Feller (1951),

Assumption 1 (Diffusion) $\kappa > 0$, $\theta > 0$, and $\sigma^2 \leq 2\kappa\theta$,

which ensures that the diffusion is in the domain $(0, \infty)$. Zero is not accessible unless as a starting value, and the process never explodes to infinity. It also implies a noncentral chi-square transitional distribution and a gamma steady-state marginal distribution.

The next assumption serves to exclude the technical arbitrage from tampering with jump information.

Assumption 2 (Jump) $\rho(\cdot) \in \mathcal{F}_t^-$ and $J(\cdot) \in \mathcal{F}_t^-$,

which says that both the jump-intensity and the jump-size functions at time t should only depend on the left limit of $r(t)$ to preserve the Markov property. In other words, if one knows the information of the exact jump timing and jump-size at an instant before a jump occurs, one can make an arbitrarily large profit with certainty.

Apart from above assumptions about the pure square-root diffusion and the pure Poisson jump processes, additional restrictions should be applied when the two parts are put together to guarantee closed-form solutions for both conditional moments and bond prices.

Assumption 3 (Conformity and Pricing) $0 \leq \rho(\cdot) < \infty$, $-r_{t_-} < J(\cdot) < \infty$.

This assumption implies that when a jump occurs, the regularity conditions of the square-root diffusion are always satisfied (nonnegativity, not accessible to zero, and non-explosion). In addition, all the moments of the jump term will have closed form

51

solutions, since both $\rho(\cdot)$ and $J(\cdot)$ are not state-dependent, or at most one of them is linearly dependent on the state variable. When the jump part is shut down ($\rho(\cdot) = 0$), it reduces to square-root diffusion.

A major proposition following Assumptions 1-3 is that the jump-diffusion process (3.1) is well-defined and can be constructed properly. To see this, consider the time period from zero until the first jump occurs. Before the jump, the process evolves as a square-root diffusion process. By the standard change of measure and change of time techniques, one can recover the Browning motion. As long as a standard Brownian motion can be constructed, the above argument can be reverted to construct the diffusion part. When the jump occurs, the process is reinitialized at a new starting value, which is within the domain of the diffusion part. Assumption 1 ensures that the diffusion is Markov; Assumption 2 ensures that the jump is Markov; and Assumption 3 ensures that the combined jump-diffusion remains a well-defined Markov process. This construction procedure extends to any point in time. The jump-diffusion process defined as above has a unique strong solution.

To price a discount bond or any interest rate derivative, one requires a well justified stochastic discount factor. It is a well-known fact that the market-price-of-risk processes must be proportional to the standard deviation of the normalized risk factor. The corresponding pricing kernel is specified as

$$\frac{d\pi(t)}{\pi(t)} = -r_t dt - \frac{\lambda_W}{\sigma}\sqrt{r_t}dW_t - \lambda_J[dN(\rho(\cdot)t) - \rho(\cdot)dt], \tag{3.2}$$

where λ_W is the diffusion risk premium parameter and λ_J is the jump risk premium parameter. The functional form of $\lambda_W/\sigma\sqrt{r_t}$ is in accordance with the square-root literature (Cox et al. 1985a), and the choice of parameter λ_J extends the constant volatility specification of Vasicek (1977) to the cases of constant or state-dependent jump-rate with independent jump-size. The instantaneous expectation of jump pre-

52

mium is proportional to instantaneous jump risk. These specifications are supported by the underlying equilibrium economy.[4] In fact, the decomposition of the total risk premium into the diffusion part and the jump part are fully identifiable, since the former will be time-varying in proportion to the square-root of short rate level while the latter will be paid constantly or linearly over time. Among the four specifications considered below, model 1 (no jump), model 2 (state-independent jump), and model 3 (state-dependent jump-rate) are fully backed by underlying economic equilibrium, but model 4 (state-dependent jump-size) is only justifiable under the no-arbitrage argument.

Given the short rate process, the bond process can be spelled out. The price of a discount bond $P(t) = P(r_t, t, T)$ at time t with $T - t$ maturity is conjectured to be log-linear $P(r_t, T - t) = A(T - t) \exp\{-B(T - t)r_t\}$ with the boundary condition $P(r_T, T, T) = 1$. Itô's formula delivers an instant bond return process

$$\frac{dP(t)}{P(t)} = \mu_P(t)dt + \sigma_P(t)dW_t + J_P(t)dN(\rho(\cdot)t),\tag{3.3}$$

where $\mu_P(t)$, $\sigma_P(t)$, and $J_P(t)$ are the instantaneous drift, diffusion, and jump functions given by

$$\mu_P(t) = \frac{P_r\kappa(\theta - r) + \frac{1}{2}P_{rr}\sigma^2 r + P_t}{P},$$

$$\sigma_P(t) = \frac{P_r\sigma\sqrt{r}}{P},$$

$$J_P(t) = \frac{P(r + J(\cdot), T - t) - P(r, T - t)}{P}.$$

To ease the notational burden, the dependence on time t is suppressed. Applying the

---

[4]A counter example is given by Cox et al. (1985a), in which a linear risk premium is specified by the no-arbitrage approach while the underlying factor is a square-root process. It implies that the equilibrium bond return over-pays a constant premium, even when the interest rate is zero, which becomes a riskless arbitrage opportunity.

martingale pricing result (Appendix A), we can show that

$$\mu_P(t) = r_t + \frac{\lambda_W}{\sigma}\sqrt{r_t}\sigma_P(t) + \rho(\cdot)(\lambda_J - 1)E[J_P(t)], \qquad (3.4)$$

which simply says that the instantaneous bond return should be equal to the "risk free" rate plus total risk premium. The risk premium from either the diffusion factor or the jump factor, is simply the product of the bond's risk exposure to that factor and the factor premium per unit of risk. Such an interpretation of the equilibrium pricing condition is in line with the Arbitrage Pricing Theory (APT) (Ingersoll 1987). Notice that the noncentral tendency of the jump process $\rho(\cdot)E[J_P(t)]$ needs to be adjusted in the jump risk premium. This tendency is indeed the "drift" of the jump process. One thus reaches the fundamental valuation equation for bond pricing

$$\frac{1}{2}\sigma^2 r P_{rr} + \kappa(\theta - r)P_r + P_t - rP - \lambda_W r P_r + \rho(\cdot)(1 - \lambda_J)P E_{J(\cdot)}[e^{-BJ(\cdot)} - 1] = 0, \quad (3.5)$$

It is clear that particular restrictions on the jump term are driven by the tractability of $E_{J(\cdot)}[e^{-BJ(\cdot)} - 1]$. To make the expectation solvable without imposing additional restrictions on the ordinary differential equation system, the choices of jump-size distribution are only normal and uniform (or constant as a degenerate case). To guarantee the nonnegativity of the jump-diffusion interest rate, only the uniform random variable can enforce no-arbitrage condition. It is also attractive to have an upper bound on the jump-size, which is meaningful in a large stable economy. A more interesting model would allow the short-rate level to feed back into the jump-rate and/or jump-size process, affecting more than the mean reversion and conditional volatility. To reduce the redundancy, we only let the jump-rate or the jump-size be state-dependent, but not both. However, since the jumps still feed into the drift, diffusion, and jump-rate functions through the short rate level, the amplitude of jumps can still be identified by the intensified mean-reversion, volatility clustering, and/or time-varying jump probability.

54

3.2.2 Jump Specifications and Closed-Form Solutions

Various jump-diffusion processes are outlined here to facilitate the empirical comparison. Model 1, 2, and 3 bellow are fully supported by the underlying equilibrium economy, with closed-form solutions. However, model 4 is only justifiable by a no-arbitrage argument, with an approximate solution.

Model 1, Square-Root Diffusion. If the jump term is shut down by letting $\rho = 0$, the model reduces to the well-known CIR specification

$$dr_t = \kappa(\theta - r_t)dt + \sigma\sqrt{r_t}dW_t,$$

and the pricing partial differential-difference equation (3.4) is reduced to

$$\frac{1}{2}B^2\sigma^2 r - B\kappa\theta + B\kappa r - \frac{A'}{A} + B'r - r + \lambda_W Br = 0.$$

The exact solution is already known (Cox et al. 1985a). Despite its general equilibrium feature and tractability for derivative pricing, the majority of empirical studies have rejected the model. We use CIR specification as a benchmark, to measure whether the jump augmentation can improve the term structure fitting. For the purpose of comparison, we retain the solutions in ordinary differential equation form,

$$B' = 1 - \frac{1}{2}\sigma^2 B^2 - (\lambda_W + \kappa)B, \tag{3.6}$$

$$\frac{A'}{A} = -\kappa\theta B. \tag{3.7}$$

The Ricatti type ODEs can be easily solved numerically with the Runge-Kutta method (Press, Teukolsky, Vetterling, and Flanney 1996).

Model 2, Square-Root Diffusion with Independent Jump. The jump term is driven by a Poisson process with a constant rate ρ, and the jump-size J follows a uniform distribution with constant lower bound a and upper bound b

$$dr_t = \kappa(\theta - r_t)dt + \sigma\sqrt{r_t}dW_t + JdN(\rho t).$$

55

The specification is consistent with the economic reality that any potential interest-rate jump should be bounded. The partial differential-difference equation (PDDE) boils down to

$$\frac{1}{2}B^2\sigma^2 r - B\kappa\theta + B\kappa r - \frac{A'}{A} + B'r - r + \lambda_W Br$$

$$+\rho(1 - \lambda_J)(\exp\{\frac{e^{-Bb} - e^{-Ba}}{-B(b-a)}\} - 1) = 0,$$

and the solution ODEs are

$$B' = 1 - \frac{1}{2}\sigma^2 B^2 - (\lambda_W + \kappa)B, \tag{3.8}$$

$$\frac{A'}{A} = -\kappa\theta B + \rho(1 - \lambda_J)(\exp\{\frac{e^{-Bb} - e^{-Ba}}{-B(b-a)}\} - 1). \tag{3.9}$$

If the jump does not occur ($\rho = 0$), the pricing formula simply reduces to the square-root model. If the market price for jump risk $\lambda_J = 1$, jump risk is "neutral" and does not contribute to the determination of the bond price, but the parameter estimates of the short rate part are still affected by adding the jump term. After de-trending the jump risk, the appropriate measure of risk premium should be $\lambda_J - 1$ instead of λ_J. If $\lambda_J = 1$, the representative agent is risk-neutral to the centered jump risk, just as is $\lambda_W = 0$ when the agent is risk-neutral to the diffusion risk.

Model 3, Square-Root Diffusion with State-Dependent Jump-Rate.[5]

The jump-size distribution is still the uniform as in model 2, while the jump rate is specified as a linear function of the state variable $\rho(\cdot) = \rho_0 + \rho_1 r_t$,

$$dr_t = \kappa(\theta - r_t)dt + \sigma\sqrt{r_t}dW_t + JdN((\rho_0 + \rho_1 r_t)t).$$

This is a potentially more interesting model, since it allows mean reversion in the jump timing, if $\rho_0 > 0$ and $\rho_1 < 0$. It also implies that the short rate is more

---

[5] Duffie, Pan, and Singleton (1999) discuss a similar class of multivariate affine jump-diffusion processes, Chernov, Gallant, Ghysels, and Tauchen (1999) extend it a non-affine class of models, and Drost, Nijman, and Werker (1998) emphasize on testing the GARCH jump-diffusion specification.

leptokurtotic at lower levels, but looks more Gaussian at higher levels. The pricing PDDE now becomes

$$\frac{1}{2}B^2\sigma^2 r - B\kappa\theta + B\kappa r - \frac{A'}{A} + B'r - r + \lambda_W Br$$

$$+(\rho_0 + \rho_1 r_t)(1 - \lambda_J)(\exp\{\frac{e^{-Bb} - e^{-Ba}}{-B(b-a)}\} - 1) = 0,$$

and the solution ODES are

$$B' = 1 - \frac{1}{2}\sigma^2 B^2 - (\lambda_W + \kappa)B - \rho_1(1 - \lambda_J)(\exp\{\frac{e^{-Bb} - e^{-Ba}}{-B(b-a)}\} - 1), \quad (3.10)$$

$$\frac{A'}{A} = -\kappa\theta B + \rho_0(1 - \lambda_J)(\exp\{\frac{e^{-Bb} - e^{-Ba}}{-B(b-a)}\} - 1). \quad (3.11)$$

One can easily see the hierarchical relationship between model 1, 2, and 3. It is interesting that even though the jump-rate is state-dependent, the pricing kernel laid out in section 3.2.1 is still compatible with underlying equilibrium, because the market price of jump risk has already incorporated the contribution of the state-dependent jump-rate $\Lambda_J[dN(\rho(\cdot)t) - \rho(\cdot)dt]$.

Model 4, Square-Root Diffusion with State-Dependent Jump-Size. Alternatively, one can let the jump-size be a linear deterministic function of the state variable $J = J_0 - r_t$. This specification is parsimonious and guarantees closed-form solutions of the conditional moments (see Section 3.3.1 for details). To retain a near closed-form pricing solution, we have to restrict the jump rate to a constant (ρ),

$$dr_t = \kappa(\theta - r_t)dt + \sigma\sqrt{r_t}dW_t + (J_0 - r_t)dN(\rho t).$$

A tricky point in solving the pricing PDDE (3.4) is to approximate the deterministic difference term by a first-order Taylor expansion

$$e^{-B(J_0 - r_t)} - 1 \approx -B(J_0 - r_t),$$

57

which is extremely accurate since the difference term is very close to zero. Hence the pricing PDDE is approximately

$$\frac{1}{2}B^2\sigma^2 r - B\kappa\theta + B\kappa r - \frac{A'}{A} + B'r - r + \lambda_W Br - \rho(1 - \lambda_J)B(J_0 - r_t) \approx 0,$$

and the solution ODEs are

$$B' \approx 1 - \frac{1}{2}\sigma^2 B^2 - (\lambda_W + \kappa)B - \rho(1 - \lambda_J)B, \qquad (3.12)$$

$$\frac{A'}{A} \approx -\kappa\theta B - \rho(1 - \lambda_J)J_0 B. \qquad (3.13)$$

It must be pointed out that the pricing kernel in section 3.2.1 is not compatible with economic equilibrium implied by this short rate specification, since the market price of jump risk $\Lambda_J[dN(\rho t) - \rho dt]$ is not proportional to the centered jump risk $(J_0 - r_t)[dN(\rho t) - \rho dt]$. Nevertheless, the approximate solution is guaranteed by the simple no-arbitrage rule.

Our empirical study will cover each of the four models to illustrate the improvement in matching the yield curve when discrete jumps are allowed. Keep in mind that the entire term structure may depend on more than one factor. What we illustrate here is only an effort to better capture the the short-end dynamics of the interest rate.

3.3 Moment Generator and Efficient Estimation

The innovation of this paper is to characterize all the conditional moments of a square-root jump-diffusion process in a convenient matrix form. In particular, the martingale property of the Generalized Itô process is exploited to derive the lower-order four conditional moments. Only conditional mean and variance will not be adequate for identifying the jump impact. An efficient estimation strategy is also constructed, together with several specification tests.

58

3.3.1 A Characterization of Conditional Moment

Let's start with an example on how to derive the conditional mean of the square-root independent jump process (model 2):

$$dr_t = \kappa(\theta - r_t)dt + \sigma\sqrt{r_t}dW_t + JdN(\rho t).$$

Applying Itô's Lemma to r_s for $s \geq t$ as a function of r_t, and taking conditional expectation with respect to t, we have

$$E_t(r_s) = r_t + E_t \int_t^s [\kappa(\theta - r_u) + \rho E_J(J)]du.$$

Notice that the conditional expectation of the local martingale $\int_t^s \sigma\sqrt{r_u}du$ is equal to zero. Interchanging the expectation operator with the integration operator, and taking derivative with respect to the time s on both sides, we arrive at

$$\frac{dE_t(r_s)}{ds} = \kappa\theta + \rho E_J(J) - \kappa E_t(r_s),$$

which is a linear first-order differential equation. Its solution is given by standard text books. Using the boundary condition $E_t(r_t) = r_t$, it is straightforward to show that

$$E_t(r_T) = r_t e^{-\kappa(T-t)} + \frac{\theta\kappa + \rho E_J(J)}{\kappa}(1 - e^{-\kappa(T-t)}).$$

For reassurance, notice that when the jump term turns off ($\rho = 0$), it reduces to the conditional mean of Cox et al. (1985a)

$$E_t(r_T) = r_t e^{-\kappa(T-t)} + \theta(1 - e^{-\kappa(T-t)}).$$

This strategy is well-known in literature (Kloeden and Platen 1992, Kushner and Dupuis 1992, Fisher and Gilles 1996), and we simply extend it to an arbitrary vector of moments.

59

Now let's focus on the first four conditional moments of the special jump-diffusion processes proposed earlier

$$dr_t = \kappa(\theta - r_t)dt + \sigma\sqrt{r_t}dW_t + J(\cdot)dN(\rho(\cdot)t).$$

Let $R_s = [r_s, r_s^2, r_s^3, r_s^4]'$ be the column vector of the first four powers of r_s for some $s \geq t$. An extended version of Itô's formula (see Kushner (1967) and Protter (1992)) delivers $E_t(R_s) = [E_t(r_s), E_t(r_s^2), E_t(r_s^3), E_t(r_s^4)]'$ in a matrix form

$$\begin{bmatrix} E_t(r_s) \\ E_t(r_s^2) \\ E_t(r_s^3) \\ E_t(r_s^4) \end{bmatrix} = \begin{bmatrix} r_t + E_t \int_t^s \{\kappa(\theta - r_u) + \rho(\cdot)E[(r_u + J(\cdot)) - r_u]\}du \\ r_t^2 + E_t \int_t^s \{2r_u\kappa(\theta - r_u) + \sigma^2 r_u + \rho(\cdot)E[(r_u + J(\cdot))^2 - r_u^2]\}du \\ r_t^3 + E_t \int_t^s \{3r_u^2\kappa(\theta - r_u) + 3r_u\sigma^2 r_u + \rho(\cdot)E[(r_u + J(\cdot))^3 - r_u^3]\}du \\ r_t^4 + E_t \int_t^s \{4r_u^3\kappa(\theta - r_u) + 6r_u^2\sigma^2 r_u + \rho(\cdot)E[(r_u + J(\cdot))^4 - r_u^4]\}du \end{bmatrix}.$$

So the conditional moment is simply the realization of the four powers of r_t at the initial date plus the expected Riemann integral of the stochastic differential generated by Itô's formula. If one observes the continuous time record, these moments of any jump-diffusion process can be calculated directly by numerical integration. Since the data is only available in discrete samples or since the continuous time record is contaminated by institution and microstructure noises, the main challenge remains to tackle the integration without relying on the actual sample path. For instance, Stanton (1997) applied a stochastic Taylor series approximation to the integral and estimated by a nonparametric kernel regression approach. Conley et al. (1997b) adopted an orthogonal series approximation in a general eigenvalue-eigenfunction framework. Fisher and Gilles (1996) proposed a quasi-maximum likelihood estimator based on closed-form conditional mean and variance function for the affine diffusion process. The approach developed below is more general for any jump-diffusion process, and more straightforward for extending to any number of moments.

Taking derivatives of $E_t(R_s)$ with respect to the future time s and writing in

60

matrix form, we have a system of ordinary differential equations

$$\frac{dE_t(R_s)}{ds} = AE_t(R_s) + g,$$ (3.14)

where g is a 4×1 vector and A is a 4×4 lower-triangle matrix. They are nonlinear functions of the structural parameter vector β, which is yet to be specified for a particular choice of the jump-rate and jump-size functions. Since any random variable inside jump-size function $J(\cdot)$ is defined as independent with both the Brownian motion and the Poisson jump process,[6] the condition expectation $E_t(\cdot)$ and $E_J(\cdot)$ can be exchanged by applying the Law of Iterated Expectations.

This is a linear first order differential equation system. It is non-homogeneous in having a forcing function g. The solution to the homogeneous part is fully determined by the fundamental matrix A. Since all the coefficients are not dependent on the time t, the system is temporally homogeneous. Its solution comprises a fundamental solution plus a particular solution. Using the boundary condition $E_t(R_t) = R_t$, we have a characterization of the conditional moments

$$
\begin{aligned}
E_t(R_T) &= e^{(T-t)A}R_t + \int_t^T e^{(T-s)A}g\,ds \\
&= e^{(T-t)A}R_t + e^{TA}A^{-1}(e^{-tA} - e^{-TA})g,
\end{aligned}
$$ (3.15)

where e^A is the matrix exponential of A.[7] The last step in equation (3.15) depends on the fact that the vector g is time-homogeneous and also exploits the communicate and

---

[6]The first four moments of $J \sim \text{Uniform}[a, b]$, are respectively $E_J(J) = \frac{(b^2-a^2)}{2(b-a)}$, $E_J(J^2) = \frac{(b^3-a^3)}{3(b-a)}$, $E_J(J^3) = \frac{(b^4-a^4)}{4(b-a)}$, and $E_J(J^4) = \frac{(b^5-a^5)}{5(b-a)}$. They will be used later for calculating the conditional moments of the short rate

[7]Matrix exponential is defined as

$$e^A = \sum_{k=0}^{\infty} \frac{A^k}{k!},$$

which is different from the exponential of each element in a matrix A.

61

derivative properties of matrix exponentials.[8] Since the data are discretely sampled, we can always set $t = 0$ and $T = 1$ to solve the system explicitly as

$$E_0(R_1) = e^A R_0 + e^A A^{-1}(I - e^{-A})g,$$

where I is a 4×4 identity matrix. This generic solution requires that matrix A is nonsingular, which is guaranteed if the mean reversion parameter κ in the drift function of r_t is nonzero.

The generic result can be written as a vector autoregressive function $E_t(R_{t+1}) = DR_t + C$, or

$$
\begin{bmatrix}
E_t(r_{t+1}) \\
E_t(r_{t+1}^2) \\
E_t(r_{t+1}^3) \\
E_t(r_{t+1}^4)
\end{bmatrix}
=
\begin{bmatrix}
d_{11} & 0 & 0 & 0 \\
d_{21} & d_{22} & 0 & 0 \\
d_{31} & d_{32} & d_{33} & 0 \\
d_{41} & d_{42} & d_{43} & d_{44}
\end{bmatrix}
\begin{bmatrix}
r_t \\
r_t^2 \\
r_t^3 \\
r_t^4
\end{bmatrix}
+
\begin{bmatrix}
c_1 \\
c_2 \\
c_3 \\
c_4
\end{bmatrix},
\tag{3.16}
$$

where the constant matrix $D = e^A$ and vector $C = e^A A^{-1}(I - e^{-A})g$ are nonlinear functions of the structural parameter vector β, which is yet to be specified.[9] Solutions for the vector ordinary differential system of various jump-diffusion specifications (models 1, 2, 3, and 4) are listed in Appendix B.

3.3.2 Estimation and Inference

The above method can be used to generate conditional moments up to an arbitrary order. Arguably, if the stochastic process has bounded conditional moments up to any order as our jump-diffusion specification (3.1), fitting these moments will approach the full recover of the distribution information. Given a particular set of moment conditions, a judicious choice of an estimation strategy can achieve the prescribed

---

[8]If matrices A and B are commutable, i.e. $AB = BA$, then $e^{A+B} = e^A e^B$. The formula for matrix derivative is $de^{tA}/dt = Ae^{tA}$.

[9]D and C can be calculated by the matrix language in MATLAB or can be derived exactly in a FORTRAN program, since the exponential and inverse of an upper or lower-triangular matrix have closed form solutions.

efficiency. The Multivariate Weighted Nonlinear Least Square for Jump-Diffusion (MWNLS-JD) estimator constructed here is in the spirit of Gallant (1987). The consistency and asymptotic normality results are standard; we simply verify several critical conditions and restate other technical assumptions. We also discuss a Lagrange multiplier test and a conditional moment test. Our focuses are on constructing the estimator and making inferences.[10]

Consistency

The moment condition (3.16) generated by Itô's formula in Section 3.3.1 justifies the nonlinear regression hypothesis

$$R_{t+1} = DR_t + C + U_{t+1},$$

where $U_{t+1} = [u_{1t+1}, u_{2t+1}, u_{3t+1}, u_{4t+1}]'$ is a vector of errors. With the data generating process (3.1) and Assumption 1-3 of section 3.1, it clearly follows that $0 < R_t < \infty$; hence $E|R_t| < \infty$. This is equivalent to the strict stationary and ergodic condition. The nonlinear functions $D = e^A$ and $C = e^A A^{-1}(I - e^{-A})g$ derived in Section 3.3.1 are obviously continuous. The parameter space $\mathcal{B}$ for $\beta \in \mathcal{B} \subset \mathcal{R}^l$ is usually assumed to be compact. Thus the consistency result is primarily driven by the identification condition

Condition 1 (Identification) *There exists $\beta_0 \in \mathcal{B} \subset \mathcal{R}^l$ such that $E_t(U_{t+1}) = 0$.*

This condition is indeed the result of Section 3.3.1. The square-root diffusion part is well-identified. Since jumps are fed back into the drift and diffusion through the short-rate level, the intensified mean reversion and volatility clustering can provide

---

[10]Our estimator is indeed a Classical Method of Moments. If all the moments are finite, parameter estimation by matching the analytical moments to an arbitrary order will approach the efficiency of maximum likelihood estimation (Gallant and Tauchen 1998a). Singleton (1999) also discusses the asymptotic efficiency of several estimators for the affine diffusion processes by exploiting the Conditional Characteristic Function.

the identification for the discrete jumps. In fact, the stationarity assumption of the square-root process ($\sigma^2 \leq 2\kappa\theta$) always restricts the empirical estimation to underfit the conditional volatility.[11] Our proposed model (3.1) circumvents this difficulty by jump-augmenting the short-rate level to enhance the conditional volatility, without pushing the diffusion parameter into a non-stationary region.

The nonlinear least-square estimator solves the minimization problem:

$$\min_{\beta \in \mathcal{B}} Q_T(\beta) \equiv \frac{1}{T} \sum_{t=1}^{T-1} \frac{1}{2} U_{t+1}(R_{t+1}, R_t, \beta)' W_T(R_{t+1}, R_t, \bar{\beta})^{-1} U_{t+1}(R_{t+1}, R_t, \beta), \quad (3.17)$$

where $\bar{\beta}$ is some consistent estimator of β, and the weighting matrix $W_T(\cdot, \bar{\beta})$ is a 4×4 symmetric, positive definite matrix with probability one, usually constructed by

$$W_T(\bar{\beta}) = \frac{1}{T} \sum_{t=1}^{T-1} U_{t+1}(\bar{\beta}) U_{t+1}(\bar{\beta})'. \quad (3.18)$$

Either a two-step or an iterated estimator may apply. If R_t is stationary, it is a typical assumption that $W_T(\bar{\beta}) \xrightarrow{a.s.} W_0$, where W_0 is a symmetric positive definite matrix. The last building block of the consistency result is a uniform law of large numbers. Assumptions 1-3 of section 3.2 deliver an invariant probability measure $P(R_{t+1}, R_t|\beta)$, which is Markovian and stationary. Let $Q(\beta) = \lim_{T \to \infty} \int Q_T(\beta) dP(\cdot|\beta)$ be the population limit of the objective function, then a standard proof gives the result $Q_T(\beta) \xrightarrow{a.s.} Q(\beta)$ uniformly on $\mathcal{B}$. In addition, assume that $Q(\beta)$ is continuous on $\mathcal{B}$. Now we have the consistency result[12]

$$\hat{\beta} \xrightarrow{a.s.} \beta_0, \quad (3.19)$$

---

[11]See Zhou (1999a) for Monte Carlo evidence and Gallant and Tauchen (1998b) for empirical evidence.

[12]The maximum likelihood estimator of the uniform distribution parameters is not consistent, if one specifies the distribution as $J \sim$ Uniform (a, b) instead of $J \sim$ Uniform $[a, b]$. However, this is not a concern for the method-of-moments estimator, because the objective function is smooth around the true parameters.

64

where $\hat{\beta} = \arg \min_{\beta \in \mathcal{B}} Q_T(\beta)$.

Asymptotic Normality

The following condition greatly simplifies the derivation of the asymptotic distribution and the construction of specification tests:

Condition 2 (No Serial Correlation) *For all t, $E_t(U_t U'_{t+j}) = 0$, for any $j \geq 1$.*

This condition is satisfied because our derived moment conditions are essentially martingale difference sequence (MDS). In other words, U_t is already in the information set at time t, therefore $E_t(U_t U'_{t+j}) = U_t E_t(U_{t+j})' = 0$ by the iterated law of expectations.

With a little more effort, one can prove that the scores

$$s_{t+1}(\cdot, \beta) = \frac{\partial}{\partial \beta} U_{t+1}(\cdot, \beta)' W_T(\cdot, \bar{\beta})^{-1} U_{t+1}(\cdot, \beta)$$

are not serially correlated (Wooldridge 1994). Assume that β_0 is an interior point of a convex set $\mathcal{B}$. Define the expected outer product of the scores as

$$J = \int s_{t+1}(\cdot, \beta) s_{t+1}(\cdot, \beta)' dP(\cdot | \beta)$$

and its empirical counterpart as

$$J_T = \frac{1}{T} \sum_{t=1}^{T-1} s_{t+1}(\cdot, \hat{\beta}) s_{t+1}(\cdot, \hat{\beta})'.$$

Standard argument leads to

$$\sqrt{T} \frac{\partial}{\partial \beta} Q_T(\beta_0) \xrightarrow{D} N(0, J_0),$$

where J_0 is J evaluated at the true parameter. It is a positive semi-definite matrix. The Hessian is defined as

$$A = \int \frac{\partial}{\partial \beta} U_{t+1}(\cdot, \beta)' W_T(\cdot, \bar{\beta})^{-1} \frac{\partial}{\partial \beta} U_{t+1}(\cdot, \beta) dP(\cdot | \beta)$$

65

and its empirical estimate as

$$A_T = \frac{1}{T} \sum_{t=1}^{T-1} \frac{\partial}{\partial \beta} U_{t+1}(\cdot, \hat{\beta})' W_T(\cdot, \bar{\beta})^{-1} \frac{\partial}{\partial \beta} U_{t+1}(\cdot, \hat{\beta}).$$

Assume that $A_T(\beta) \xrightarrow{P} A$ uniformly in a neighborhood of β_0 and A is continuous at β_0. Let A_0 be the Hessian evaluated at the true parameter; we arrive at the asymptotic normality result[13]

$$\sqrt{T}(\hat{\beta} - \beta_0) \xrightarrow{D} N(0, A_0^{-1} J_0 A_0^{-1}). \tag{3.20}$$

To make inference about the parameter value, the asymptotic variance should be estimated by the usual sandwich formula

$$\widehat{\text{Avar}}(\hat{\beta}) = \frac{1}{T} A_T^{-1} J_T A_T^{-1}, \tag{3.21}$$

which is a White's heteroskedasticity robust estimator. Because of Condition 2, no serial correlation need to be accounted for.

Conditional Moment Test

A conditional moment type test (Newey 1985, Tauchen 1985) can be constructed from the errors of the fitted moments. It only requires the estimation of the restricted model. Here we first introduce the asymptotic distribution of the test statistics when true parameter is known, then discuss the correction of asymptotic variance when using the estimated parameter value.

The error vector of the first two moments is

$$V_{t+1} = \begin{bmatrix} r_{t+1} - E_t(r_{t+1}) \\ r_{t+1}^2 - E_t(r_{t+1}^2) \end{bmatrix},$$

---

[13]The maximum-likelihood estimator of uniform distribution parameters is not asymptotically normal, because the convergence comes from only one side (above or below). This is not a problem in our method-of-moments estimator, since the sample estimates are scattered evenly around the population truth. The statistical inference based on the Wald standard error is valid.

66

which has a distribution with $E_t(V_{t+1}) = [0, 0]'$ and $\Omega_{t+1} = \mathrm{Var}_t(V_{t+1})$. By construction, Ω_{t+1} is symmetric and positive definite. With the first four conditional moments at hand, we can construct Ω_{t+1} as

$$\Omega_{t+1} = \begin{bmatrix} E_t(r_{t+1}^2) & E_t(r_{t+1}^3) \\ E_t(r_{t+1}^3) & E_t(r_{t+1}^4) \end{bmatrix} - \begin{bmatrix} E_t(r_{t+1})E_t(r_{t+1}) & E_t(r_{t+1})E_t(r_{t+1}^2) \\ E_t(r_{t+1}^2)E_t(r_{t+1}) & E_t(r_{t+1}^2)E_t(r_{t+1}^2) \end{bmatrix}.$$

Applying a suitable version of central limit theorem, one has

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{T-1} \Omega_{t+1}^{-\frac{1}{2}} V_{t+1} \xrightarrow{D} \mathrm{Normal}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\right), \tag{3.22}$$

or equivalently,

$$\frac{1}{T} \sum_{t=1}^{T-1} V_{t+1}' \Omega_{t+1}^{-1} V_{t+1} \xrightarrow{D} \mathrm{Chi\text{-}Square}(2). \tag{3.23}$$

The regularity condition for the central limit theorem is very mild, since the error vector is not serially dependent and since the conditional heteroskedasticity, as a function of stationary r_t, is well-bounded. The latter chi-square test provides a convenient diagnostics for the jump-diffusion modeling.

As pointed out by Newey (1985) and Tauchen (1985), the asymptotic variance of the test statistic (3.23) is not correct if the estimated parameter value is used instead of the population truth. The adjusted asymptotic variance can be shown as

$$\Sigma = I_2 - J_{VS}(J_{SS})^{-1} J_{SV},$$

where I_2 is a 2×2 identity matrix, i.e., the asymptotic variance in (3.23) with true parameter. $J_{SS} = E[s_{t+1} s_{t+1}'] = J$, $VS = E[\Omega_{t+1}^{-1/2} V_{t+1} s_{t+1}']$, and $SV = VS'$ are simply the variance-covariance matrices of the moment condition vector

$$E\begin{bmatrix} s_{t+1} \\ \Omega_{t+1}^{-\frac{1}{2}} V_{t+1} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

67

Then the correct asymptotic distribution of the test statistic (3.23) is given by

$$\frac{1}{T} \sum_{t=1}^{T-1} V'_{t+1} \Omega_{t+1}^{-\frac{1}{2}} \Sigma^{-1} \Omega_{t+1}^{-\frac{1}{2}} V_{t+1} \xrightarrow{D} \text{Chi-Square}(2). \qquad (3.24)$$

It is clear that the adjusted asymptotic variance Σ is smaller than the unadjusted one I_2; hence only the upper bound of the test statistic (3.23) is following a Chi-square (2) distribution. Therefore rejecting a model by test (3.23) guarantees a rejection by test (3.24), but accepting a model by test (3.23) invites more type II error than test (3.24).

We will implement test (3.23) to sift out the rejected specification. Instead of calculating a complicated variance-covariance matrix of test (3.24), we complement our inference with a robust Lagrange multiplier test.

Lagrange Multiplier Test

The Lagrange Multiplier or Rao's Score test is a favorable choice for model specification testing, since it only requires the estimation under the null and since it is robust to heteroskedasticity. Under the identification condition (1) and the no-serial-correlation condition (2), the limiting distribution of scores are asymptotically normal; hence

$$\text{LM}_T = \frac{1}{T} \sum_{t=1}^{T-1} \hat{s}'_{t+1} A_T^{-1} \sum_{t=1}^{T-1} \hat{s}_{t+1} \xrightarrow{D} \mathcal{X}^2(l), \qquad (3.25)$$

where l is the number of parameters in the unrestricted model under the alternative minus the one in the restricted model under the null. A_T is the empirical estimate of the Hessian, and $\hat{s}_{t+1}$ is the empirical estimate of the score. Although under correct specification, the Hessian can be replaced by the outer product of gradient, it requires additionally the correct specification of conditional variance and may have poor finite sample properties (Wooldridge 1994). The validity of the LM test primarily resides

68

in the fact that non-serially-correlated errors lead to non-serially-correlated scores (Wooldridge 1994).

The Lagrange Multiplier test is equivalent to the GMM objective function test or the J-test, however, it can not fully substitute the conditional moment test discussed earlier. Since one can arbitrarily specify the dimension of the unrestricted model, the score tests on the same candidate models may yield different results. On the other hand, the conditional moment test does not rely on the dimension of the unrestricted model, hence the test result is robust to the choice of the alternative specification.

3.4   Empirical Application

Four jump-diffusion term structure models are implemented in this section. We first estimate the short-rate process using the MWNLS-JD estimator constructed in Section 3.3 and then perform the classical specification tests. Next we estimate the market prices of diffusion and jump risks by minimizing the pricing errors between an average, observed yield curve and a model-implied one. Finally we use the estimated parameters to calculate term structure and to study some comparative statics.

3.4.1   Data Description

The weekly three-month Treasury Bill rate is used to approximate the instantaneous short rate, because it is most widely traded on the secondary market and has new issue every week. The data set, which has a sample mean of 5.66%, and a sample standard deviation of 2.93%, is summarized in Table 3.1. The positive skewness indicates that there are more values below sample average than above, while the large kurtosis indicates heavy tails at both ends of the distribution. Looking at the first difference, it is well-centered but has extremely heavy tails. These are the typical features of financial time series data. We conjecture that any smooth sample path

69

distribution may not fit the extraordinary skewness and kurtosis. Either discrete change in regimes or large unexpected jumps must be introduced to cope with the seemingly non-stationary characteristics.

The time series level and first difference are plotted in Figure 3.1. It is clear that the short rate shows both smooth continuous movements and discrete large changes. Also, when the short rate jumps, the mean reversion and volatility persistence are intensified. This is one of the benefits of explicit jump-diffusion modeling. Incorporating the third and fourth moments enable us to capture the impact of large outliers. High volatility clustering may come from either the frequent but small zig-zags or the rare but large swings. Therefore conditional mean and variance alone are not sufficient to distinguish between a distribution with wide shoulders and one with fat tails. Shorter maturity yields like the seven-day EURO dollar rates or the overnight Federal Fund rates are more close to the short rate than the three month t-bill; but these yields have many liquidity shocks and too much institutional noise. Increasing the sampling frequency to daily may invite more microstructure distortion and day-of-week effect. Considering the weekly issuance structure and the deep market liquidity, weekly three month t-bill rate from the secondary market may be the best choice to represent the short-term interest rate.

3.4.2 Estimation of Jump-Diffusion Processes

The estimation results are summarized in Table 3.2. They include diffusion parameter estimates, jump parameter estimates, a conditional moment test, and a Lagrange multiplier test. The estimates of long-ran mean parameter θ are very similar across the four models, ranging from 4.03% to 4.86%. In jump-diffusion model, the "total long run mean" should include both the contribution from the diffusion (θ) and the contribution from the jumps ($\rho(\cdot)E_J[J(\cdot)]/\kappa$) (see the example in Section 3.3.1).

70

Since the sample mean is 5.66%, the pure diffusion (model 1) and dependent jump-size (model 4) underfit, but the independent jump (model 2) and the dependent jump-rate (model 3) are getting closer. The Student-t tests of individual parameters are mostly significant. The mean reversion parameter estimation is also very close, within the range from 0.0045 to 0.0052. It corresponds to a extremely persistent yet stationary component of the spot-rate process. The local variance parameter estimate is varying across the four models and is rejected by the dependent jump-size (model 4).

For model 2 (independent jump), the jump-rate and jump-size estimates seem reasonable. The jump amplitude follows a uniform distribution with a lower bound of -2.87% and an upper bound of 3.52%. The expected jump-size is about 30 basis points if one jump occurs. The jump-rate is 0.0120, i.e., roughly one jump in less than two years. The expected jump impact at any instant is only about 0.4 basis point. The jump term is statistically very significant.

For model 3 (state-dependent jump-rate), the interpretation of jump-rate and jump-size is slightly different. The jump lower bound parameter seems reasonable at -3.92%, while the upper bound of 11.92% seems too high. The instantaneous jump probability now becomes state dependent $0.0036 - 0.0266r_t$, featuring high interest-rate, low jump probability. The mean jump-size averaged by the jump probability, has an expected jump impact between 0 and 2 basis points. When the short-rate level is higher, the jump probability is lower, and the expected jump impact would be even smaller. All the jump parameters are very significant.

For model 4 (state-dependent jump-size), the jump-size is driven into unrealistically negative region $-0.2001 - r_t$, while the jump-rate is forced to be roughly one jump every 200 years. However the instantaneous expected jump impact is about -0.02 basis points. The near zero jump-rate with a large standard error indicates that

71

state-dependent jump-size specification may be rejected. In fact, the square-root parameter estimation in Model 4 is very close to that in Model 1 which has no jumps at all.

Turning to the overall specification, the conditional moment test (3.23) introduced in Section 3.2.3 strongly rejects the square-root diffusion process with p-value 0.0002. All the jump-diffusion models easily pass this test, with p-value ranging from 0.1031 to 0.4391. Since the unadjusted conditional moment test (3.23) may underreject the false alternatives, we need to investigate this possibility with the more robust LM test (3.25). The results in Table 3.2 suggest that all the candidate models are not rejected, however, the p-value of the square-root model is only 15%, while the p-values for the jump-diffusion models are about 82-99%. Recall that the LM test is sensitive to the dimension of the unrestricted VAR specification (3.16).

The parameter estimation and specification test suggest that the jump augmentation is statistically significant. Square-root model 1 is rejected. Both independent jump-rate model 2 and state-dependent jump-rate model 3 have reasonable parameter estimates. State-dependent jump-size model 4, although it passes the specification tests, has an insignificant jump term.

3.4.3 Estimating the Risk Premiums

The market prices of diffusion and jump risks must be obtained to complete the term structure calculation. Following Aït-Sahalia (1996a), we adopt a nonlinear least-square approach to minimize the pricing errors between a target yield curve and the model-implied yield curve. The target yield curve in Figure 3.2 is based on the time averages of discount bond yields from June 30, 1964 to December 29, 1995. The data set is obtained from the Center for Research in Security Prices (CRSP). There is a total of 379 monthly observations, with nine maturities (months 1, 3, 6, and 9 and

years 1, 2, 3, 4, and 5). These discount bond yields were first constructed by Fama (1984) and Fama and Bliss (1987), and then subsequently updated by CRSP. The short maturity yields are based on the average of bid and ask prices of Treasury bills and are normalized to reflect a standard month of 30.4 days. The long maturity yields are computed by interpolation, assuming the same forward rates between successive-maturity treasury bonds.

We take the time average as the target yield curve and use the model-implied yield curve to estimate the market prices of risks. Since there are nine maturities, we use the shortest maturity (1 month) to invert from the yield to the short rate. The objective function to be minimized is simply

$$\min_{\lambda_W, \lambda_J} \frac{1}{8} \sum_{\tau=2}^{9} [Y(\hat{\beta}; \tau; \lambda_W, \lambda_J) - \hat{Y}(\tau)]^2,$$

where $Y(\hat{\beta}; \tau; \lambda_W, \lambda_J)$ is the model-implied yield and $\hat{Y}(\tau)$ is the observed average yield. If the equilibrium price is observed without noise, the pricing error should equal zero. Without an economic theory for the measurement error, the most reasonable assumption is that the error is just white noise. The estimation results are shown in Table 3.3, together with the robust standard error and the minimized pricing errors.

For the square-root model 1, the market price of diffusion risk λ_W is negative with a reasonable magnitude, and is statistically significant. However, the fitted-yield curve shown in Figure 3.2 is very flat and slightly convex, not conforming to the historical average. On the other hand, all the jump-diffusion models achieve dramatic decreases in the minimized pricing error (see Table 3.3), and the implied yield curves are much more flexible in tracking down the observed yield curve (Figure 3.2). In the short end of the term structure, the state-dependent jump-size model 4 fits better. In the long end, the state-independent jump-rate model 2 seems closer. On average, the state-dependent jump-rate model 3 stands out as the best match.

73

It is also revealing to look at the parameter estimates of risk premiums. The market price of diffusion risk for model 2 (independent jump) is -0.0012, which is much smaller than the square-root model in absolute magnitude. The slack is picked up by the jump-risk parameter -545.5183, which is seemingly too large. Considering that the correct measure of jump-risk premium is $\rho(\lambda_J - 1)E_J(J)$, the expected jump-risk premium (-0.0213) is still in the normal range. The state-dependent jump-rate model 3 is the best fit. The positive diffusion risk price (0.0030) is somewhat counter-intuitive. The jump-risk price is -111.2223. The total jump-risk premium is now time-varying $(\rho_0 + \rho_1 r_t)(\lambda_J - 1)E_J(J)$, ranging from -0.0029 to -0.0136. The total risk premium of bond return is still within the reasonable range, 23 to 82 basis points. For the state-dependent jump-size model 4, the diffusion and jump-risk prices both have the wrong signs, and the bond return premium becomes negative. Even though the yield-curve fitting is not bad (Figure 3.2), its economic interpretation is hard to gauge. Overall, the state-dependent jump-rate model achieves the smallest pricing error.

3.4.4 Economic Comparative Statics

Relationships between bond prices or yields with respect to parameters changes are also of interest to financial economists. However, global analysis of comparative statics are usually quite involved, especially for the jump-diffusion processes. Even in the simple square-root model, some relationships have been disputed for decades (Sun 1992). Here we restrict to local comparative statics of the estimated parameters and numerically plot the yield curves.

It is common knowledge that the yields of all maturities rise as the current interest rate increases, with the short end of the term structure being more volatile (Cox et al. 1985a). However, the shape of the yield curve as a function of time-to-maturity can

74

be anything from concave increasing to convex decreasing, from flat-shaped to hump-shaped (Figure 3.2). Our study of comparative statics will focus on the structural parameters around their estimated values.

Figure 3.3 investigates the standard square-root model 1. The yield curve turns out to be a linear function of the long-run mean parameter θ, and the scale of the y-axis suggests that the yield changes are very small. In contrast, the yield change with regard to the mean reversion parameter κ is significantly convex and increasing. The local variance parameter σ dramatically pulls down the yield curve initially and then makes it almost flat. The diffusion risk price λ_W shows the most interesting shape: first concave and increasing below zero, then concave and decreasing above zero. These stylized facts of the square-root yield curve are quite standard.

Comparative studies of the jump-diffusion yield curves (models 2, 3, and 4) are reported in Figures 3.4, 3.5, and 3.6. The characteristics of the diffusion part parameters are very similar. The yield curves on long-run mean θ are all linearly increasing and very flat, just like that of the square-root process (model 1). For the mean reversion κ, yield curves are all linearly decreasing, exactly opposite to the square-root process. This is possibly caused by the positive expected jump-size, which captures most of the upward mean-reversion and leaves the downward mean-reversion to the diffusion part. All the yield curves on local variance σ are decreasing and concave, which is the same as in the square-root model. Lastly, the market price of diffusion risk λ_W drives all the jump-diffusion yield curves downward, with the independent jump model being more linear, the state-dependent jump-rate model more concave, and the state-dependent jump-size model most concave. In contrast, the square-root counterpart is almost quadratic.

Since the jump-term specifications are different across the three models, detailed comparisons are given here. The yield curves on lower-bound a and upper-bound

75

b are linearly increasing in model 2 (independent jump), while they are concave increasing in model 3 (state-dependent jump-rate). From the pricing functions one can see that the expect jump impact not only feeds into the intercept (model 2), but also into the slope (model 3) (Recall that yield curve is a linear function of the short rate). The constant jump-rate ρ in model 2 drags down the bond return, similar to the linear jump-rate function in model 3. This is no surprise, since the mean jump-size has a positive impact on the yield level, and the jump-rate as a measure of risk has a negative impact on yield level. In other words, jump-size plays mostly the role of drift, and jump-rate plays mostly the role of local variance. The impact from the jump risk premium parameter λ_J is linearly increasing in models 2 and 3. This is consistent with the pure square-root model when the λ_W is less than zero. The comparative statics of the jump term in model 4 is counter-intuitive. The impact of the intercept in the state-dependent jump-size is negative, in contrast to the usual positive sign of the expected jump-size. The jump-rate has an almost flat impact at first, then dramatically oscillates into positive and negative regions, without any sensible explanation. One should be reminded that the state-dependent jump-size model 4 is not an equilibrium compatible specification, and its pricing solution can only be approximated (see Section 3.2.3). The impact from the jump risk premium parameter is similar to those in models 2 and 3. In short, both the empirical estimation and the comparative statics suggest that the state-dependent jump-size model is not a good candidate for modeling the short interest rate.

Overall, the yield-curve comparative statics have economically meaningful signs and magnitudes and tend to be more sensitive to both the factor risk parameters (σ and ρ) and the risk aversion parameters (λ_W and λ_J). The specifications of independent jump-diffusion and state-dependent jump-rate models stand out with convincing economic interpretations.

3.5 Conclusion

In this paper we have proposed a class of jump-diffusion models of equilibrium interest rate. Our contribution is to design an efficient estimation strategy with valid specification tests. We can provide analytical solutions for both the conditional moments of short-rate dynamics and the entire term structure of different maturities. The parameter estimation and statistical inference suggest that jump impact is very significant. A better capturing of the short-rate volatility can reduce the pricing error of the entire yield curve. Numerical analysis of comparative statics is also consistent with the economic intuitions. In general, the jump-diffusion term structure is very sensitive to the factor risks and the attitudes toward risk. The state-dependent jump-rate model and the independent jump-rate model are favored by both the econometric diagnostics and the economic analysis.

The challenge of fitting the short-term interest rate is to accommodate the rich volatility feature. Jump-diffusion modeling is one of numerous efforts that are working in this direction. An important lesson from this study is that, in order to estimate the volatility more accurately and more efficiently, one must incorporate the third and fourth moments in addition to conducting an extensive specification search of the first and second moments. This necessity is due to the fact that the fat-tail characteristics of interest rate data can not be well explained by a smooth sample path distribution with stationarity restrictions.

3.6 Tables and Figures

Table 3.1: Summary Statistics of Three Month T-Bill Rate

The weekly three month T-bill yield data is obtained from the Federal Reserve Bank of St. Louis. The time series ranges from January 8, 1954 to June 5, 1998, with total observations of 2318. The skewness and kurtosis are 0 and 3 for a standard normal distribution.

Maturity	Annualized Yield	First Difference
Moments		
Mean	0.0566	0.0000
Std Deviation	0.0293	0.0022
Skewness	1.1138	-0.6146
Kurtosis	4.6870	23.1248
Quantiles		
Minimum	0.0059	-0.0189
05 percent	0.0198	-0.0027
25 percent	0.0353	-0.0006
50 medium	0.0522	0.0000
75 percent	0.0726	0.0007
95 percent	0.1143	0.0027
Maximum	0.1736	0.0203

78

Table 3.2: Parameter Estimates and Specification Tests
The details of models 1-4 are discussed in section 3.2.2; the MWNLS-JD estimator is introduced in section 3.3.1 and 2.3.2; and the two specification tests are laid out in section 3.3.2. White's robust standard errors are reported in the parentheses for all parameter estimates.

Parameter	Model 1	Model 2	Model 3	Model 4
Diffusion				
Long-Run Mean θ	0.0477	0.0448	0.0403	0.0486
	(0.0014)	(0.0013)	(0.0006)	(0.0002)
Mean Reversion κ	0.0049	0.0052	0.0043	0.0050
	(0.0001)	(0.0001)	(0.0001)	(0.0001)
Diffusion σ	0.0061	7.8e-6	0.0028	0.0046
	(0.0014)	(0.0029)	(0.0009)	(0.0012)
Jump-Size				
Independent a		-0.0287	-0.0392	
		(0.0011)	(0.0007)	
Independent b		0.0352	0.1193	
		(0.0010)	(0.0006)	
Dependent J_0				-0.2001
				(0.0630)
Jump-Rate				
Independent ρ		0.0120		1.3e-5
		(0.0011)		(1.2e-5)
Dependent ρ_0			0.0036	
			(0.0002)	
Dependent ρ_1			-0.0266	
			(0.0010)	
Conditional Moment Test				
Chi-Square(2)	17.5726	1.6461	4.5435	2.9851
p-value	0.0002	0.4391	0.1031	0.2248
Lagrange Multiplier Test				
Chi-Square	15.3962	0.5394	3.6657	2.2924
Degree of Freedom	(11)	(8)	(7)	(9)
p-value	0.1651	0.9998	0.8174	0.9860

79

Table 3.3: Market Prices of Risks

The pricing error is the standard deviation of the minimized residue between the average observed yield curve and the model-implied yield curve. White's robust standard errors are given in parentheses. The pricing functions for models 1, 2, 3, and 4 are discussed in section 3.2.2.

	Diffusion Risk λ_W	Jump Risk $\lambda_J - 1$	Pricing Error
Model 1	-0.0870 (0.0069)		0.0102
Model 2	-0.0012 (0.0001)	-545.5183 (60.6029)	0.0015
Model 3	0.0030 (0.0003)	-140.5368 (16.3018)	0.0009
Model 4	6.2829 (0.2709)	131910 (5418)	0.0011
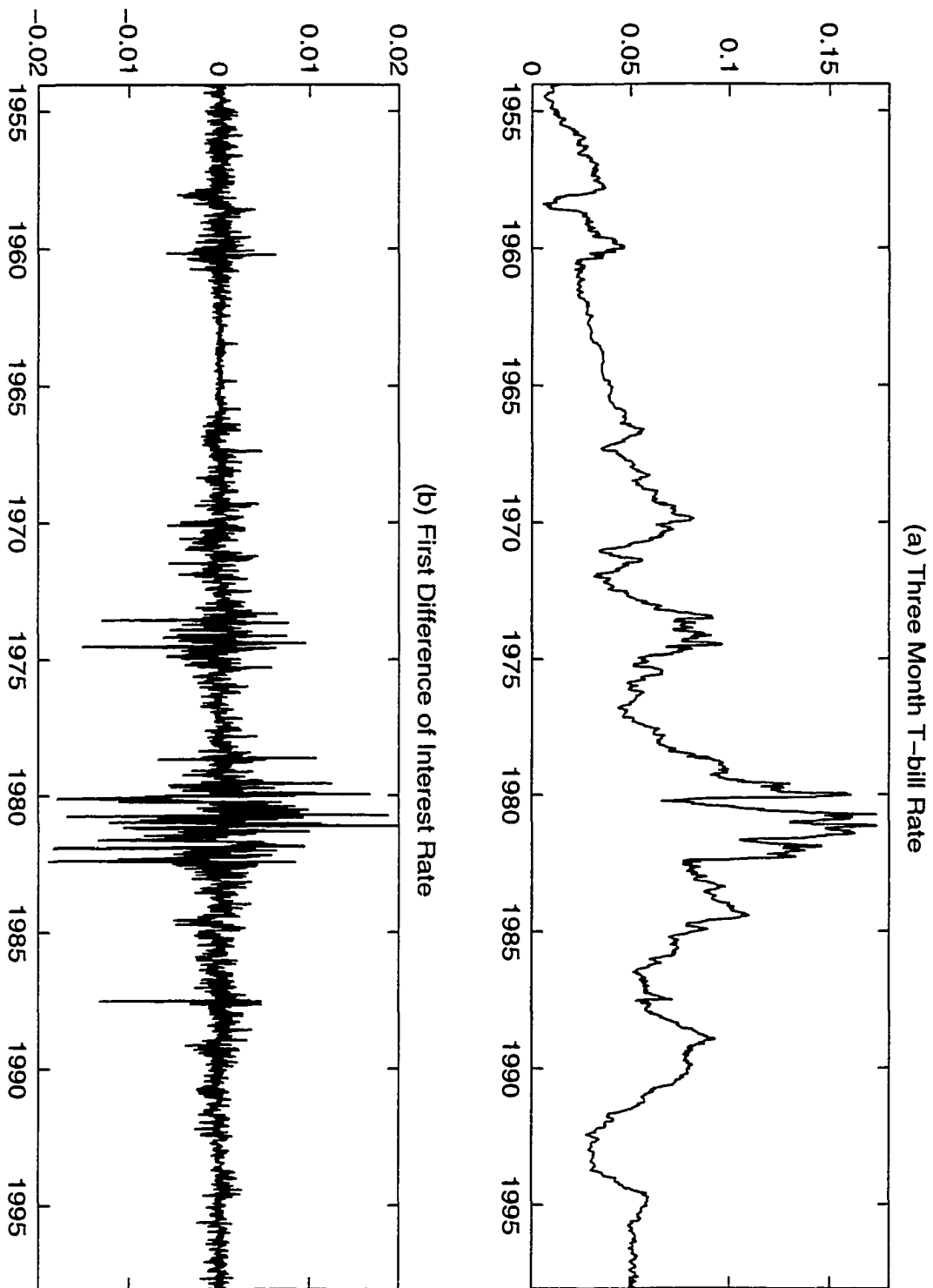
Figure 3.1: Time Series Plot of Weekly Three Month T-Bill Rate.

81
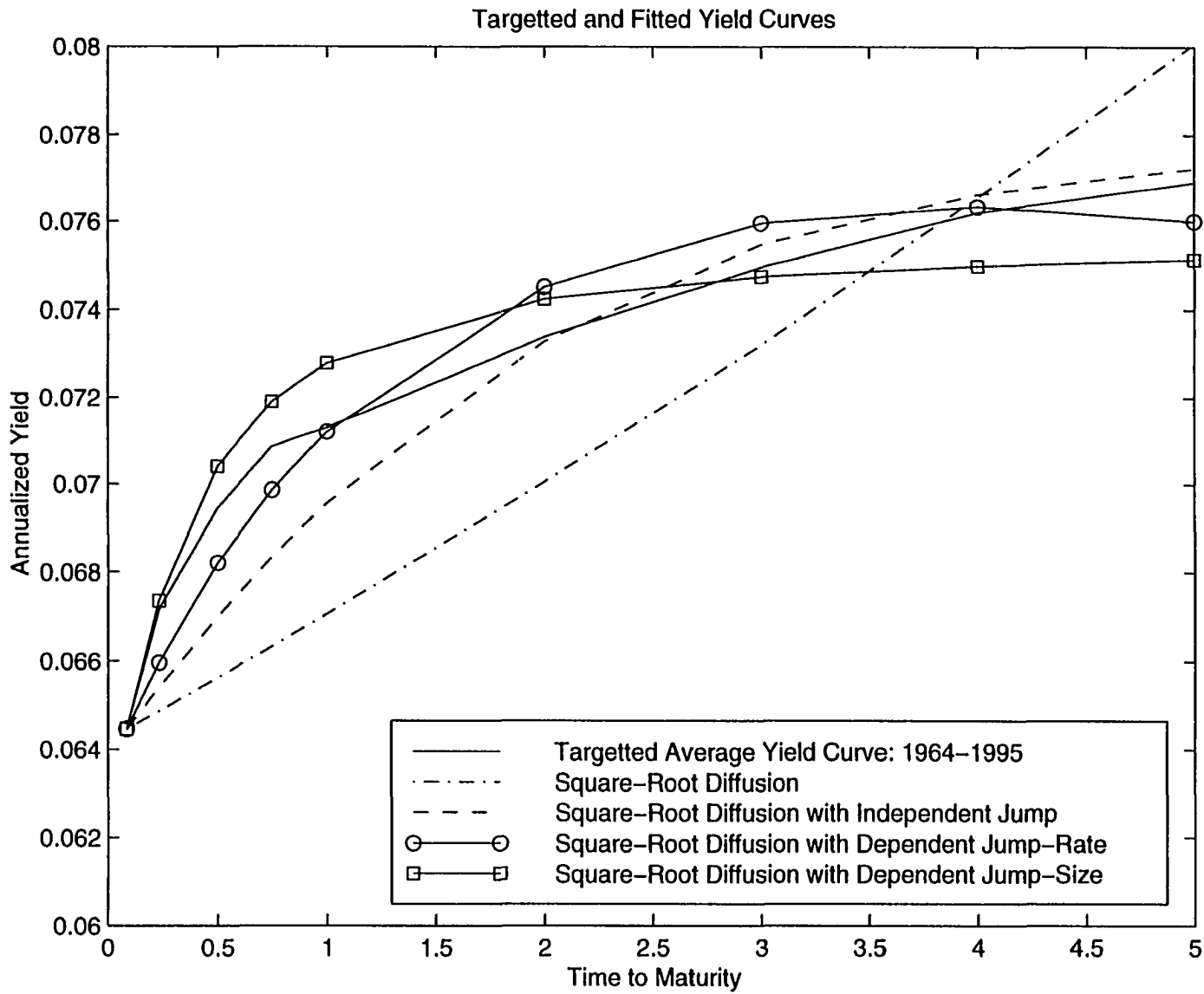
Figure 3.2: Fitting Term Structure of Different Interest Rate Models.
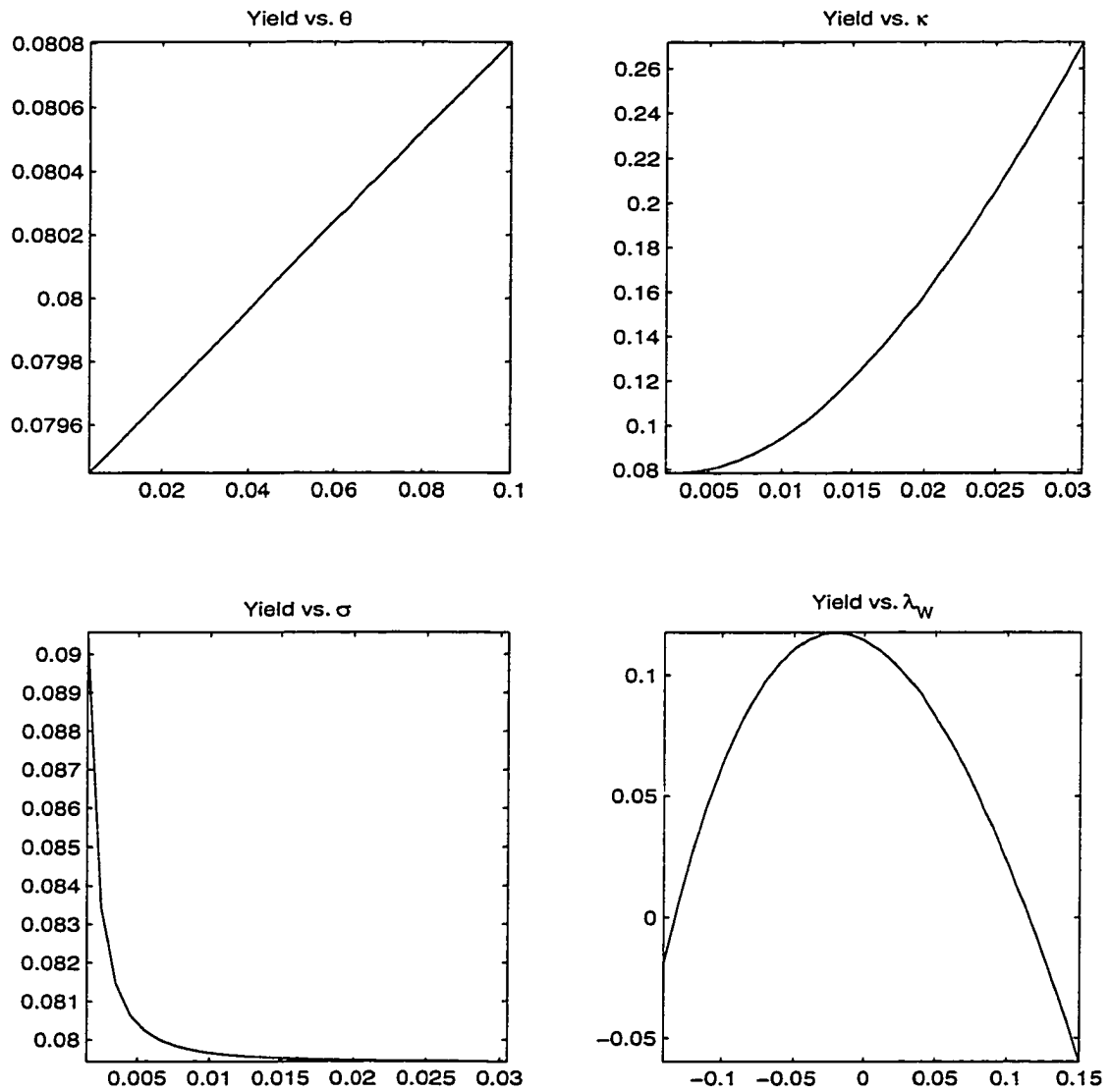
Figure 3.3: Comparative Statics for Model 1 with $r_t = 5.66\%$ and YTM $= 5$ Year.
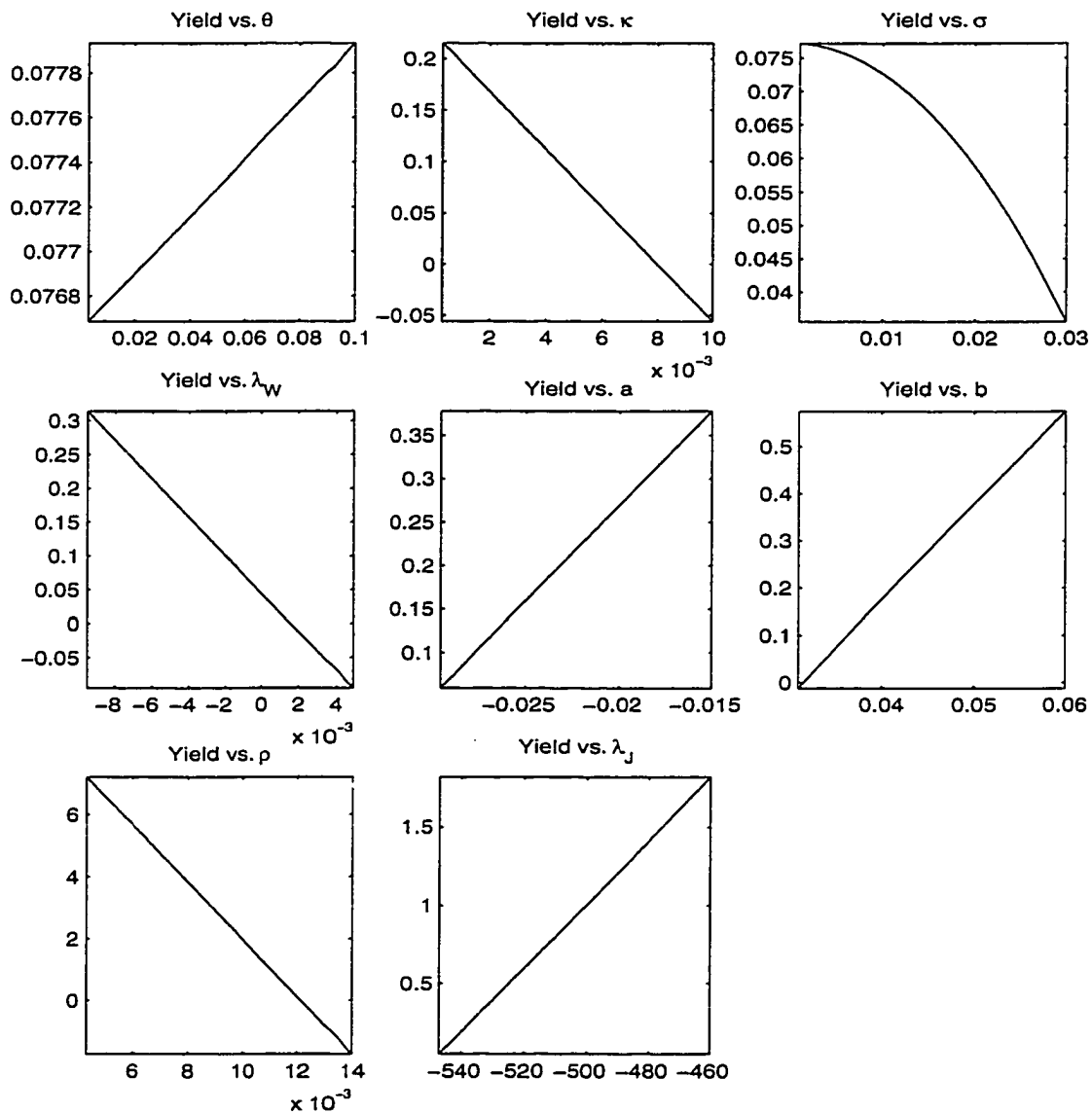
Figure 3.4: Comparative Statics for Model 2 with $r_t = 5.66\%$ and YTM $= 5$ Year.

84
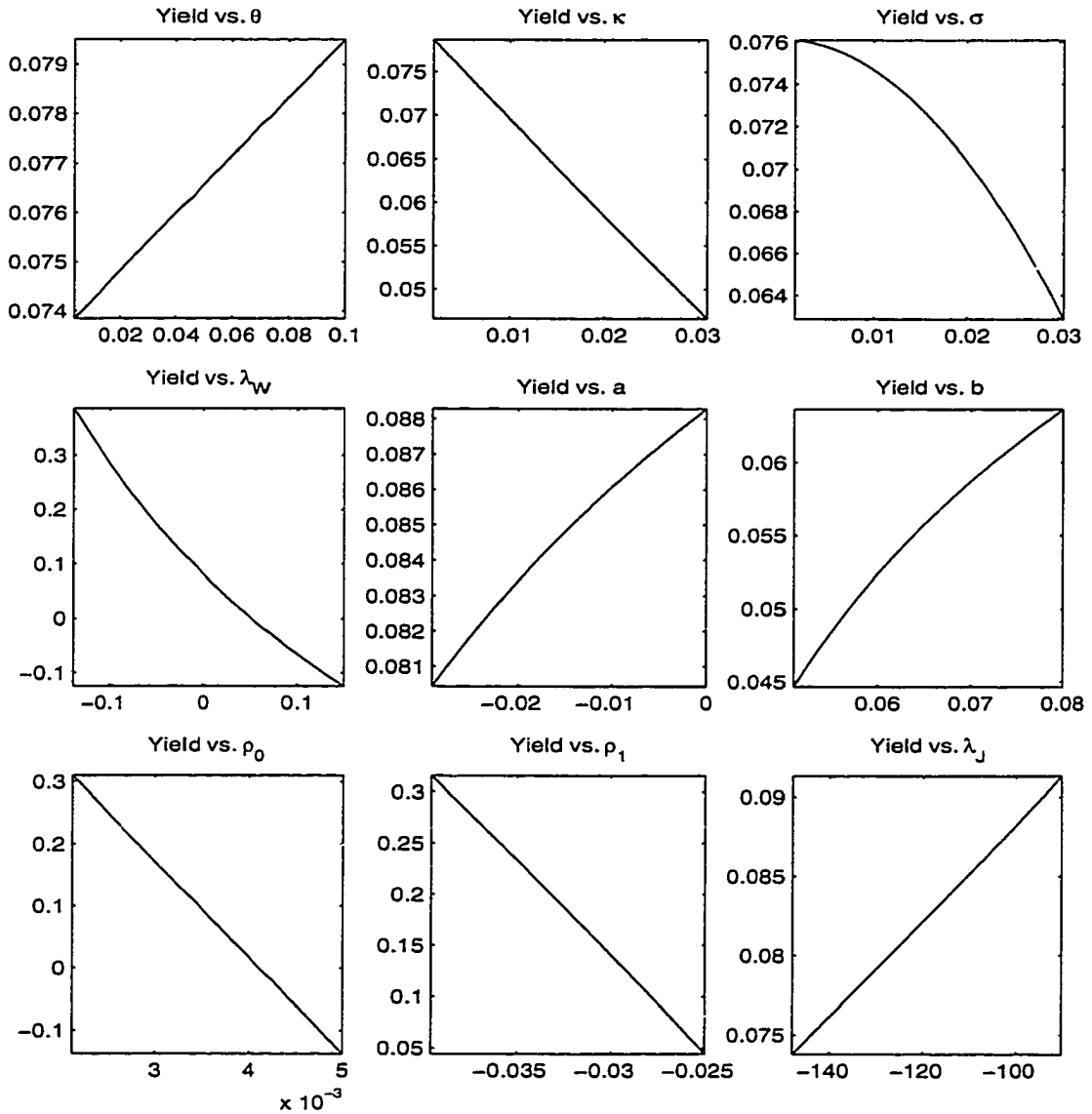
Figure 3.5: Comparative Statics for Model 3 with $r_t = 5.66\%$ and YTM $= 5$ Year.

85
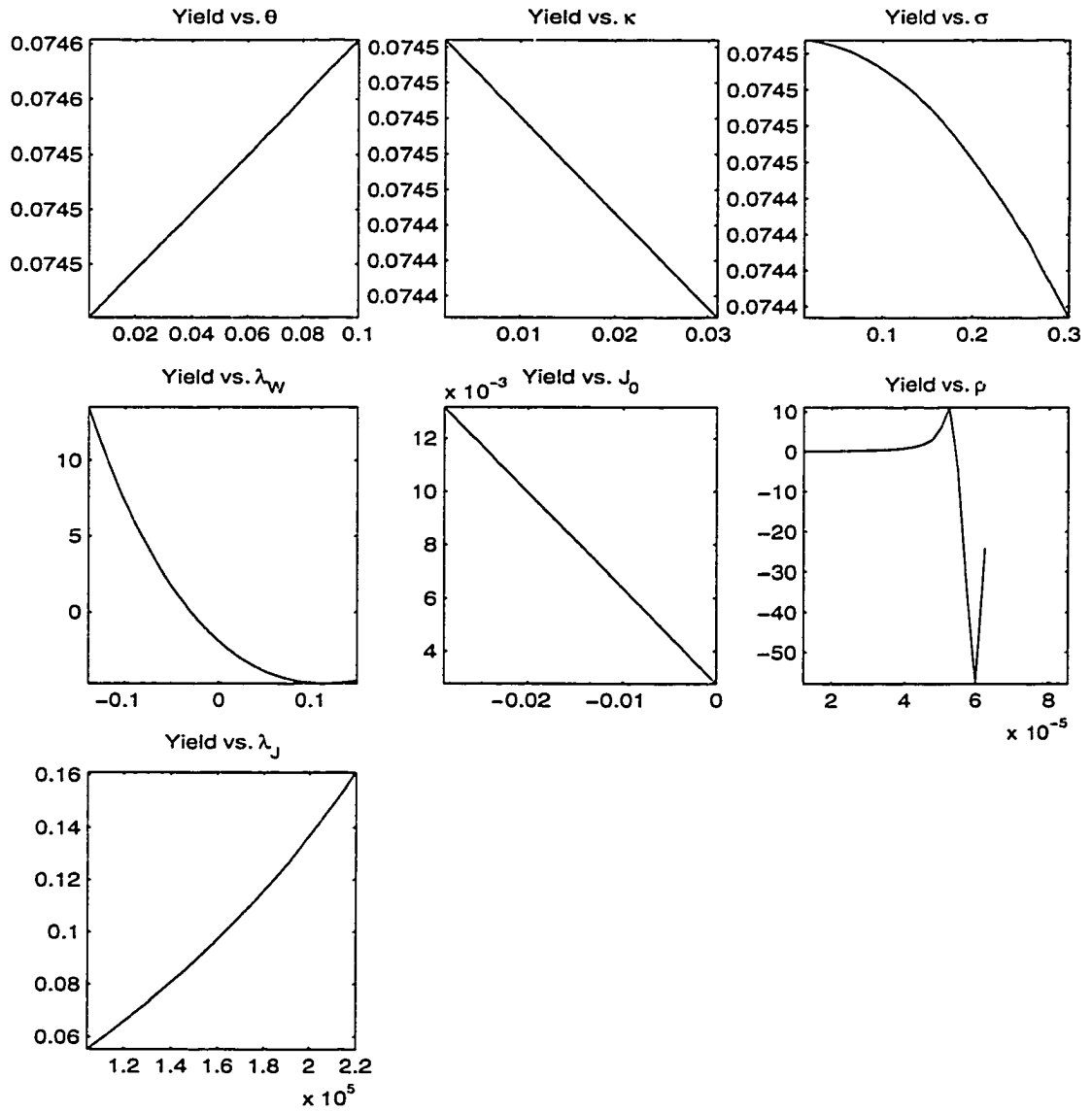
Figure 3.6: Comparative Statics for Model 4 with $r_t = 5.66\%$ and YTM = 5 Year.

86

Chapter 4

Term Structure of Interest Rate with Regime Shifts

[1]In this chapter I explore the possibility that changes in the conduct of interest rate policy have important effects on the term structure of interest rates. To capture this idea, we develop an equilibrium term structure model with the underlying short term interest rate and the market price of risk are determined by latent regime-switching square-root processes. I will combine the regime-switching methodology proposed by Hamilton (1989, 1990, 1996) with the recent literature of affine term structure models as in Duffie and Singleton (1997) and Dai and Singleton (2000), and provide analytical solutions for bond prices of different maturity. Exploiting the analytical solutions, one can estimate the model via the Efficient Method of Moments (EMM), using short and long interest rate data from 1964-1995. The empirical results show that a regime shifts two-factor square-root model, finds considerable support in the data. Standard term structure models (with up to three factors), which do not allow for regime shifts, are sharply rejected in the data. The diagnostics show that only the preferred regime switching model can reasonably mimic the observed conditional volatility and correlation of yields. In addition, the estimated regime-switching model produces the smallest absolute pricing error in fitting the yield curve. I find that the extracted regimes are intimately related to business cycles, and the key observable difference across regimes is in the yield spread.

---

[1]Part of the material of this chapter is in Bansal and Zhou (1999).

87

4.1 Introduction

Many recent papers find that the univariate short interest rate process can be reasonably well modeled as a regime switching process (see Hamilton (1988, 1996), Gray (1996), Garcia and Perron (1996), and Ang and Bekaert (1998)). In addition to this statistical evidence, there are economic reasons as well to believe that regime shifts are an important aspect of the interest rate data. The conduct of monetary policy, which has first order effects on interest rates, is subject to discrete changes in policy regimes. The discrete changes in policy regimes, for example, may reflect the differences in conduct of policy during recessions and expansions. Such shifts in regimes, on economic grounds, have significant impact on not only the behavior of the short interest rate, but on the entire term structure of interest rates. Standard term structure models, such as the Cox et al. (1985a) (CIR) model, do not permit the possibility of discrete changes in regimes despite the potential for sizable effects from such regime shifts. In the context of the U.S. treasury yield curve, the poor empirical performance of the various versions on the CIR model, as documented in Brown and Dybvig (1986), Gibbons and Ramaswamy (1993), Chen and Scott (1993), and Pearson and Sun (1994), may well be due to the fact that they do not permit the possibility of discrete regime shifts.[2]

In this essay we take seriously the idea that changes in regimes have sizable effects on the term structure and incorporating them can better account for the observed behavior of the term structure of interest rates. Motivated by this possibility we develop a discrete time model of the term structure that incorporates persistent regime shifts. As in the CIR model, one can provide an analytical solution for the

---

[2]In particular, evidence provided by Brown and Dybvig (1986), shows that the parameters of the model change considerably across time; a feature which is consistent with the premise of regime shifts. In the context of swap yields Duffie and Singleton (1997) and Dai and Singleton (2000) estimate affine term structure models. Aït-Sahalia (1996b), Andersen and Lund (1997), estimate different short interest rate models.

entire term structure of interest rates in the presence of discrete and persistent regime shifts and time-varying risk premium. The time-varying risk premium is an important aspect of the specification given the overwhelming violations of the expectations hypothesis in term structure data, as documented in Campbell and Shiller (1991). It is useful to note that many of the above mentioned papers, model the univariate short rate process as a regime switching process, however they do not derive or explore the implications of regime shifts for the joint conditional distribution of multiple yields (i.e., the term structure), which is an important focus of this paper.

To provide a consistent method to estimate various models under consideration one can rely on a simulation based estimator for the model. In particular I use the Efficient Method of Moments, developed in Gallant and Tauchen (1996b), and Bansal et al. (1995) to estimate all the models under consideration. Tests of over-identifying restrictions based on the EMM method provide a way to compare different, potentially non-nested models. This estimation technique forces the model to confront several important aspects of the data, such as the conditional volatility and correlation across different yields. This permits one to analyze whether a given term structure model can account for these important time-series and cross-sectional aspects of the data. To provide diagnostics which permit sharp discrimination across models I rely on the reprojection methods developed by Gallant and Tauchen (1998b). The reprojected conditional density, is the density for yields conditional on their lags for a given term structure model. Using this conditional density, one can evaluate if different model specifications imply conditional volatility and cross-correlations of yields comparable to those found in the data.

Exploiting the derived analytical solution, I estimate the regime shifts model using monthly term structure data on U.S. treasury bills and bonds for the period from 1964-1995. In addition, I also estimate several versions of a benchmark CIR

model. In virtually all empirical dimensions, our model, which incorporates discrete regime shifts, does better than the multi-factor versions of the CIR model. Indeed the evidence shows that only the regime shift model can account for the observed conditional volatility and cross-correlation of yields. Formal statistical tests cannot reject our model which incorporates regime shifts, while various specifications of the CIR are sharply rejected in the data.

In terms of more specific empirical results, I find that the one, two, and three factor CIR models are sharply rejected with p-values of zero. While the three factor model provides some improvement over the two factor model, the diagnostics show that this model cannot justify the observed conditional volatility and conditional cross-correlations across yields. The only model specification that finds support in the data is the two-factor regime switching model. Tests of overidentifying restrictions do not reject this model (with p-value of 14%). Further, the implications of this model for conditional volatility and correlation across the short and long interest rates are surprisingly close to those found in the data. Given a set of estimated parameters I also explore the ability of a given specification to capture all observed yields in the cross-section, at each date. I show that the two factor regime switching model produces the smallest cross-sectional pricing errors across all the specifications considered in the paper. The empirical evidence suggests that regimes differ primarily in terms of the volatility of the short interest rates, and in the slope of the yield curve. Further, regime indicators are intimately related to the phases of the business cycle—regime with low yield spreads seems to occur prior to business contractions.

The remaining of this paper is organized in the following manner. Section 4.2 presents the benchmark and the regime shifts term structure model. Section 4.3 discusses the EMM estimation strategy and Section 4.4 presents the non-parametric density for yields used to estimate the various term-structure models. Section 4.5

90

discusses the empirical evidence related to different models and presents an array of diagnostics. Section 4.6 provides cross-sectional implications on pricing errors and regime classification from the observed yield data and Section 4.7 presents concluding comments.

4.2 Term Structure Models of Interest Rates

In this section, I first derive the term structure implications for the specification considered in Sun (1992), Backus and Zin (1994), and Campbell, Lo, and MacKinlay (1997). The solution methods used for this benchmark case are then extended to derive results in the case where the state variables follow a regime switching process[3]. The derivation's focus is on a single factor (i.e., state variable) as the extensions to the multi-factor case follow immediately from the single factor case.

4.2.1 Benchmark Model: The CIR model

The key building block of this benchmark case is the state variable x which follows the following square-root process.

$$x_{t+1} - x_t = \kappa(\theta - x_t) + \sigma\sqrt{x_t}u_{t+1}, \tag{4.1}$$

where $u_{t+1} \sim N(0,1)$ is a white noise, κ is the mean reversion parameter, θ is the long-run mean parameter, and σ is the local variance parameter. This is the discrete time counterpart to the CIR model, discussed in considerable detail in Sun (1992) and Campbell et al. (1997). Sun (1992) also shows that this discrete time specification converges to the continuous-time specification when the interval across

---

[3]In a similar general equilibrium setting, Evans (1998) investigates the impact of regime-switching on the nominal and real interest rates. An earlier work by Naik and Lee (1997) solves the term structure model with regime shifts in continuous time and provides empirical evidence that a one factor regime-switching model can out-perform a two factor Affine model.

91

adjacent time periods shrinks. Based on 4.1, it follows that $E_t[x_{t+1}] = x_t + \kappa(\theta - x_t)$, and $Var_t(x_{t+1}) = \sigma^2 x_t$.

I also assume that the pricing kernel (the inter-temporal marginal rate of substitution in equilibrium) is $M_{t+1} = \exp\{-r_{f,t} - (\alpha\sigma)^2 \frac{x_t}{2} - \alpha\sigma\sqrt{x_t}u_{t+1}\}$, where α affects compensation for aggregate risk. For deriving many of the formulas it is also convenient to re-write the pricing kernel in terms of $\lambda \equiv \alpha\sigma^2$—compensation for systematic risk is proportional to this parameter. Consequently, the pricing kernel can, equivalently be written as

$$M_{t+1} = \exp\{-r_{f,t} - (\frac{\lambda}{\sigma})^2 \frac{x_t}{2} - \frac{\lambda}{\sigma}\sqrt{x_t}u_{t+1}\} \qquad (4.2)$$

Note that $E_t[M_{t+1}] = \exp(-r_{f,t})$, where $r_{f,t}$ is the continuous one period risk-free rate—this implies the usual restriction that the conditional mean of the pricing kernel must equal the price of the one period default-free discount bond. Further, the asset pricing restrictions on bond returns imply that, $E_t[M_{t+1}h_{n,t+1}] = 1$; where $h_{n,t+1} = [\frac{P(t+1,n-1)}{P(t,n)}]$ is the one period gross return on a pure discount bond with n periods to maturity at date t. The price of this bond at t is $P(t,n)$.

To solve for the discount bond prices at date t which mature at $t+n$, I conjecture that the bond price, $P(t,n)$, is a function of the state variable x_t and the time to maturity n,

$$P(t,n) = \exp\{-A(n) - B(n)x_t\}.$$

No-arbitrage conditions imply that instantaneous value of a dollar should be one, hence for $n = 0$, $A(0) = B(0) = 0$. In addition, I also conjecture that $A(1) = 0$ and $B(1) = 1$, that is $r_{f,t} = x_t$. Based on this conjecture, and the assumed process for x, the continuous bond return $\ln[\frac{P(t+1,n-1)}{P(t,n)}]$, is conditionally normally distributed with mean

$$\mu_{n,t} = -A(n-1) - B(n-1)E_t[x_{t+1}] + A(n) + B(n)x_t$$

92

and variance

$$\sigma_{n,t}^2 = B(n-1)^2 Var_t(x_{t+1}).$$

Exploiting the asset pricing condition, $E_t[M_{t+1}h_{n,t+1}] = 1$, the joint lognormality of M_{t+1} and $h_{n,t+1}$, and our conjectured solution, I derive

$$\exp\{-r_{f,t} + \mu_{t,n} + \frac{\sigma_{n,t}^2}{2} + B(n-1)\lambda x_t\} = 1 \qquad (4.3)$$

The third term in this expression, $\frac{\sigma_{n,t}^2}{2}$, is the Jensen's adjustment for continuous returns and the last term, $B(n-1)\lambda x_t$ is the risk-premium associated with the bond return. Using the restriction $r_{f,t} = x_t$, the definitions for $\mu_{t,n}$, and $\sigma_{t,n}^2$, and taking logs of 4.3 it follows that

$$-x_t - A(n-1) - B(n-1)E_t[x_{t+1}] + A(n) + B(n)x_t + \frac{\sigma_{n,t}^2}{2} + B(n-1)\lambda x_t = 0 \quad (4.4)$$

It is useful to recognize that 4.4 can also be derived by using $\exp(y) - 1 = \sum_{k=1}^{k=\infty} \frac{y^k}{k!} \approx y$. I use this relation latter in the paper. While this is an approximate result in discrete time, it is exact in continuous time—over small intervals of time, each term in 4.3 is scaled by the length of the time interval Δt, consequently higher order term in the approximation (i.e. terms y^k, $k > 1$) have negligible effects on the solution. For greater details regarding this approximation see Sun (1992), Merton (1990), Backus and Zin (1994), and Ingersoll (1987). Equation 4.4 says that expected excess returns, in equilibrium, must equal the compensation for systematic risk.

Using 4.4 it follows that $A(n)$ and $B(n)$ satisfy,

$$B(n) = (1 - \lambda - \kappa)B(n-1) - \frac{1}{2}\sigma^2 B^2(n-1) + 1,$$

$$A(n) = A(n-1) + \kappa\theta B(n-1) \qquad (4.5)$$

with boundary conditions $A(0) = B(0) = 0$. The solution for $A(n)$ and $B(n)$, implied by 4.5, satisfies the asset pricing conditions 4.4 and is consistent with our conjectured

93

solution. This solution is comparable to the continuous time CIR model. Note that the yield-to-maturity of a discount bond in discrete time is defined as

$$Y(t, n) = -\frac{\ln P(t, n)}{n} = \frac{A(n)}{n} + \frac{B(n)x_t}{n} \tag{4.6}$$

Extending this analysis to multiple independent square root factors is straight forward, and is discussed briefly below.

4.2.2 Introducing Regime-Switching

We are interested in deriving the implications for the term structure when the economy is subject to regime shifts. As discussed earlier, the motivation for this is to try and capture, in an analytically tractable manner, the potential effects of shifts is policy. To keep things tractable I will model the regime shifts process as a two state Markov process as in Hamilton (1989, 1990).

Suppose that the evolution of tomorrow's regime $s_{t+1} = 0, 1$ given today's regime $s_t = 0, 1$ is governed by the transitional probability matrix of a Markov chain

$$\Pi = \begin{bmatrix} \pi_{00} & \pi_{01} \\ \pi_{10} & \pi_{11} \end{bmatrix}, \tag{4.7}$$

where $\sum_{j=0,1} \pi_{ij} = 1$ and $0 < \pi_{ij} < 1.$[4] In addition to the discrete regime shifts the economy, as in the benchmark case, the economy is also affected by a continuous state variable,

$$x_{t+1} - x_t = \kappa_{s_{t+1}}(\theta_{s_{t+1}} - x_t) + \sigma_{s_{t+1}}\sqrt{x_t}u_{t+1}, \tag{4.8}$$

where $\kappa_{s_{t+1}}$, $\theta_{s_{t+1}}$, and $\sigma_{s_{t+1}}$, are the regime-dependent mean reversion, long run mean, and volatility parameters respectively. All these parameters are subject to

---

[4]When, the transition probability is independent of the regime, that is when $\pi_{00} = \pi_{10}$ I get an non-persistent (an iid) regime switching process. Note this probability restriction implies that $\pi_{01} = 1 - \pi_{00} = \pi_{11}$, and $\pi_{00} + \pi_{11} = 1$. Further $\pi_{00} + \pi_{11} - 1$, as shown in Hamilton (1989), determines the persistence in the regime process, which in the iid case is zero.

94

discrete regime shifts. Specifically, $x_{t+1} - x_t = \kappa_0(\theta_0 - x_t) + \sigma_0\sqrt{x_t}u_{t+1}$ if the regime $s_{t+1} = 0$, and $x_{t+1} - x_t = \kappa_1(\theta_1 - x_t) + \sigma_1\sqrt{x_t}u_{t+1}$ if the regime $s_{t+1} = 1$. Note that the innovation in the process characterized in 4.8, u_{t+1}, is conditionally normal given x_t and s_{t+1}. In general, there are two systematic risks, one is square-root factor risk and the other regime shifts risk.

For analytical tractability I assume that the process for regime shifts s_{t+1} is independent of x_{t+1-l}, $l = 0 \cdots \infty$, this is similar to the assumptions made in Hamilton's regime switching models. It is also assumed that the agents in the economy observe the regimes, though the econometrician may possibly not observe the regimes.

The pricing kernel for this economy, is similar to that in the benchmark case, save for incorporating regime shifts

$$M_{t+1} = \exp\{-r_{f,t} - (\frac{\lambda_{s_{t+1}}}{\sigma_{s_{t+1}}})^2\frac{x_t}{2} - \frac{\lambda_{s_{t+1}}}{\sigma_{s_{t+1}}}\sqrt{x_t}u_{t+1}\} \qquad (4.9)$$

For much of the analysis it is useful to define two information sets; $\mathcal{F}_t = \{s_l, x_l; -\infty \leq l \leq t\}$ is the information known till time t, and $\mathcal{F}_t^a = \mathcal{F}_t \cup \{s_{t+1}\}$ is an augmented information. $\mathcal{F}_t^a$ additionally includes information regarding the regime, s_{t+1}, but does not include x_{t+1}. I use the notation $\mathcal{F}_{j,t}^a$, to emphasize that $s_{t+1} = j$, $j = 0, 1$.

Using 4.9 it follows that $E(M_{t+1}|\mathcal{F}_t) = \exp(-r_{f,t})$—the price of a one period pure discount bond.[5] The asset pricing conditions that pure discount bond returns must satisfy, as before is $E[M_{t+1}h_{n,t+1}|\mathcal{F}_t] = 1$. With regime shifts, I conjecture that the bond price with n periods to maturity, at date t depends on the regime $s_t = i$,

---

[5]To see this, I exploit the law of iterated expectations given that $s_t = i$,

$$E[E(M_{t+1}|\mathcal{F}_t^a)|\mathcal{F}_t] = \exp(-r_{f,t}) \sum_{j=0,1} \pi_{ij}E[\exp\{-(\frac{\lambda_j}{\sigma_j})^2\frac{x_t}{2} - \frac{\lambda_j}{\sigma_j}\sqrt{x_t}u_{t+1}\}|\mathcal{F}_{j,t}^a] \qquad (4.10)$$

The expectation based on $\mathcal{F}_{j,t}^a$ uses the additional information regarding the regime at s_{t+1}, and the probability averaging integrates out the dependence on the future regime s_{t+1}. The conditional mean, $E[\exp\{-(\frac{\lambda_j}{\sigma_j})^2\frac{x_t}{2} - \frac{\lambda_j}{\sigma_j}\sqrt{x_t}u_{t+1}\}|\mathcal{F}_{j,t}^a] = \exp(0)$ for $j = 0, 1$ and the probabilities sum to one. Hence the result follows.

$i = 0, 1$, and x_t

$$P_i(t, n) = \exp\{-A_i(n) - B_i(n)x_t\}.$$

The one period ahead bond price, analogously depends on s_{t+1} and x_{t+1}

$$P_{s_{t+1}}(t+1, n-1) = \exp\{-A_{s_{t+1}}(n-1) - B_{s_{t+1}}(n-1)x_{t+1}\}.$$

In addition I impose the no-arbitrage condition $A_i(0) = B_i(0) = 0$, and also conjecture that $A_i(1) = 0$, $B_i(1) = 1$, for $i = 0, 1$, that is, $r_{f,t} = x_t$.

The continuous one period ahead bond return, $\ln(1 + h_{n,t+1}) = \ln[\frac{P_{s_{t+1}}(t+1,n-1)}{P_i(t,n)}] = -A_{s_{t+1}}(n-1) - B_{s_{t+1}}(n-1)x_{t+1} + A_i(n) + B_i(n)x_t$. Given x_t and $s_{t+1} = j$, the continuous bond return is normally distributed with mean

$$\mu_{n,j,t} = -A_j(n-1) - B_j(n-1)E[x_{t+1}|\mathcal{F}_{j,t}^a] + A_i(n) + B_i(n)x_t, \qquad (4.11)$$

and conditional variance

$$\sigma_{n,j,t}^2 = B_j(n-1)^2\sigma_j^2 x_t. \qquad (4.12)$$

The asset pricing condition, using law of iterated expectations is

$$E[E(M_{t+1}h_{n,t+1}|\mathcal{F}_t^a)|\mathcal{F}_t] = \sum_{j=0,1} \pi_{ij}E[M_{t+1}h_{n,t+1}|\mathcal{F}_{j,t}^a] = 1 \qquad (4.13)$$

For given x_t and s_{t+1}, the process $M_{t+1}h_{n,t+1}$ is conditionally log-normal, and exploiting this feature 4.13 implies

$$\sum_{j=0,1} \pi_{ij}E[M_{t+1}h_{n,t+1}|\mathcal{F}_{j,t}^a] = \sum_{j=0,1} \pi_{ij}\exp\{-r_{f,t} + \mu_{n,j,t} + \frac{\sigma_{n,j,t}^2}{2} + B_j(n-1)\lambda_j x_t\} = 1$$

$$(4.14)$$

Using the approximation $\exp(y) - 1 \approx y$, and $r_{f,t} = x_t$, 4.14 can be re-stated as

$$\sum_{j=0,1} \pi_{ij}[-x_t + \mu_{n,j,t} + \frac{\sigma_{n,j,t}^2}{2} + B_j(n-1)\lambda_j x_t] = 0 \qquad (4.15)$$

96

with $i = 0, 1$. Note that 4.15 depends on the current state i, which affects $\mu_{n,j,t}$ and the probabilities used in 4.15. Consequently, for each $s_t = i$ the analog for 4.15 must hold.

With discrete regime shifts, 4.15 is the asset pricing restriction on the risk premium. Here, the continuous bond returns, from the perspective of economic agents, are distributed as a conditional mixture of normals. In the standard case, bond returns are conditionally normal. The risk premium restriction, 4.15, also shows that agents anticipate the effects of regime shifts and are compensated for it. For example, if the probability of the high risk regime is greater when $s_t = 0$, then agents will demand a greater risk premium in regime $s_t = 0$, relative to when they are in regime $s_t = 1$. Further, if regimes are fairly persistent, then the effects of a regime switch on the risk premium and hence the term structure will be infrequent, but potentially large when a regime switch does happen. This feature, in addition to implied distribution for bond returns (and yields) distinguishes the regime switching model from the standard model discussed in the previous section.

Substituting the expression for $\mu_{n,j,t}$ and $\sigma^2_{n,j,t}$ (see 4.11 and 4.12) in 4.15, I can solve for the regime dependent coefficients of the bond price function;

$$\begin{bmatrix} B_0(n) \\ B_1(n) \end{bmatrix} = \begin{bmatrix} \pi_{00} & \pi_{01} \\ \pi_{10} & \pi_{11} \end{bmatrix} \begin{bmatrix} (1 - \kappa_0 - \lambda_0)B_0(n-1) - \frac{1}{2}\sigma_0^2 B_0^2(n-1) + 1 \\ (1 - \kappa_1 - \lambda_1)B_1(n-1) - \frac{1}{2}\sigma_1^2 B_1^2(n-1) + 1 \end{bmatrix} \quad (4.16)$$

and

$$\begin{bmatrix} A_0(n) \\ A_1(n) \end{bmatrix} = \begin{bmatrix} \pi_{00} & \pi_{01} \\ \pi_{10} & \pi_{11} \end{bmatrix} \begin{bmatrix} A_0(n-1) + \kappa_0\theta_0 B_0(n-1) \\ A_1(n-1) + \kappa_1\theta_1 B_1(n-1) \end{bmatrix} \quad (4.17)$$

with initial conditions $A_0(0) = A_1(0) = B_0(0) = B_1(0) = 0$. As these coefficients depend on which regime the economy is in, so does the term structure of bond prices. Note that bond price coefficients are mutually dependent on both the regimes— current bond prices anticipate the likelihood of regime shifts in the future. The

97

effects of the transition probabilities, as suggested by the above solution, are likely to be more pronounced at the short end of the term structure. Hence, one would suspect that the effects of regime shifts on the term structure are more evident at the short end of the term structure.

4.2.3 Extension to Multifactor Term Structure

Extensions of the standard model and the model with regime shifts to multiple factors is quite simple. For the two-factor model without regime-switching, I assume that there are two independent square-root processes $x_{1t+1} - x_{1t} = \kappa_1(\theta_1 - x_{1t}) + \sigma_1\sqrt{x_{1t}}u_{1t+1}$ and $x_{2t+1} - x_{2t} = \kappa_2(\theta_2 - x_{2t}) + \sigma_2\sqrt{x_{2t}}u_{2t+1}$. The risk-free interest rate is the sum of two factor returns $r_t^f = x_{1t} + x_{2t}$.[6] The pricing kernel is similar to the univariate case, except that the risk premium is now the sum of two square-root components. The bond price is log linear, $P(t,n) \equiv \exp\{-A(n) - B_1(n)x_{1t} - B_2(n)x_{2t}\}$ and using arguments similar to those used in the single factor case, it follows that

$$B_1(n) = (1 - \lambda_1 - \kappa_1)B_1(n-1) - \frac{1}{2}\sigma_1^2 B_1^2(n-1) + 1,$$

$$B_2(n) = (1 - \lambda_2 - \kappa_2)B_2(n-1) - \frac{1}{2}\sigma_2^2 B_2^2(n-1) + 1,$$

$$A(n) = A(n-1) + \kappa_1\theta_1 B_1(n-1) + \kappa_2\theta_2 B_2(n-1),$$

with boundary conditions $A(0) = B_1(0) = B_2(0) = 0$. The same argument applies to the case of three or larger independent square root processes.

In the case of regime-shifting, I assume that the regime shifts process independent of the factor square-root processes, and that the same regime shifts process applies to all square root factors. The bond prices are state dependent and log linear, $P_0(t,n) \equiv \exp\{-A_0(n) - B_{10}(n)x_{1t} - B_{20}(n)x_{2t}\}$, *if* $s_t = 0$ and $P_1(t,n) \equiv \exp\{-A_1(n) - $

---

[6]Subject to identification constraints, more flexible linear specification of the short rate process has been proposed in the literature (Dai and Singleton 2000, Duffie and Kan 1996)

98

$B_{11}(n)x_{1t} - B_{21}(n)x_{2t}\}$, if $s_t = 1$. The pricing kernel, as discussed above depends on the risk free rate, and the multiple factors with different $\lambda_{s_{t+1}}$ for each factor. The solutions are systems of vector differential equations, for factor k, is

$$
\begin{bmatrix} B_{k0}(n) \\ B_{k1}(n) \end{bmatrix} = \begin{bmatrix} \pi_{00} & \pi_{01} \\ \pi_{10} & \pi_{11} \end{bmatrix} \begin{bmatrix} (1 - \kappa_{k0} - \lambda_{k0})B_{k0}(n-1) - \frac{1}{2}\sigma_{k0}^2 B_{k0}^2(n-1) + 1 \\ (1 - \kappa_{k1} - \lambda_{k1})B_{k1}(n-1) - \frac{1}{2}\sigma_{k1}^2 B_{k1}^2(n-1) + 1 \end{bmatrix},
$$

and

$$
\begin{bmatrix} A_0(n) \\ A_1(n) \end{bmatrix} = \begin{bmatrix} \pi_{00} & \pi_{01} \\ \pi_{10} & \pi_{11} \end{bmatrix} \begin{bmatrix} A_0(n-1) + \kappa_{10}\theta_{10}B_{10}(n-1) + \kappa_{20}\theta_{20}B_{20}(n-1) \\ A_1(n-1) + \kappa_{11}\theta_{11}B_{11}(n-1) + \kappa_{21}\theta_{21}B_{21}(n-1) \end{bmatrix}
$$

with initial conditions $A_0(0) = A_1(0) = B_{10}(0) = B_{11}(0) = B_{20}(0) = B_{21}(0) = 0$. Note that the above solution restrictions apply to each factor $k = 1 \cdots K$, where K is the number of independent factors. Finally, the yields in the various models is computed from the given solution to the bond price equation. Yields in a K factor regime switching model are computed by using the relation,

$$
Y_s(t, n) = -\frac{\ln P_s(t, n)}{n} = \frac{A_s(n)}{n} + \sum_{k=1}^{K} \frac{B_{ks}(n)x_{kt}}{n}. \tag{4.18}
$$

It is worth noting that the yield curve restrictions, such as 4.18, can be derived from a representative agent based general equilibrium model of the type considered in Lucas (1978). In such an specification, the underlying aggregate consumption growth and inflation process would be subject to regime shifts.[7] Our empirical exercise, treats the consumption and inflation processes as latent factor processes.

4.3   Estimation by Efficient Method of Moments

In estimation the focus will be to evaluate if the different models can justify the observed behavior of two interest rates—the six month U.S. t-bill rate , and the five

---

[7]The implications for the yield curve from the general equilibrium exercise are the same as in 4.18. To keep the discussion compact I have not included this exercise in the paper, however, it is available from the authors upon request.

99

year U.S. t-bond rate. We explore the ability of up to a three factor benchmark model, and up to a two factor regime switching model to justify the observed conditional distribution of the two interest rates under consideration. To utilize a consistent approach for evaluation and estimation across the different models I rely on simulation based EMM (efficient method of moments) estimator, developed in Bansal et al. (1995), Gallant and Tauchen (1996b), Tauchen (1996), and Gallant and Tauchen (1998a).

The EMM estimator consists of two steps. First, the true conditional density of observed interest rates is approximated by a seminonparametric (SNP) series expansion, with the model dimension carefully chosen by an appropriate information criterion. Second, the score functions of the SNP density are used as moments, to be minimized using the simulation output from the structural economic model. In this sense the model is forced to match the conditional distribution of the observed 6-month and 5-year yields. This is a GMM-type estimator which in addition to providing comparable (across models) measures for specification tests, also permits a series of interesting diagnostics to understand the differences and shortcomings of the different models under consideration.

4.3.1 Projecting True Density onto Auxiliary Model

Let the invariant conditional density of the observed interest rate data generating process be the p-model. By assumption, direct maximum likelihood estimation of the p-model is not available. However, any smooth conditional density function can be approximated arbitrarily close by a Hermite polynomial expansion (see Gallant and Tauchen (1996b)). The empirical evidence and details regarding estimating a conditional density using this semi-nonparametric approach are discussed below.

Let y be the vector of interest rates under consideration, x be the vector of lagged

100

y. The auxiliary f-model has a density function defined by a modified Hermite polynomial,

$$f(y|x,\theta) = C\{[\mathcal{P}(z,x)]^2 \phi(y|\mu_x, \Sigma_x)\},$$

where $\mathcal{P}$ is a polynomial with degree K_z in z, which is a standardized transformation $z = R_x^{-1}(y - \mu_x)$ with $\Sigma_x = R_x R_x'$. The square of $\mathcal{P}$ makes the density positive and argument of the polynomial is z. The coefficients of the polynomial are allowed to be another polynomial of degree K_x in x. The constant in polynomial of z is set to 1 for identification. C is a normalizing factor to make the density integrating to one and $\phi(\cdot)$ is a normal density of y with conditional mean μ_x and conditional variance Σ_x, where μ_x is estimated by using a VAR, and Σ_x is estimated by using an ARCH specification, which parameterizes R_x. Note that both μ_x and R_x depend only on lags of y, that is the vector x.

The length of the auxiliary model (i.e., the SNP density) parameter is determined by the number of lags in the VAR mean specification, L_μ, lags in the ARCH specification L_r, lags of x used in constructing the coefficients of the polynomial L_p. The degree of the polynomial in z is K_z, and the degree for x is K_x. The specific choice for these tuning parameters, and consequently the SNP conditional density is done by relying on the BIC information criterion.

Let $\{\tilde{y}_t\}_{t=1}^n$ be the observed data, and $\tilde{x}_{t-1}$ be the lagged observations. The sample mean log likelihood function is defined by

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

where θ are the unknown parameters of this conditional density, which need to be estimated. A quasi-maximum-likelihood estimator is obtained by

$$\tilde{\theta}_n = \arg\max_\theta \mathcal{L}_n(\theta, \{\tilde{y}_t\}_{t=1}^n). \tag{4.19}$$

101

The dimension of the auxiliary f-model, the length of θ, is selected by the Schwarz's Bayesian Information Criterion (BIC)

$$BIC = s_n(\tilde{\theta}) + \frac{l_\theta}{2n} \log(n),$$

where $s_n(\tilde{\theta}) = -\mathcal{L}_n(\tilde{\theta}, \{\tilde{y}_t\}_{t=1}^n)$ is the negative maximized objective function. For greater details regarding SNP see Gallant and Tauchen (1989), Gallant and Tauchen (1992), Gallant and Tauchen (1997), and Gallant, Hsieh, and Tauchen (1997).

4.3.2 Matching Auxiliary Scores with Minimum Chi-Square

From the first stage seminonparametric estimates one obtained the fitted scores as the moment conditions,

$$m_n(\tilde{\theta}) = \frac{1}{n} \sum_{t=1}^n \frac{\partial}{\partial\theta} \log f(\tilde{y}_t | \tilde{x}_{t-1}, \tilde{\theta}).$$

In the second stage, a SMM-type estimator is implemented in the following way. Let $\{\hat{y}_t\}_{t=1}^N$ be a long simulation from a candidate value of ρ, the parameter vector of the maintained structural model. The auxiliary score functions can be re-evaluated at the simulated data,

$$\hat{m}_N(\rho, \tilde{\theta}) = \frac{1}{N} \sum_{t=1}^N \frac{\partial}{\partial\theta} \log f(\hat{y}_t | \hat{x}_{t-1}, \tilde{\theta}),$$

and the minimum chi-square estimator is simply minimizing quadratic objective function,

$$\hat{\rho}_n = \arg\min_\rho \{\hat{m}_N(\rho, \tilde{\theta})' \tilde{\mathcal{I}}^{-1} \hat{m}_N(\rho, \tilde{\theta})\}, \tag{4.20}$$

where the weighting matrix $\tilde{\mathcal{I}}^{-1}$ is estimated by the mean-outer-product of SNP scores

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

102

Under regularity conditions on the data generating process, the gradients of the moment condition and the information matrix are convergent. Let ρ^0 be the truth of the structural parameter and θ^0 the isolated solution of the moment condition $m(\rho^0, \theta^0) = 0$, then it is straightforward that $\lim_{n \to \infty} \hat{M}_n = M^0$ almost surely, where $M(\rho, \theta) = (\partial/\partial\rho')m(\rho, \theta)$ are the gradients of moment conditions. $\hat{M}_n = M(\hat{\rho}_n, \tilde{\theta}_n)$ and $M^0 = M(\rho^0, \theta^0)$ are just the sample and population counterparts. Given the population information matrix

$$I^0 = \int \int [\frac{\partial}{\partial\theta} \log f(y_t | x_{t-1}, \theta^0)][\frac{\partial}{\partial\theta} \log f(y_t | x_{t-1}, \theta^0)]' p(y_t, x_{t-1} | \rho^0) dx_{t-1} dy,$$

a standard argument leads to $\lim_{n \to \infty} \tilde{I}_n = I^0$ almost surely. Therefore we can derive the consistency and asymptotic normality results

$$\lim_{n \to \infty} \hat{\rho}_n = \rho^0 \text{ almost surely}$$

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

The asymptotic variance can be estimated by its empirical counterpart,

$$\widehat{VAR}(\hat{\rho}) = \frac{1}{n}[(\hat{M}_n)'(\tilde{I}_n)^{-1}(\hat{M}_n)]^{-1}. \tag{4.21}$$

The normalized criterion function value in the EMM estimation forms a specification test for the overidentifying restrictions

$$N \hat{m}_N(\rho, \tilde{\theta})' \tilde{\mathcal{I}}^{-1} \hat{m}_N(\rho, \tilde{\theta}) \sim \mathcal{X}^2(l_\theta - l_p), \tag{4.22}$$

with the degree of freedom equals $l_\theta - l_p$, the number of scores (i.e., moment conditions) in the p model less the number of structural parameters, l_p. It is assumed that l_p is smaller than l_θ.

Note that the distance matrix $\tilde{\mathcal{I}}$ used in constructing the $\mathcal{X}^2(l_\theta - l_p)$ specification test is identical across different structural model specifications (the null hypothesis).

103

Consequently, the p-values based on this specification test can be directly compared across different structural models to identify the best model specification.[8] It is well recognized in the literature, that tests for the absence of regime-shifts against a regime shift alternative require non-standard approaches (see Hansen (1992) and Garcia (1992)). The approach of comparing all the considered models to a common non-parametric density (the SNP density), allows us to rank order all the considered models according to the p-values implied by the EMM criterion function. The advantage of using the non-parametric SNP is that it can asymptotically converge to virtually any smooth distributions, including mixture distributions (as is the case with a model of regime shifts), as documented in Gallant and Tauchen (1998a). Further, Gallant and Tauchen (1998a) show EMM can provide more efficient estimates relative to alternative method of moment estimators such as simulated method of moments (see Duffie and Singleton (1993)).[9]

4.4 Empirical Results from EMM Estimation

4.4.1 Data Description

The data set ranging from June 1964 to December 1995, is obtained from the Center for Research in Security Prices (CRSP). There are total 379 monthly observations, with nine maturities 1, 3, 6, 9 month and 1, 2, 3, 4, 5 year. These discount bond yields were first constructed by Fama (1984) and Fama and Bliss (1987), and subsequently updated by CRSP. Longstaff and Schwartz (1992), also use this data set to estimate their term structure model.

It is important to recognize that the data period 1964-1995 contains different

---

[8] For a discussion of the importance of having the same distance matrix, for a consistent comparison across models, see Hansen and Jagannathan (1997).

[9] As is typical with GMM-type estimators, several Monte Carlo studies have documented the over - rejection bias of EMM (Chumacero 1997, Andersen et al. 1999c, Zhou 1999a), in small to moderate sample sizes, therefore the test is more conservative and cautious in model selection.

Federal Reserve monetary policies, which as stated earlier provides the economic motivation for incorporating regime shifts. From 1970-79 the Federal Reserve by and large pursued a policy of interest rate targeting though monetary targets were also taken into account. However, beginning from 1979, for a three year period, the Federal Reserve abandoned the policy of targeting interest rates and focused primarily on monetary targets. This policy, which saw very volatile interest rates, was abandoned in mid 1982 in favor of targeting interest rates and monetary aggregates. From 1993, the Fed's focus has primarily been on interest rates (see Froyen (1996) for greater details). One would suspect that when interest rate targets are pursued the interest rate volatility would be low, while targeting monetary quantities would lead to greater interest rate volatility. In general, there have been important shifts in monetary regimes which seem to affect the behavior of the term structure of interest rates.

The summary statistics of these monthly yields are give in Table 4.1. It is clear from the mean statistics that on average, the yield curve is upward sloping and concave. The standard deviation, positive skewness, and kurtosis are systematically higher for short maturities than for long ones.

To incorporate important time-series and cross-sectional aspects of term structure data I focus on a short term and a long term yield—the yield on the six month bill and the five year note. All the structural models, one, two, three factor with or without regime-switching are forced to match the conditional bivariate joint dynamics of these two yields. I did not use the one month or three month yield to represent the short end, because they are more likely to be influenced by liquidity needs (see Bansal and Coleman (1996) for such a model). The time series and first difference plots of the basis yields are reported in Figure 4.1. Many of the interesting dynamics of the two yields are recovered by the estimated SNP density, which I discuss next.

105

4.4.2 SNP Projection

The estimation results of auxiliary model are reported in Table 4.2 and 4.3. In the first step of EMM procedure, I tried to fit a seminonparametric (SNP) density of bivariate joint dynamics of the two basis yields. This SNP is a Hermite polynomial expansion of a normal density with the leading term as VAR-in-mean and ARCH-in-standard deviation. Table 4.2 reports the choice of SNP density and the BIC criterion, based on which I choose our preferred specification. The leading term of the bivariate SNP density, being conditionally normal, has 1 lag in the VAR based conditional mean ($L_\mu = 1$) and 5 lags in ARCH specification ($L_r = 5$). If all other terms in the SNP density where zeroed out, then the SNP density specification would indeed be this conditional normal density. However, the preferred specification allows for departures from conditional normality, in particular, it chooses a polynomial of order 4 ($K_z = 4$) in z; this choice leads to a "semiparametric ARCH" specification similar to that proposed by Engle and González-Rivera (1991). This specification allows for skewness and kurtosis in the SNP distribution. Note that in our preferred specification the interaction terms above degree 2 are suppressed ($I_z = 2$), and the coefficients of the polynomial do not depend on x, that is $K_x = 0$[10].

The total number of parameters for this semi-parametric ARCH specification is 28, that is $l_\theta = 28$. Hence the total number of scores that I have from this SNP density is also 28. The information criterion for choosing the preferred SNP specification is the minimum value of BIC provided in Table 4.2. Table 4.3 gives the parameter estimates of the SNP density. The preferred specification, for future reference is, $L_\mu = 1$, $L_r = 5$, $K_z = 4$, and $I_z = 2$. Figure 4.2 presents a plot of the bivariate density.

---

[10] When the $K_x = 0$, it also follows that the x-polynomial as coefficients of z-polynomial are all constants, hence these coefficients are not depending on lag state x, i.e., $L_p = 0$. For greater details see Gallant and Tauchen (1997).

106

It is fairly instructive to focus on some specific aspects of the estimated bivariate density. The Top panel in Figure 4.4 and Figure 4.5 gives the estimated conditional volatility of the 6-month and 5-year yields, which seems to be very persistent and fairly volatile. The short interest rate has a wide range for the conditional volatility and the volatility peaks around 1980. The range for the five year yield volatility is narrow, relative to that of the 6-month yield. The Top Panel in Figure 4.7, shows the conditional correlation between the 6-month and 5-year yields, the range for this correlation is from about 40% to above 80%—a wide range indeed. The most volatile period for yields, the early 80's sees, is associated with a considerable drop in the conditional correlation. The behavior of the conditional variance and the cross-correlation, as documented below, poses a serious challenge to the various term structure models under consideration.

4.5 Estimation Results and Diagnostics

4.5.1 EMM Specification Test Results

Table 4.4 gives the main EMM estimation results of for different models: one-factor square-root (1-Factor[CIR]), one-factor regime-switching (1-Factor[RS]), two-factor square-root (2-Factor[CIR]), two-factor regime-switching (2-Factor[RS]), and three-factor square-root (3-Factor[CIR]). The result reported here are for simulation size of 75,000. The square-root factors are generated monthly, and the bond price formula given in 4.18 is applied with one month being the time interval—consequently, the five year bond is treated as a 60 month bond (similarly for the 6-month bond). The data that was used for estimation was the monthly yield multiplied by 12, analogously the simulated monthly yields are also multiplied by 12.

The single factor CIR model (1-Factor[CIR]) is sharply rejected with p-value less than 0.0000. The magnitudes and signs of the parameter estimates are close to those

107

reported in the literature. Keep in mind that the long-run mean θ, local variance σ, and risk premium λ should be multiplied by 12 to match the annualized model. The mean reversion κ is not affected by sampling frequency, and is very persistent as typically reported. All the parameter estimates are significant.

The one factor model with regime-shifts is also rejected with a p-value less than 0.0000. It is worth noting the regimes do not differ much in the mean reversion coefficient or the long run mean. The key difference is in the volatility; regime 0 volatility σ_0, is more that twice that of in the low volatility regime, regime 1. In addition, the risk-premium parameter λ is larger in the high volatility regime. All the parameters of this specification are estimated fairly accurately.

Even without regime-switching, introducing a second factor make the test statistics down to 62.376 with p-value smaller than 0.0000. Since the factors additively determine the short rate, the long run mean levels are about half of the one factor model. The first factor has higher mean reversion and a larger variance parameter, relative to the second factor. Other than the risk premium parameter for the first factor, all parameter estimates are significant. The risk premium parameter for the first factor is not significantly different from zero. The inclusion of a second factor provides a significant improvement over the one factor benchmark model, though relative to the regime shifts one factor model, the improvement at best is very small.

The best model amongst all specification is the two-factor regime switching specification. This specification finds considerable support in the data, as the p-value for this specification is about 15%, which is well above conventional levels of significance. In terms of the parameters of interest, note that the first factor, for both regimes (i.e., 0 and 1), relative to the second factor has far greater mean reversion. For both regimes, the first factor is also more volatile relative to the second factor. The regime 1 volatility parameters is larger for both the factors, and in this sense it represents

108

the more volatile regime. Additionally, the risk premium parameter, for both factors, is larger for regime 1, the more volatile regime. Given the nature of the mean reversion, it seems that the second factor represents the level factor and the first, the slope factor. The regime probabilities (standard error) are $\pi_{11} = 0.91$ (0.22) and $\pi_{00} = 0.94$ (0.18). All the parameters of the model are estimated fairly accurately. These transition probabilities reported for the 2-Factor[RS] specification, are comparable to those found in other studies, for example, 0.99 and 0.97 in Gray (1996), 0.99 and 0.91 in Hamilton (1988), and 0.99 and 0.94 in Cai (1994).

The 2-Factor[RS] model can be viewed as the 3 factor model with the regime switching factor being multiplicative or nonlinear switching process. For a fair comparison of this two factor regime switching model, I also estimate a three factor CIR model. As Table 4.4 shows, 3-Factor[CIR] model is also sharply rejected with chi-square being 45.607 and p-value of 0.0001. All the parameter estimates are significant, and all factors are associated with a positive risk premium. The first factor has a high long run mean, relatively high mean reversion of 0.033, and a relatively smaller variance. The other two factors have greater persistence and volatility. Our empirical evidence suggests that the standard three factor CIR model cannot account for the observed behavior of the short and long yields used in this empirical exercise.

Using the estimated parameter values and the variance-covariance matrix, one

109

can perform the t-test for some hypothesis of interest.[11] These test statistics suggest that the regime shifts are not an *iid* mixture ($t_{\pi_{00}=1-\pi_{11}} = 2.1507$).

The first factor has significant regime switches in level, mean reversion, and volatility parameters—the t-ratios are $t_{\theta_{10}=\theta_{11}} = 7.4972$, $t_{\kappa_{10}=\kappa_{11}} = 2.7699$, $t_{\sigma_{10}=\sigma_{11}} = 45.9025$, respectively. The risk premium of the regime switching factor, as measured by the difference of the risk premiums across the two regimes, is also sizable with $t_{\lambda_{10}=\lambda_{11}} = 35.1633$. Evidence for significant differences in regimes in the second factor is muted, with the t-ratios for the level, mean reversion and volatility parameters, being $t_{\theta_{20}=\theta_{21}} = 1.7521$, $t_{\kappa_{20}=\kappa_{21}} = 5.6031$, $t_{\sigma_{20}=\sigma_{21}} = 3.4267$, respectively. Further, the difference in the risk-premium parameter across regimes is not significantly different from zero as the t-ratio, $t_{\lambda_{20}=\lambda_{21}} = 0.3763$. In all there seems to be greater differences across regimes in the first factor, which as documented later in the paper, is intimately related to volatility or spread.

A recent study by Dai and Singleton (2000) found that a three-factor affine process model, similar to that proposed by Chen (1996), passes the standard statistical tests of over-identifying restrictions. The Dai and Singleton (2000) uses swap yields data after 1987 for their empirical exercise. By confining to the swap interest rate data beginning in 1987, much of the information regarding some of the most volatile periods of interest rates, e.g., 1979-1982, and in 1973-76, does not bear on the esti-

---

[11] For example, the t-test on whether the regime shifts are *iid* mixtures can be formulated as

$$H_0 : \pi_{00} = 1 - \pi_{11},$$

$$t = \left| \frac{\hat{\pi}_{00} + \hat{\pi}_{11} - 1}{\sqrt{\mathrm{var}(\hat{\pi}_{00}) + \mathrm{var}(\hat{\pi}_{11}) - 2\mathrm{cov}(\hat{\pi}_{00}, \hat{\pi}_{11})}} \right|,$$

and the test on whether the two regimes are equally persistent can be formulated as

$$H_0 : \pi_{00} = \pi_{11},$$

$$t = \left| \frac{\hat{\pi}_{00} - \hat{\pi}_{11}}{\sqrt{\mathrm{var}(\hat{\pi}_{00}) + \mathrm{var}(\hat{\pi}_{11}) - 2\mathrm{cov}(\hat{\pi}_{00}, \hat{\pi}_{11})}} \right|.$$

The formulas for testing other parameter restrictions across regimes are similar.

mation. The sample from 1964-1995, has several features that are missing from the post 1988 period. Indeed, the most challenging obstacle in term structure modeling, as pointed out by Brown and Dybvig (1986), is to account for the apparent regime shift of Federal Reserve monetary policy during 1978-1982. Further, as shown by Bansal et al. (1998) the model specification used by Dai and Singleton (2000) when applied to US Treasury yields, for the sample period of 1964-1995 is sharply rejected. In addition, as shown in this paper, a related three factor CIR model is also sharply rejected. In all, our specification tests provide considerable support for the two factor, regime switching model. All the benchmark CIR model specification (with up to three factors) are sharply rejected.

4.5.2 Diagnostics for Different Model Specifications

In this section I provide a range of diagnostics which allow us to understand the strengths and weaknesses of the different model specifications. In particular, these diagnostics reveal why the 2-Factor[RS] specification finds considerable empirical support, while the other specifications discussed above, do not. The diagnostics include looking at the t-ratios for the fitted scores, the re-projected density with its implications for the conditional volatility and cross-correlation for observed yields, and a comparison of the model implied term structure relative to that seen in the data.

For a reasonable structural model, all the 28 scores (moment conditions) used to estimate the model parameters should be close to zero. Table 4.5 reports the t-tests for the 28 moment conditions for the various models. These 28 scores (moment conditions), should for a reasonable model specification, should be close to zero. The size of the t-ratios gives a sense of the scores which different model specifications have difficulty is fitting.

For reference note that for the Hermite polynomial part, $a(i,j)$ refers to the

111

parameter before the polynomial term with i-th degree of power on the first variable (6 month yield) and j-th degree of power on the second variable (5 year yield). For the VAR mean part, $\mu(i,j)$ represents the parameter on variable i ($i = 1,2$) with j lag. For the ARCH standard deviation part, $R(i,j)$ denotes the parameter on variable i with j lags, and the last term $R(3)$ is simply the constant covariance parameter. If the structural model under consideration fits the particular moment under consideration, then at conventional 5% levels of significance the t-ratio should be smaller than 1.96.

Overall it is clear that introducing additional factors improves the model's ability to match the scores. In the context of the benchmark model, the three factor (see column 3-Factor[CIR]) model provides considerable improvement over the one (1-Factor[CIR]) and two factor (2-Factor[CIR]) specifications. The t-ratios for almost all scores are relatively smaller for the three factor specification. However, in an absolute sense the three factor model (3-Factor[CIR]) has considerable difficulty is matching many of the Hermite polynomial scores and the ARCH specification scores. This suggests that the benchmark three factor model cannot capture the conditional skewness, kurtosis, and the second moment properties of the bivariate interest rate process. For the most preferred 2-Factor[RS] specification, all the moments in the conditional mean, all but one in the Hermite polynomial, all but four in the conditional standard deviation, are fitted. The magnitude of the highest t-ratio is 3.13 (see ARCH, $R(2,5)$), which relative to the size of the t-ratio for other specifications is quite small. Overall, this preferred specification also seems to have the greatest difficulty is matching the conditional volatility and correlation, i.e., the ARCH scores—though, it provides considerable improvement over all other models. The 2-Factor[RS] specification provides large gains in fitting the non-Gaussian polynomials (i.e., the Hermite parameters), suggesting that conditional skewness and kurtosis in the standardized innovations can be accounted for by this term structure model.

112

4.5.3 Reprojected Density and Conditional Second Moments

The idea behind the reprojection technique (Gallant and Tauchen 1998b), is to characterize the dynamics of a given vector of observed variables (i.e., observed yields) conditional on its lags. In models where there are latent factors or regimes, the re-projection density provides a way to characterize the conditional density strictly in terms of observables (i.e., the yields). The re-projected density can be estimated by relying on simulated data for the yield series from a given estimated structural model. In our context, the reprojected density is the bivariate conditional density for the two yields under consideration, the 6-month and the 5-year yield.

Figure 4.3 compares the reprojected densities of marginal 6 month and 5 year yields with their projected ones. Given the estimated model and the simulated output for yields, the re-projected conditional density is estimated by re-estimating the parameters of the SNP density, using the same specification as was used to characterize the bivariate density of the 6-month and 5-year yields earlier (referred to as the unrestricted conditional density). Indeed, an reasonably good term structure model would imply a re-projected conditional density which is very similar to the unrestricted conditional bivariate density for the 6-month and 5-year yields. Once the re-projected density is estimated, specific moments, such as the conditional variances and correlations, implied by the model specification can be computed. These conditional moments are simply continuous functions of the conditioning information (i.e., lagged 6-month and 5-year yields) used to estimate the re-projected density. Given the conditioning information, the implications of a given model for any conditional moment of interest can be tracked down in the data and compared to the conditional moment implied by the unrestricted SNP distribution.

It is worth noting that the re-projected density, in our context, is the density for observed data estimated from a null model. For some specification, such as the three

113

factor CIR model, or the two factor regime switching model, the conditional density for yields cannot be characterized by using only two yields (the 6-month, and 5-year yields). Hence the re-projected density provides a convenient and comparable benchmark to characterize the conditional density for the two yields under consideration, across different null models.

Figure 4.3 provides the comparison of re-projected density of the different models to the bivariate SNP conditional distribution estimated earlier (referred to as the unrestricted density). All the conditional densities are evaluated at the sample mean. The unrestricted 6 month SNP density has high peak and right tail relative to a normal density with same mean and variance. The 1-Factor[CIR] model over fits the skewness, and the 1-Factor[RS] model over fits the two tails. Both 2-Factor[CIR] and 3-Factor[CIR] model have relatively high peakedness and the variances are relatively small. Our preferred 2-Factor[RS] model has a re-projected density that is fairly close to that observed in the data.

Turning to the 5 year yield, the unrestricted SNP density has a lower peak and skewed to the left, relative to the normal distribution. The 1-Factor[CIR], and the 1-Factor[RS] model seriously fail to capture the observed density in the data. The peak of the 2-Factor[CIR] is too low, and its variance too high. The 3-Factor[CIR] model is slightly better, but the variance relative to the data is still small. Only the 2-Factor[RS] model gets the peak, tails, skewness, and kurtosis correct. Overall, the one factor models over fit the variance and the symmetry, while the multifactor models under fit the variance and skewness, except for the 2-Factor[RS] model.

Figures 4.4 and 4.5 compare the conditional volatility (the plot provides standard deviation) for the various model specifications—this reveals some sharp and important differences across models. First note that, in the data, the process for conditional variance for the 6-month yield is quite different from that of the 5-year

114

yield. The range for the conditional volatility for the 6-month yield rate is much larger than for the 5-year yield—the high end being almost three times the lowest for the 6-month, and two times for the 5-year yield. The short yield volatility is more persistent, while the long yield volatility seems more choppy. The single factor square-root model misses entirely the dynamics feature of the conditional variance. The near constant conditional volatility was also reported by Gallant and Tauchen (1998b), when a univariate square-root model is fitted to a single interest rate series. It should be pointed out that introducing regime-switching to the single factor model, does not provide visible improvement in capturing the term structure volatility. In the two factor single regime model, the short yield volatility is only marginally better than the no-regime case, while the long yield volatility has significant improvement. In our fitted 2-Factor[RS] model, the reprojected volatility mimics the dynamics of the projected volatility extremely well. The long yield volatility almost completely matches that observed in the data. Finally the three factor CIR model seems to capture general shape of the volatility much better than the 2-Factor[CIR] model, however, in an absolute sense it does not capture the volatility dynamics as well as the 2-Factor[RS] specification.

Figure 4.6 and 4.7 provide evidence regarding the conditional covariance and correlation between the 6-month and 5-year yield. Again, it is very clear that only the 2-Factor[RS] model succeeds in capturing the wide range of the covariance and correlation observed across these yields. The correlation varies from 40% to 80%. Note that while the conditional covariance increases during the very volatile period of the early 80's, the correlation decreases—suggesting that the volatilities of the two yields rise more rapidly relative to the conditional covariance. The 3-Factor[CIR] model fails to capture the range or the conditional correlation, the 2-Factor[RS] model comes surprisingly close to capturing virtually all the observed dynamics of

115

the conditional correlation between these yields.

Based on all the diagnostics and the specification tests it seems that regime shifts are an important aspect of the behavior of the term structure. In particular, they are quite important to account for the conditional volatility and cross-correlation across yields, along with the conditional higher moments such as skewness and kurtosis.

4.6   Cross-Sectional Implications and Regimes

In this section I document the cross-sectional differences across models—that is, the ability of different specifications to reproduce the observed yield curve at each date in the data. To avoid clutter I focus only on the empirically plausible specifications— the 2-Factor[CIR] model, 3-Factor[CIR] model, and the preferred specification, the 2-Factor[RS] specification.

To derive the model implied term-structure for each date, in the context of the 2-Factor[CIR] and 3-Factor[CIR] models, I simply recover the latent factors, by using the bond pricing function, and the estimated parameters,

$$\mathbf{X_t} = \mathbf{B}^{-1}[\mathbf{Y_t} - \mathbf{A}]. \tag{4.23}$$

The K vector of latent factors $\mathbf{X_t}$, can be recovered by using K observed basis yields, the estimated $K \times K$ matrix of bond pricing parameters $\mathbf{B}$, and the vector $\mathbf{A}$. Given the vector $\mathbf{X_t}$, and the estimated parameters, all the entire yield curve can be computed for each date in the data. By construction, the computed yield curve will pass through the basis yields. This approach to extract the latent factors is similar to that used in Chen and Scott (1993), and Duffie and Singleton (1997), amongst others. Note that in the 2-Factor[CIR] model, the computed yield curve will pass through two basis yields and in the 3-Factor[CIR] model through three basis yields.

To compare models, for each date I also compute the absolute average cross-sectional pricing error, with $\hat{Y}(n)_t$ being the computed yield at date t for a given

116

model specification,

$$PE_t = \frac{\sum_{n=1}^{N} |\hat{Y}(n)_t - Y(n)_t|}{N} \qquad (4.24)$$

Smaller the pricing error, better is the model specification. The pricing errors will arise due to model mis-specification.[12]

To compute the yield curve for each date, under the null specification of 2-Factor[RS], one also needs to know the regime at each date. Recall, the yield curve, in addition to the two latent factors, also depends on the latent (i.e., to the econometrician) regime. To deal with this, I first extract the two latent factors (i.e., $K = 2$) using 4.23, the estimated $\mathbf{B}_{s_t}$ and $\mathbf{A}_{s_t}$, for $s_t = 0$ and $s_t = 1$. Given the latent factors and the estimated parameters, I compute two potential yield curves for each date, one for $s_t = 0$ and another for $s_t = 1$. For each regime, it is straight-forward to compute the pricing error PE_t, for $s = 0, 1$. However, under the null of the 2-Factor[RS] model specification, the pricing error (i.e., PE_t) associated with the regime that truly prevails in the economy at date t will be close to zero [13]. Hence, the regime that prevails at date t is simply the one which has the smallest pricing error across regimes. With the knowledge of the regime at date t, the model implied yield curve and the pricing error also become available. Note that the computed yield curves and the pricing errors for 3-Factor[CIR] and the 2-Factor[RS] can be fairly compared—the 3-Factor[CIR] specification uses three basis yields, whereas the 2-Factor[RS] specification relies on two basis yields and an identification procedure for the regime classification.

Table 4.7 provides information of the average pricing error (i.e., $(\sum_t^T PE_t)/T$). It is clear from the sample statistics that the 2-Factor[RS] model has the smallest

---

[12]An additional source could be measurement errors in the data.

[13]Note that this procedure under the null of the model, will detect the regime correctly. For example, in our simulations of the 2-Factor[RS] model, this procedure would indeed identify the correct regime—with the pricing error for the true regime being zero.

average pricing error and also the smallest standard deviation in the pricing error. The maximal pricing error associated with the 2-Factor[RS] classification is also the smallest. Further, on average the pricing error is only about 23 basis points for the annualized percentage yields—a small number indeed. The 3-Factor[CIR] specification has a average pricing error of 25 basis points, which in an absolute sense is also quite small. Fig 10 provides the computed yield curves for dates which are 30 months apart—these graphs re-enforce the view that the 2-Factor[RS] specification is the preferred one. The earlier evidence from our specification tests and the re-projection based diagnostics showed that the 2-Factor[RS] is by far the preferred model—this view is further corroborated by the cross-sectional evidence.

Figure 4.9 provides the results from the regime classification. Most of the time, it seems that the economy is in regime 1. The total number of regime switches recovered from the sample period is 44, which is larger enough for identifying the regime probabilities. The regime classification by and large coincides with the NBER business cycles. It seems that regime 0, obtains during or around recessions in the economy. The relatively frequent regime shifts between 1991-1995, shown in Fig 9, reflects almost identical average pricing error across regimes at those dates. Virtually identical pricing error across regimes for a given date implies that the data is not particularly informative regarding the regime classification.

The estimation results discussed earlier suggested that the first factor has considerable differences, especially in volatility, across regimes. Figure 4.8 shows that the recovered first factor tracks the short yield very well, while the second tracks the long yield. The correlation between the NBER recession (recession is regime 0 and boom is regime 1) indicator and recovered regime indicator is 0.1523, excluding the early 1990s where regime identification is poor (the analogous correlation for the entire sample is 0.0762). Correlation between NBER recession indicator and the

118

yield spread (5 year yield minus 6 month yield) is as high as 0.2342. Correlation between the model based regime indicator and yield spread is 0.2299–that is, high volatility regimes have high yield spreads. In general, considerable caution should be exercised in interpreting the regime classification as there is considerable error in this classification due to parameter uncertainty and model misspecification.[14]

Table 4.8 provides some simple statistics relating to the 6-month and 5-year yields, obtained from the simulated yields from the preferred 2-Factor[RS] model. The different statistics are computed by regime $s = 0, 1$. The sample counterparts of the various statistics are computed by relying of the regime classification, discussed above, and the observed yield data. The absolute value of first difference in the yield (a measure of the ex post standard deviation) and the average yield spread (slope) is higher in the more volatile regime $s = 1$—both in the population (based on simulated data from the model) and in the sample. The mean is not very different across the two regimes. The average convexity,[15] is larger in regime 0. Figure 4.11 plots the yield curves for different combinations of the state variables,[16] and in each case the slope of the yields curve is larger in the more volatile regime (i.e., regime 1). This suggests that the main observable differences across regimes lie in the yields spread and volatility of the yields. Considerable caution should be exercised in interpreting the sample statistics as they are based on a regime classification which potentially contains considerable error.

The differences in volatility across regimes is consistent with the economic implications for volatility of interest rates when the monetary policy switches from one of

---

[14]The fact that there is positive pricing error for all dates in both regimes suggests that the regime switching model is misspecified, from the cross-sectional perspective.

[15]The convexity is computed as $[(Y_t^{3yr} - Y_t^{2yr}) - (Y_t^{2yr} - Y_t^{1yr})]$, which is an discrete approximation to the second derivative or convexity taking the time interval to be one.

[16]The state variables are the x's. Note that in different combinations of the two x's, the sum equals to the risk-free rate. All the chosen combinations imply the same risk-free rate. This is done for comparability.

119

primarily targeting money growth rather than short term interest rates. As discussed earlier, in 1979, monetary policy in the U.S. did switch to targeting money growth, and then in 1982 switched back to targeting primarily interest rates. These episodes, along with other less prominent policy switches are a part of the yield data used in our empirical exercise.

4.7 Concluding Remarks

There is considerable statistical evidence regarding the presence of regime shifts in the short interest rate data (see Hamilton (1988, 1996)). In addition, there are economic reasons as well to believe that interest rates are subject to regime shifts. The conduct of monetary policy has first order effects on the term structure, and as has been well documented, is subject to discrete changes is regimes(see Froyen (1996)). Despite the potential for important effects from discrete regime shifts, received term structure models, such as the Cox et al. (1985a) model do not incorporate them. The absence of this important component in these models may well explain their poor empirical performance (see Brown and Dybvig (1986), etc.).

The main contribution of this paper is to show that an internally consistent model of the term structure, which incorporates regime shifts, provides significant improvements over multi-factor versions of the CIR model. More specifically, I develop and estimate a model for the term structure which permits regime shifts. One can show, that a model which incorporates regime shifts is essential to account for the conditional joint dynamics (i.e., the conditional distribution) of short and long yields. For comparison with an commonly used model, I also estimate multi-factor versions of the CIR model. The empirical exercise is conducted by relying on nominal U.S. treasury bill and bond yields from 1964-1995. For estimation, and specification tests, of the various models I use the Efficient Method of Moments technique developed

120

in Bansal et al. (1995) and Gallant and Tauchen (1996b). I also provide a battery of diagnostics to evaluate the various model specifications, in particular, I rely on the re-projection method of Gallant and Tauchen (1998b), to recover the conditional density for yields (conditional on lagged yields) to evaluate the merits of the different model specifications.

The empirical evidence shows that standard Cox-Ingersoll-Ross term structure models, with up to three factors are sharply rejected in the data—these models cannot account for the conditional volatility and cross-correlation across yields observed in the data. This standard model also fails to account for the skewness and kurtosis found in the interest rate data. A two factor regime switching model finds considerable support in the data and does surprisingly well in accounting for the conditional volatility and cross-correlation of the short and long yields. Further, the absolute pricing error in reproducing the observed yield curves at different dates in the data is smallest for the two factor regime-switching model. This preferred model seems to capture the conditional distribution of yields fairly well. The evidence shows that there is an intimate link between business cycles and regimes extracted from the term structure model—the main differences across extracted regimes is in the short rate volatility and the yield spread.

4.8   Tables and Figures

121

Table 4.1: Summary Statistics of Monthly Yield Data
There are 379 monthly observations of the yields with nine maturities. The data is obtained from CRSP (Center for Research in Security Prices) Fama (1984) and Fama and Bliss (1987) bond files, ranging from June 1964 to December 1995.

Maturity	1mn	3mn	6mn	9mn	1yr	2yr	3yr	4yr	5yr
Mean	0.0645	0.0672	0.0694	0.0709	0.0713	0.0734	0.0750	0.0762	0.0769
Std Dev	0.0265	0.0271	0.0270	0.0269	0.0260	0.0252	0.0244	0.0240	0.0237
Skewness	1.2111	1.2118	1.1518	1.1013	1.0307	0.9778	0.9615	0.9263	0.8791
Kurtosis	4.5902	4.5237	4.3147	4.1605	3.9098	3.6612	3.5897	3.5063	3.3531
Minimum	0.0265	0.0277	0.0287	0.0299	0.0311	0.0366	0.0387	0.0397	0.0398
05% Qntl	0.0313	0.0337	0.0356	0.0371	0.0376	0.0403	0.0427	0.0446	0.0447
25% Qntl	0.0457	0.0485	0.0510	0.0530	0.0537	0.0546	0.0572	0.0585	0.0599
50% Qntl	0.0579	0.0599	0.0634	0.0650	0.0666	0.0691	0.0709	0.0721	0.0735
75% Qntl	0.0782	0.0806	0.0821	0.0836	0.0830	0.0851	0.0857	0.0858	0.0861
95% Qntl	0.1197	0.1264	0.1294	0.1311	0.1286	0.1278	0.1298	0.1271	0.1272
Maximum	0.1640	0.1612	0.1655	0.1644	0.1581	0.1564	0.1556	0.1582	0.1500

Table 4.2: SNP Score Generator

The SNP score generator has a leading ARCH term with L_μ lags in conditional mean, L_r in conditional standard deviation. The standardized innovation has a normal density stretched by a squared Hermite polynomial with degree of K_z. Since it is a bivariate SNP density, the interaction polynomial term above the I_z degree is suppressed as zero. Similarly the coefficient of the z-polynomial may depend on the lagged history through a K_x degree polynomial with the interaction term above I_x degree being suppressed as zero. The lag order L_p on x-polynomial is inoperative if $K_x = 0$. The total number of parameters is l_θ. The BIC preferred choice is 1514200. The last two columns report the minimized objective function

$$s_n(\tilde{\theta}) = -\frac{1}{n} \sum_{t=1}^{n} \log[f(\tilde{y}_t | \tilde{x}_{t-1}, \tilde{\theta})]$$

and BIC information criterion

$$BIC = s_n(\tilde{\theta}) + \frac{l_\theta}{2n} \log(n).$$

	L_μ	L_r	L_p	K_z	I_z	K_x	I_x	l_θ	$s_n(\theta)$	BIC
	1	0	1	0	0	0	0	9	-0.79431	-0.72073
	2	0	1	0	0	0	0	13	-0.82011	-0.71383
	3	0	1	0	0	0	0	17	-0.82763	-0.68865
	1	1	1	0	0	0	0	11	-0.96388	-0.87395
	1	2	1	0	0	0	0	13	-0.98904	-0.88276
	1	3	1	0	0	0	0	15	-1.01872	-0.89609
	1	4	1	0	0	0	0	17	-1.07256	-0.93358
	1	5	1	0	0	0	0	19	-1.09188	-0.93656
	1	6	1	0	0	0	0	21	-1.09328	-0.92160
	1	7	1	0	0	0	0	23	-1.09922	-0.91119
	1	5	1	4	3	0	0	27	-1.14488	-0.92416
→	1	5	1	4	2	0	0	28	-1.17202	-0.94311
	1	5	1	4	1	0	0	30	-1.17230	-0.92704
	1	5	1	4	0	0	0	33	-1.20102	-0.93124
	1	5	1	5	4	0	0	29	-1.15141	-0.91433
	1	5	1	5	3	0	0	30	-1.17326	-0.92801
	1	5	1	5	2	0	0	32	-1.18655	-0.92494
	1	5	1	5	1	0	0	35	-1.21327	-0.92714
	1	5	1	5	0	0	0	39	-1.21690	-0.89807
	1	5	1	6	5	0	0	31	-1.17734	-0.92391
	1	5	1	6	4	0	0	32	-1.18399	-0.92239
	1	5	1	6	3	0	0	34	-1.18758	-0.90962
	1	5	1	6	2	0	0	37	-1.19769	-0.89521
	1	5	1	6	1	0	0	41	-1.22820	-0.89302
	1	5	1	6	0	0	0	46	-1.23516	-0.85910
	1	5	1	4	2	1	0	48	-1.24496	-0.85256
	1	5	1	4	2	2	1	68	-1.28923	-0.73332
	1	5	1	4	2	2	0	78	-1.30853	-0.67087

123

Table 4.3: Parameter Estimates of Projected SNP Density

The SNP density is described in Section 4.3.1, and the specification search is reported in Table 4.2. The parameter in the Hermite polynomial function $a(i, j)$ stands for the term with i'th power on the short yield and j'th power on the long yield. The parameter in the conditional mean function is respectively $\mu(1, 0)$ constant in short yield, $\mu(1, 0)$ in long yield, $\mu(1, 1)$ lag one short yield in short yield equation, $\mu(2, 1)$ lag one long yield in short yield equation, $\mu(1, 2)$ lag one short yield in long yield equation, and $\mu(2, 2)$ lag one long yield in long yield equation. The parameter in ARCH standard deviation function is $R(k, l)$ is the short yield ($k = 1$) or long yield ($k = 2$) with lag equals to l. $R(3)$ is the constant off-diagonal term, a partial contribution to the covariance. The negative sample mean log-likelihood s_n and BIC information criterion are the same as Table 4.2.

Parameter	Estimate	Standard Error
Hermite $a(0, 0)$	1.00000	(0.00000)
$a(0, 1)$	-0.03861	(0.08103)
$a(1, 0)$	0.49154	(0.10908)
$a(0, 2)$	-0.13739	(0.06996)
$a(1, 1)$	0.17944	(0.07162)
$a(2, 0)$	-0.00415	(0.09781)
$a(0, 3)$	0.01735	(0.01406)
$a(3, 0)$	-0.06015	(0.04077)
$a(0, 4)$	0.02044	(0.00742)
$a(4, 0)$	-0.01098	(0.01556)
Mean $\mu(2, 0)$	-0.09681	(0.01521)
$\mu(1, 0)$	-0.02068	(0.01273)
$\mu(2, 2)$	0.95940	(0.02201)
$\mu(2, 1)$	0.02563	(0.01576)
$\mu(1, 2)$	-0.01187	(0.01954)
$\mu(1, 1)$	0.01529	(62.097)
ARCH $R(1, 0)$	0.04595	(0.01034)
$R(2, 0)$	0.08426	(0.01093)
$R(3)$	0.19853	(0.07145)
$R(1, 1)$	0.10511	(0.01548)
$R(2, 1)$	0.12883	(0.04020)
$R(1, 2)$	-0.01868	(0.05845)
$R(2, 2)$	0.13730	(0.04346)
$R(1, 3)$	0.02818	(0.06784)
$R(2, 3)$	0.07998	(0.04836)
$R(1, 4)$	0.25753	(0.07586)
$R(2, 4)$	0.12819	(0.05029)
$R(1, 5)$	0.03728	(0.05829)
$R(2, 5)$	0.22468	(0.05443)
Spec = s1514200	s_n = -1.17202	BIC = -0.94311

124

Table 4.4: Model Estimation by Efficient Method of Moments
The five term structure models are laid out in Section 4.2. The simulation size in EMM is 75,000 for all the five models.

	1-Factor[CIR]	1-Factor[RS]	2-Factor[CIR]	2-Factor[RS]	3-Factor[CIR]
Factor 1 Regime 0					
θ_{10}	.00612(.00006)	.00526(.00009)	.00296(.00025)	.00437(.00003)	.00348(.00006)
κ_{10}	.01638(.00057)	.01644(.00068)	.02888(.00392)	.03388(.00199)	.03309(.00127)
σ_{10}	.00404(.00003)	.00544(.00012)	.00424(.00014)	.00421(.00002)	.00328(.00007)
λ_{10}	-.00578(.00047)	-.00337(.00049)	.00157(.00354)	.01155(.00112)	-.00254(.00221)
Factor 1 Regime 1					
θ_{11}		.00589(.00026)		.00273(.00001)	
κ_{11}		.01536(.00115)		.04375(.00136)	
σ_{11}		.00224(.00006)		.00629(.00002)	
λ_{11}		-.01284(.00068)		-.03145(.00086)	
Factor 2 Regime 0					
θ_{20}			.00171(.00015)	.00076(.00001)	.00051(.00002)
κ_{20}			.00489(.00015)	.01501(.00019)	.03031(.00190)
σ_{20}			.00303(.00018)	.00392(.00003)	.00698(.00033)
λ_{20}			-.02157(.00024)	-.02079(.00005)	-.01658(.00233)
Factor 2 Regime 1					
θ_{21}				.00540(.00144)	
κ_{21}				.00344(.00102)	
σ_{21}				.00409(.00004)	
λ_{21}				-.02287(.00075)	
Factor 3 Regime 0					
θ_{30}					.00050(.00001)
κ_{30}					.02377(.00359)
σ_{30}					.00452(.00014)
λ_{30}					-.04793(.00327)
Transitional Probability $Pr\{s_{t+1}|s_t\}$					
π_{00}		.97564(.00002)		.94990(.18118)	
π_{11}		.94489(.00000)		.91993(.22329)	
Specification Test					
$\mathcal{X}^2$	132.017	72.245	62.376	14.717	45.607
p	0.0000	0.0000	0.0000	0.1427	0.0001
d.o.f.	24	18	20	10	16

125

Table 4.5: Diagnostic T-Ratios

The score generator is s1514200 as reported in Table 4.2, and the parameters have same explanations as in the note of Table 4.3. The T-ratios are testing whether the fitted sample moments are equal to zero, as predicted by by population moments of the SNP density.

Parameter	1-Factor[CIR]	1-Factor[RS]	2-Factor[CIR]	2-Factor[RS]	3-Factor[CIR]
Hermite $a(0,1)$	-0.084	1.708	-0.792	-0.583	-1.019
$a(1,0)$	-2.601	-2.309	-1.395	0.432	-0.784
$a(0,2)$	5.275	2.909	3.712	1.704	3.312
$a(1,1)$	-2.607	-0.776	1.247	1.041	2.868
$a(2,0)$	5.667	4.116	5.284	2.557	5.117
$a(0,3)$	0.667	2.364	-0.309	-0.209	-0.372
$a(3,0)$	1.106	-0.658	0.519	0.680	0.470
$a(0,4)$	4.532	2.349	3.443	1.647	3.013
$a(4,0)$	4.007	2.239	2.821	1.622	2.837
Mean $\mu(2,0)$	-4.324	-4.186	-1.862	0.112	-0.086
$\mu(1,0)$	-0.362	0.757	-0.742	-1.773	-1.350
$\mu(2,2)$	-1.157	-0.191	0.313	-1.099	-0.351
$\mu(2,1)$	1.295	1.861	1.004	1.409	1.712
$\mu(1,2)$	-1.630	-0.906	0.048	-0.841	-0.348
$\mu(1,1)$	1.135	2.032	0.613	1.576	1.477
ARCH $R(1,0)$	6.736	4.682	5.765	2.943	4.902
$R(2,0)$	-4.992	-2.708	0.039	0.437	2.584
$R(3)$	4.850	3.916	2.487	0.985	1.717
$R(1,1)$	6.807	4.063	2.695	1.796	1.925
$R(2,1)$	4.759	4.404	5.709	2.791	5.095
$R(1,2)$	3.420	2.106	-0.214	0.326	0.151
$R(2,2)$	4.522	3.410	4.122	2.326	3.961
$R(1,3)$	4.479	3.838	1.775	0.678	1.061
$R(2,3)$	5.275	3.662	5.382	2.898	5.238
$R(1,4)$	6.447	4.237	3.196	2.252	2.334
$R(2,4)$	4.643	3.727	3.913	2.603	3.421
$R(1,5)$	3.858	2.541	1.020	1.984	1.117
$R(2,5)$	5.678	5.066	4.848	3.131	4.709

Table 4.6: Comparing Summary Statistics of Structural Simulations
The observed data of 6 month and 5 year yields are the same as those in Table 4.1.
The number of simulations from the structural models is 50,000.

6mn Yield	1-Factor[CIR]	1-Factor[RS]	2-Factor[CIR]	2-Factor[RS]	3-Factor[CIR]
Mean	0.0716	0.0643	0.0588	0.0625	0.0547
Std Dev	0.0193	0.0217	0.0214	0.0192	0.0122
Skewness	0.4877	0.6690	1.2659	0.9793	0.6478
Kurtosis	3.2425	3.9234	5.6704	5.2883	3.7180
Minimum	0.0180	0.0071	0.0112	0.0175	0.0194
05% Qntl	0.0428	0.0324	0.0314	0.0358	0.0371
25% Qntl	0.0577	0.0494	0.0442	0.0494	0.0463
50% Qntl	0.0701	0.0625	0.0552	0.0600	0.0535
75% Qntl	0.0835	0.0766	0.0695	0.0730	0.0618
95% Qntl	0.1059	0.1034	0.0983	0.0970	0.0770
Maximum	0.1576	0.1776	0.1818	0.2071	0.1182
5yr Yield	1-Factor[CIR]	1-Factor[RS]	2-Factor[CIR]	2-Factor[RS]	3-Factor[CIR]
Mean	0.0811	0.0737	0.0780	0.0788	0.0715
Std Dev	0.0146	0.0169	0.0305	0.0212	0.0130
Skewness	0.4877	0.6113	1.7656	1.8291	1.2199
Kurtosis	3.2425	3.8216	7.6753	9.7194	5.1770
Minimum	0.0405	0.0289	0.0298	0.0416	0.0451
05% Qntl	0.0593	0.0484	0.0443	0.0547	0.0552
25% Qntl	0.0706	0.0620	0.0569	0.0641	0.0623
50% Qntl	0.0800	0.0725	0.0709	0.0741	0.0689
75% Qntl	0.0901	0.0834	0.0911	0.0884	0.0779
95% Qntl	0.1071	0.1040	0.1355	0.1176	0.0968
Maximum	0.1462	0.1577	0.2737	0.2874	0.1564

Table 4.7: Average Absolute Pricing Error
The pricing error is in the unit of basis point. There are 9 maturities (1, 3, 6, 9 month; 1, 2, 3, 4, 5 year) for each of 379 dates. The absolute pricing error of 1-Factor[CIR] model is over 8 points, 1-Factor[RS] and 2-Factor[CIR] over 7 points, 2-Factor[RS] and 3-Factor[CIR] over 6 points. The summary statistics of the absolute pricing errors are calculated over the 379 dates for each of the five models.

	1-Factor[CIR]	1-Factor[RS]	2-Factor[CIR]	2-Factor[RS]	3-Factor[CIR]
Mean	47	43	30	23	25
Std Dev	28	27	18	16	21
Minimum	5	4	3	3	1
Maximum	174	175	121	114	133

127

Table 4.8: Identifying Regimes by Model Free Characteristics
The numbers in the table are basis points. The comparison is made across the two regimes in the population 2-Factor[RS] model and sample observations. The convexity is calculated using 1, 2, and 3 year adjacent yields, and the results are similar for other maturity combinations.

	6 Month Yield		5 year Yield	
Volatility or Mean $|Y_t - Y_{t-1}|$	Regime 0	Regime 1	Regime 0	Regime 1
Population	27	37	23	31
Sample	32	38	33	30
	6 Month Yield		5 year Yield	
Level or Mean Y_t	Regime 0	Regime 1	Regime 0	Regime 1
Population	626	624	763	828
Sample	721	691	718	775
Slope or Mean Spread	Regime 0		Regime 1	
Population	137		204	
Sample	-3		84	
Steepness or Mean Convexity	Regime 0		Regime 1	
Population	0.52		-13.00	
Sample	16.00		-7.76	

128
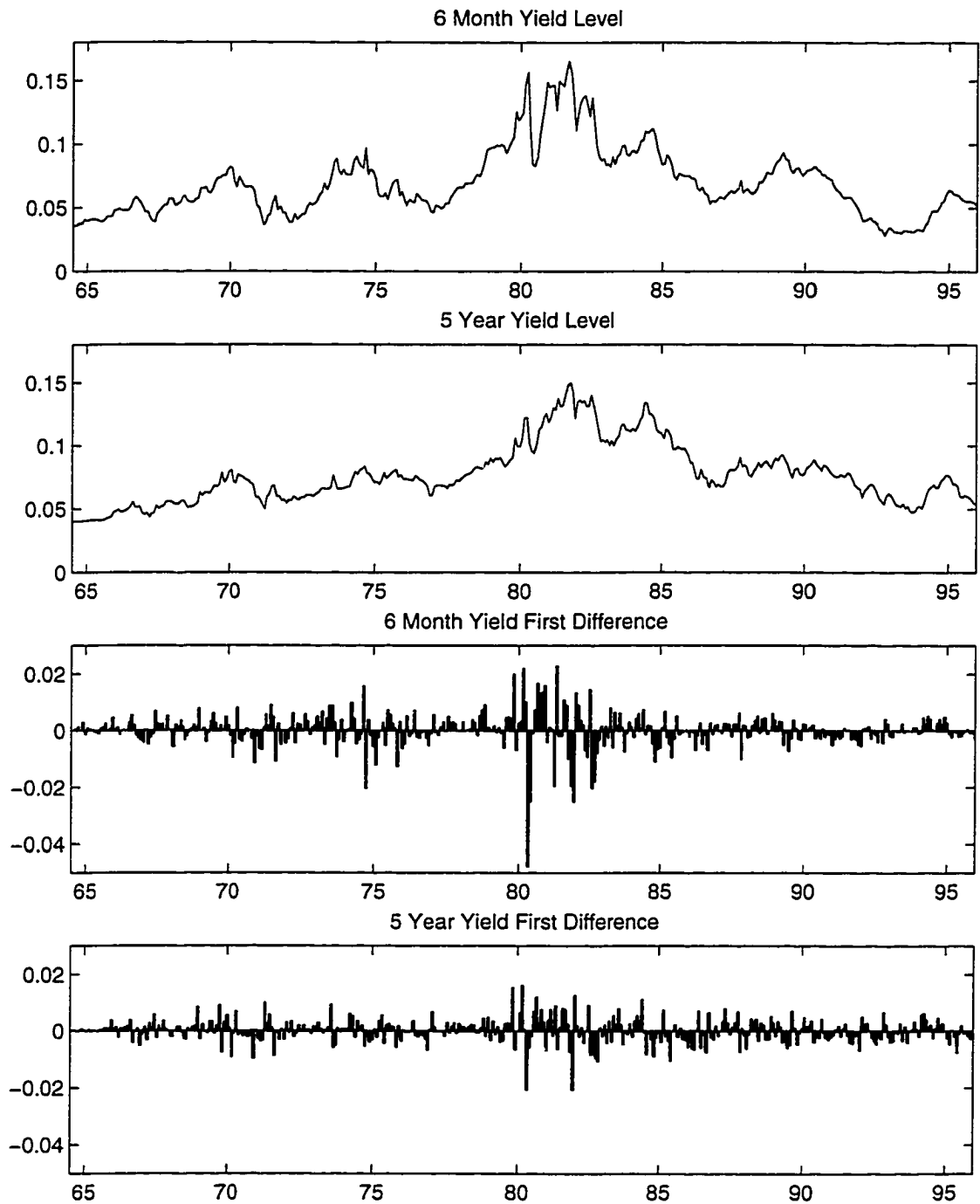
Figure 4.1: Observed Yields and First Differences

129

Joint Density

x 10⁴

Contour at 10%

x 10⁴

Marginal Density

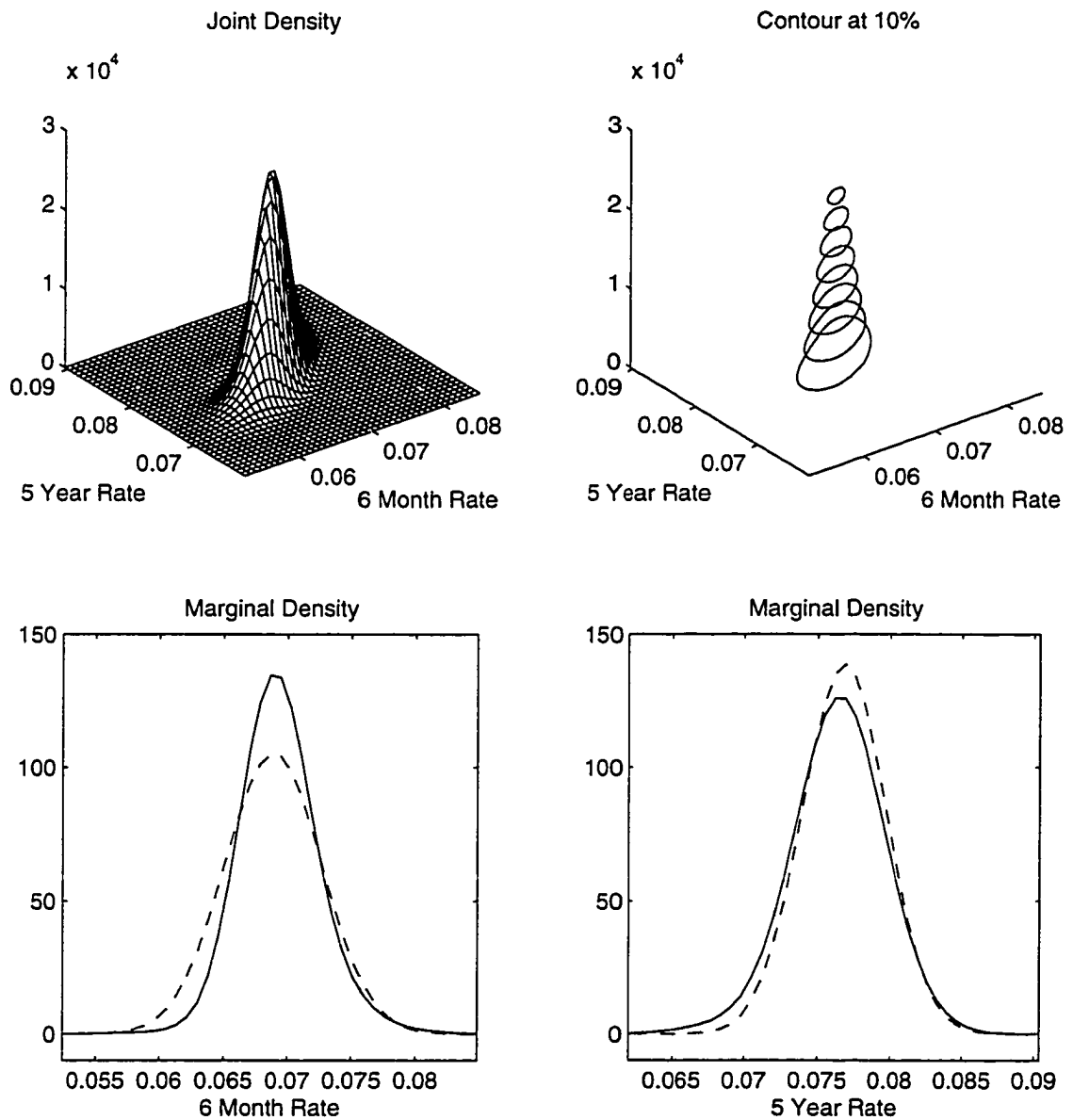**Marginal Density**

Figure 4.2: SNP Joint and Marginal Densities
The top-left panel is the joint SNP density, and the top-right is quantile contour plot
at 10% intervals. For bottom panels, the solid lines "——" are SNP marginal densities
and the dashed lines "- - -" are normal densities with same mean and variance.
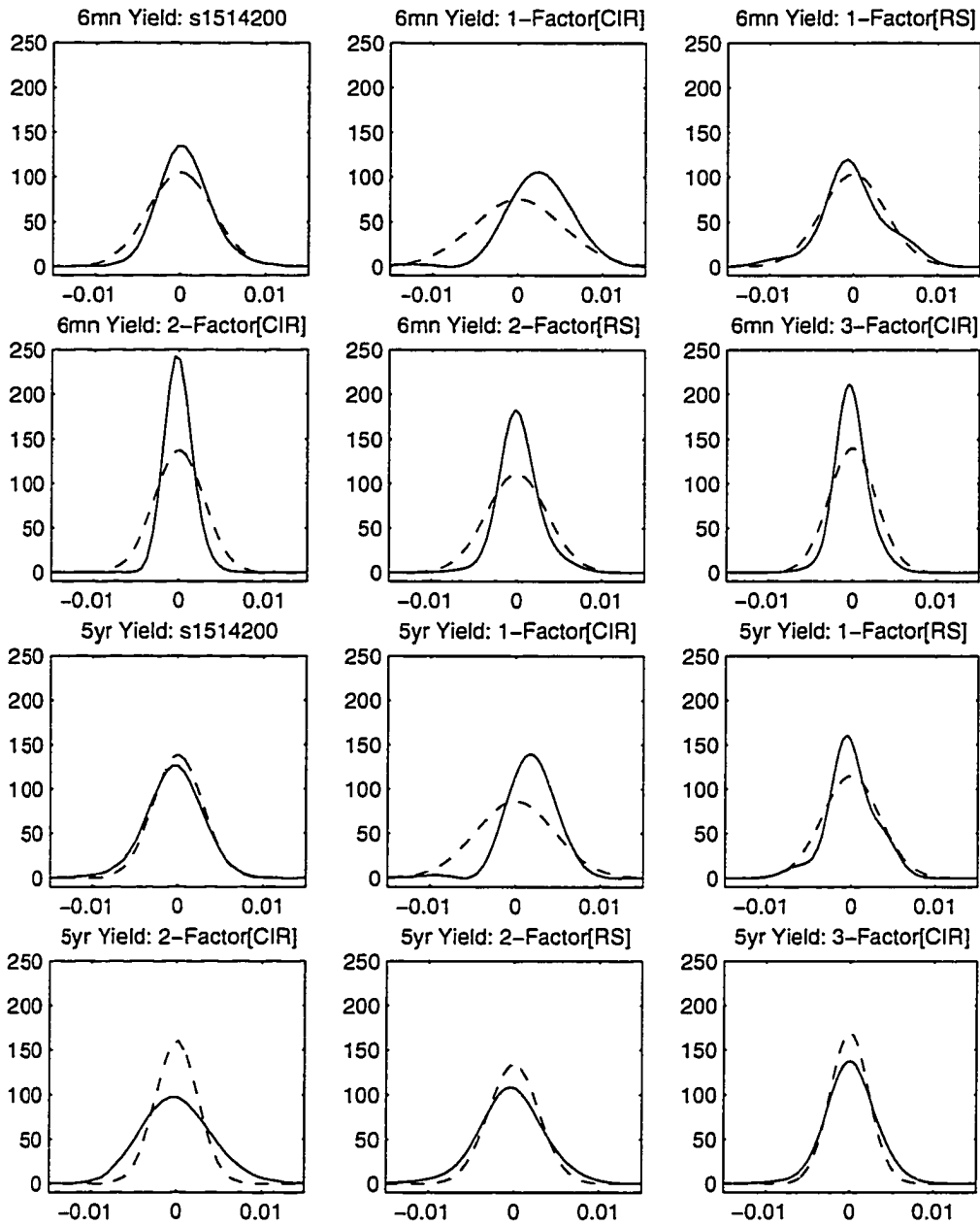
130

Figure 4.3: Projected and Reprojected Densities

The symbols are: "——" Gaussian densities with same mean and variance, "- - -" projected or reprojected densities.

131

Figure 4.4: Matching Conditional Short Volatility
Projected and Reprojected Conditional Standard Deviation of 6 Month Yield.

132

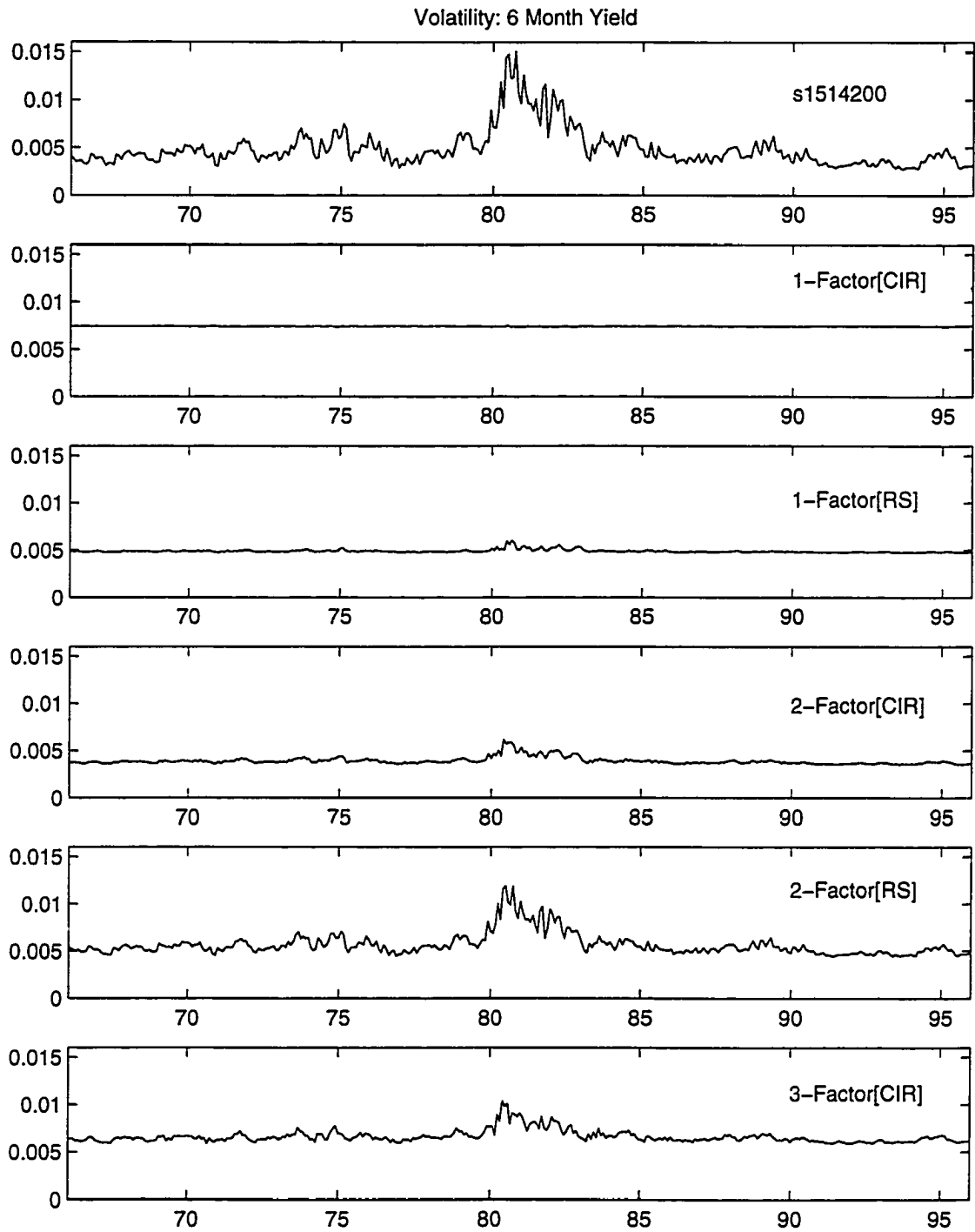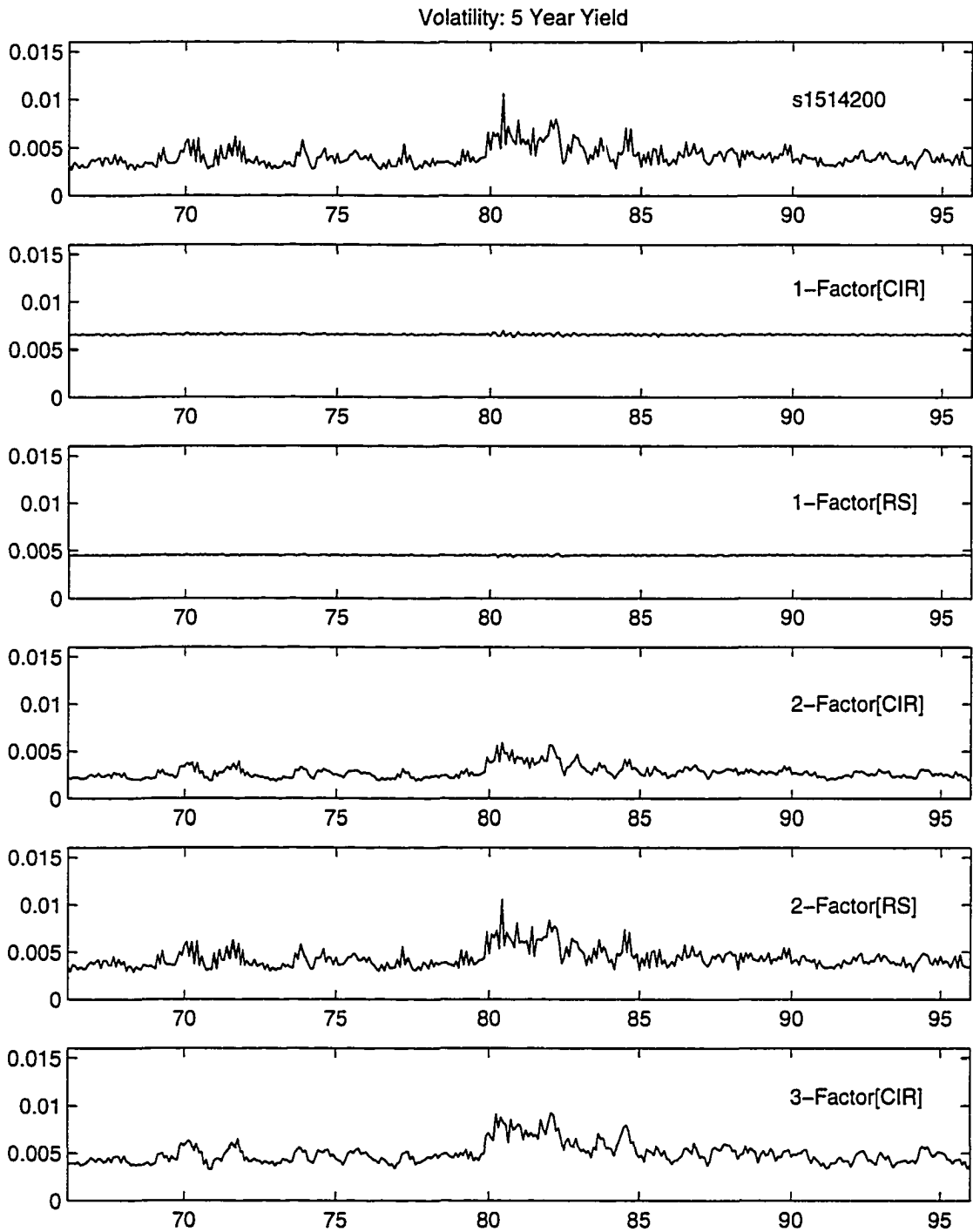**Figure 4.5**: Matching Conditional Long Volatility
Projected and Reprojected Conditional Standard Deviation of 5 Year Yield.

133
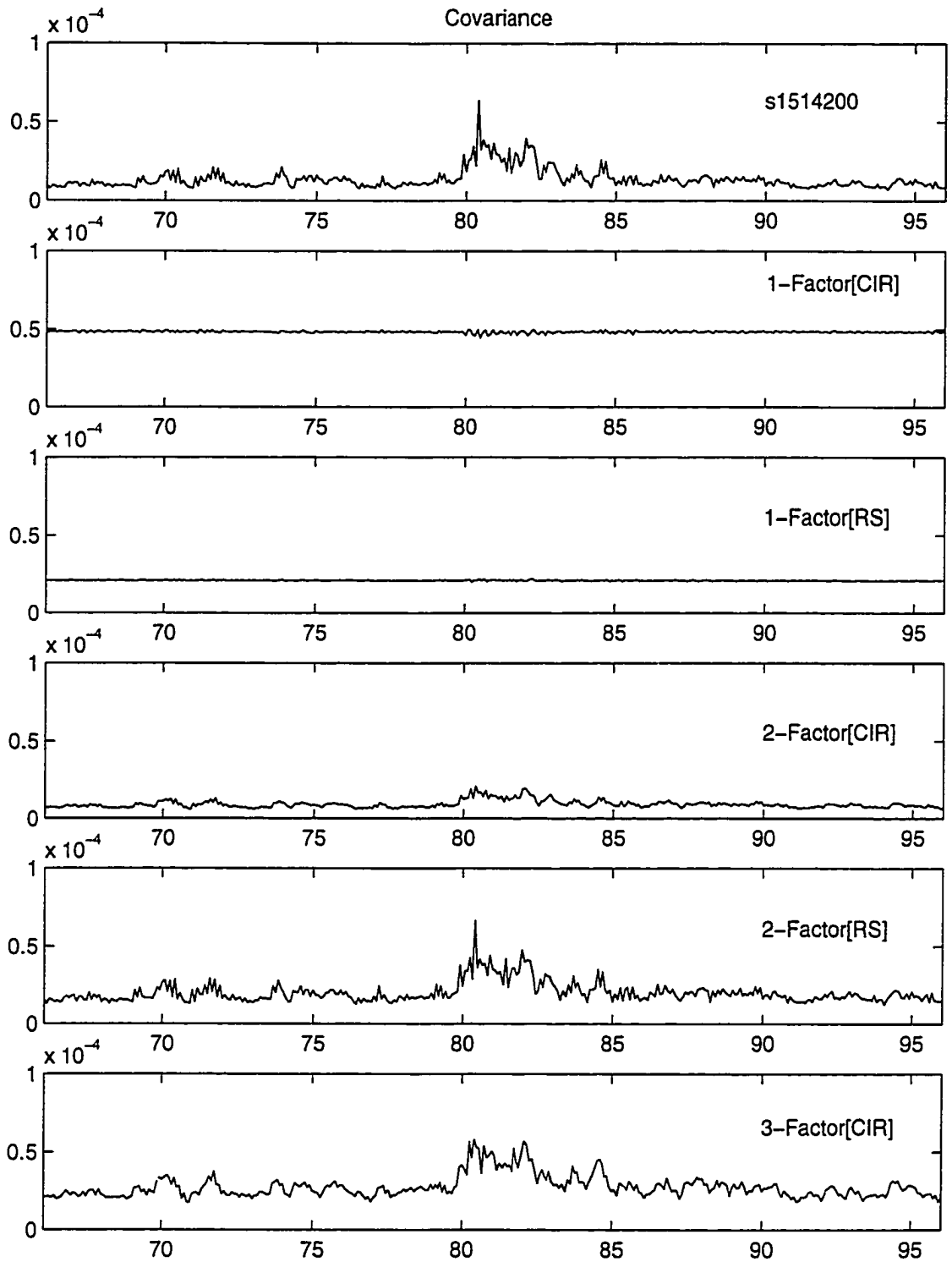
Figure 4.6: Matching Conditional Variance
Projected and Reprojected Conditional Covariance of 6 Month and 5 Year Yields.

134
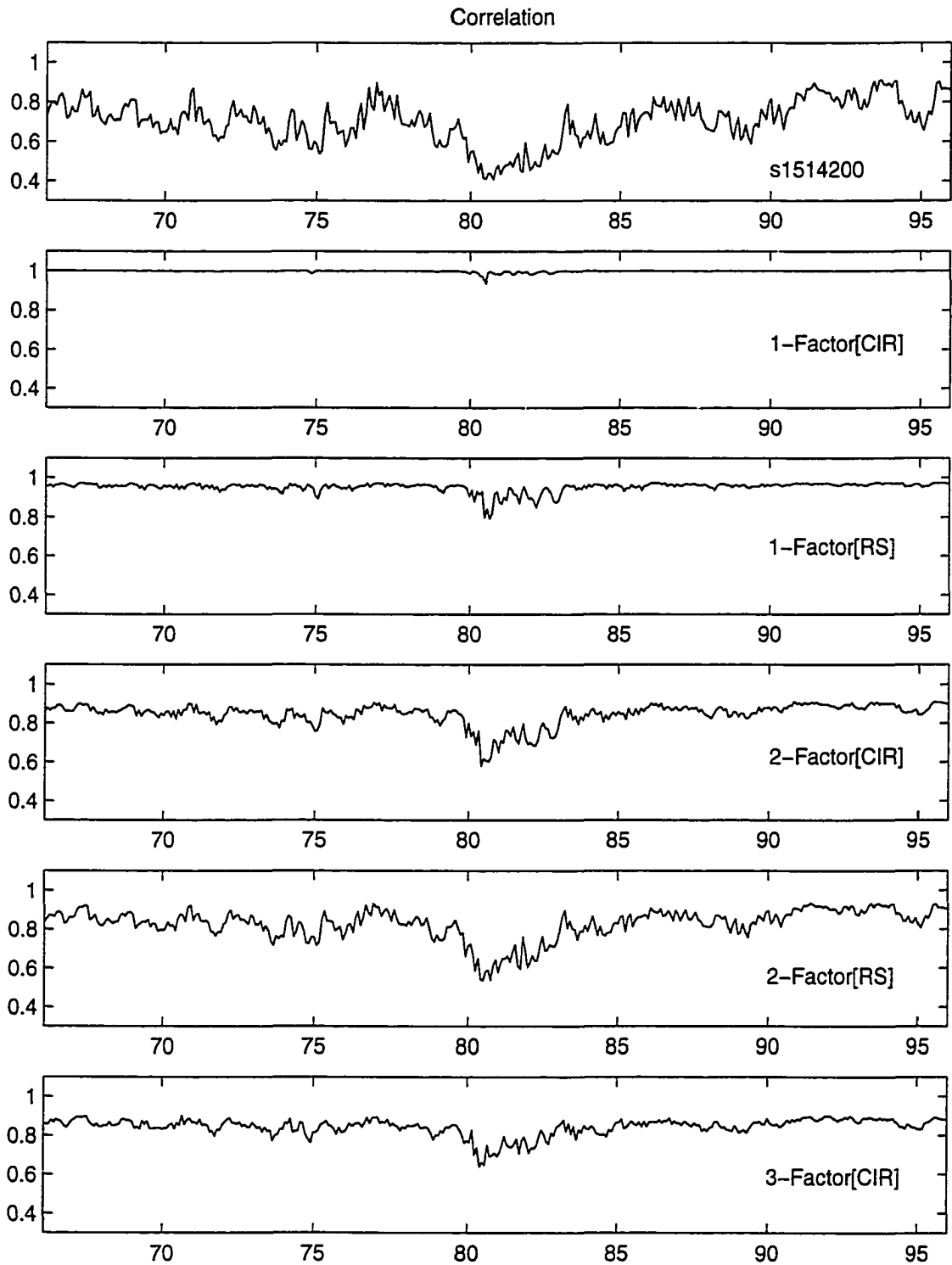
Figure 4.7: Matching Conditional Correlation
Projected and Reprojected Conditional Correlation of 6 Month and 5 Year Yields.

135
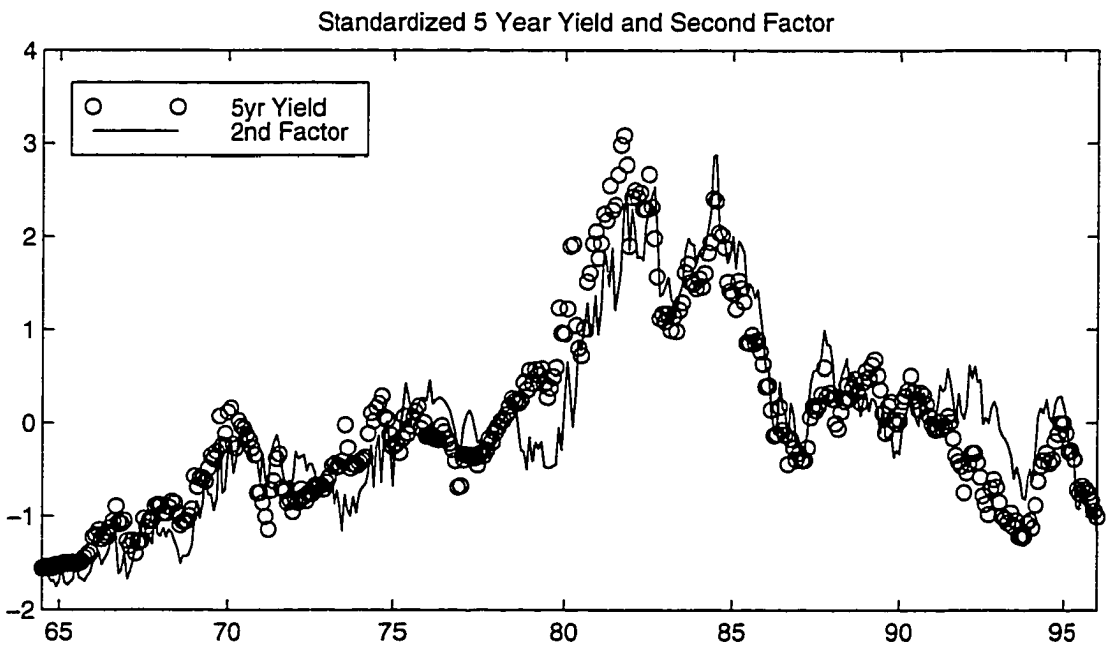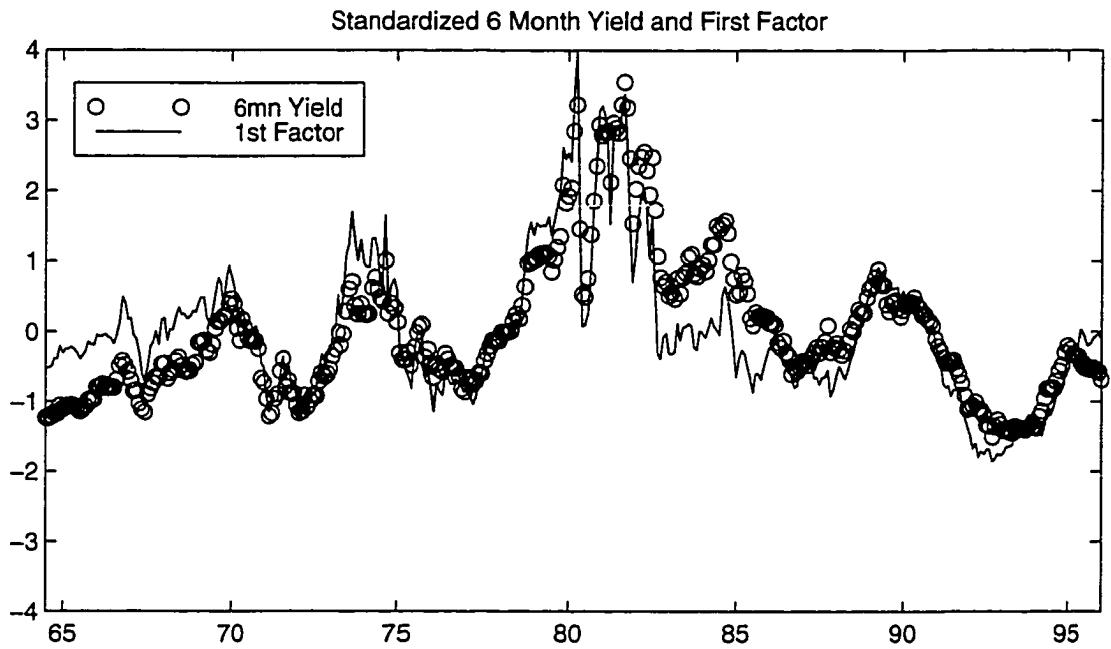
Figure **4.8**: CIR Factors and Observed Yields in 2-Factor[RS] Model.

136
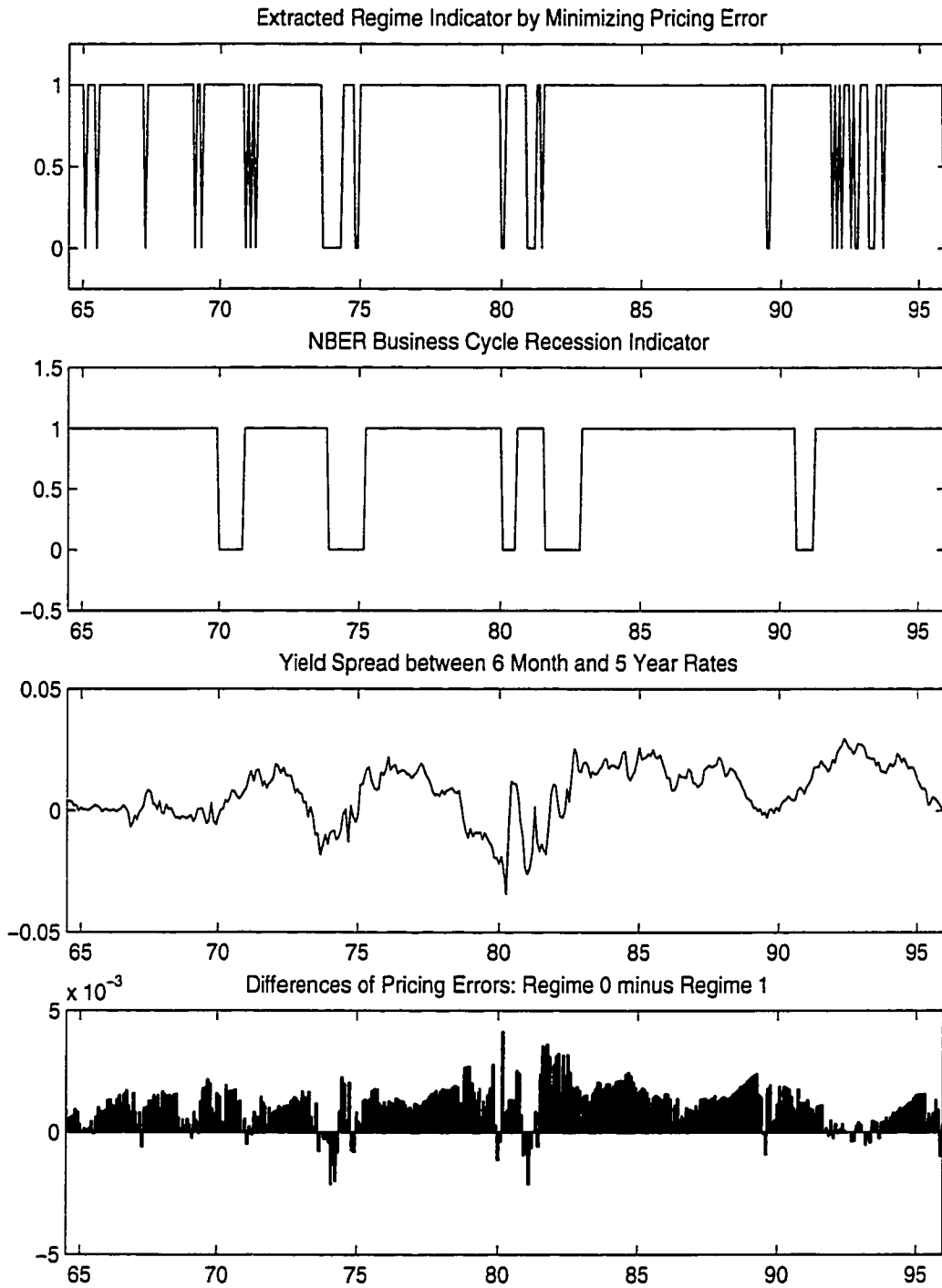
Figure 4.9: Regime Classifications and Economic Indicators in 2-Factor[RS] Model.

137
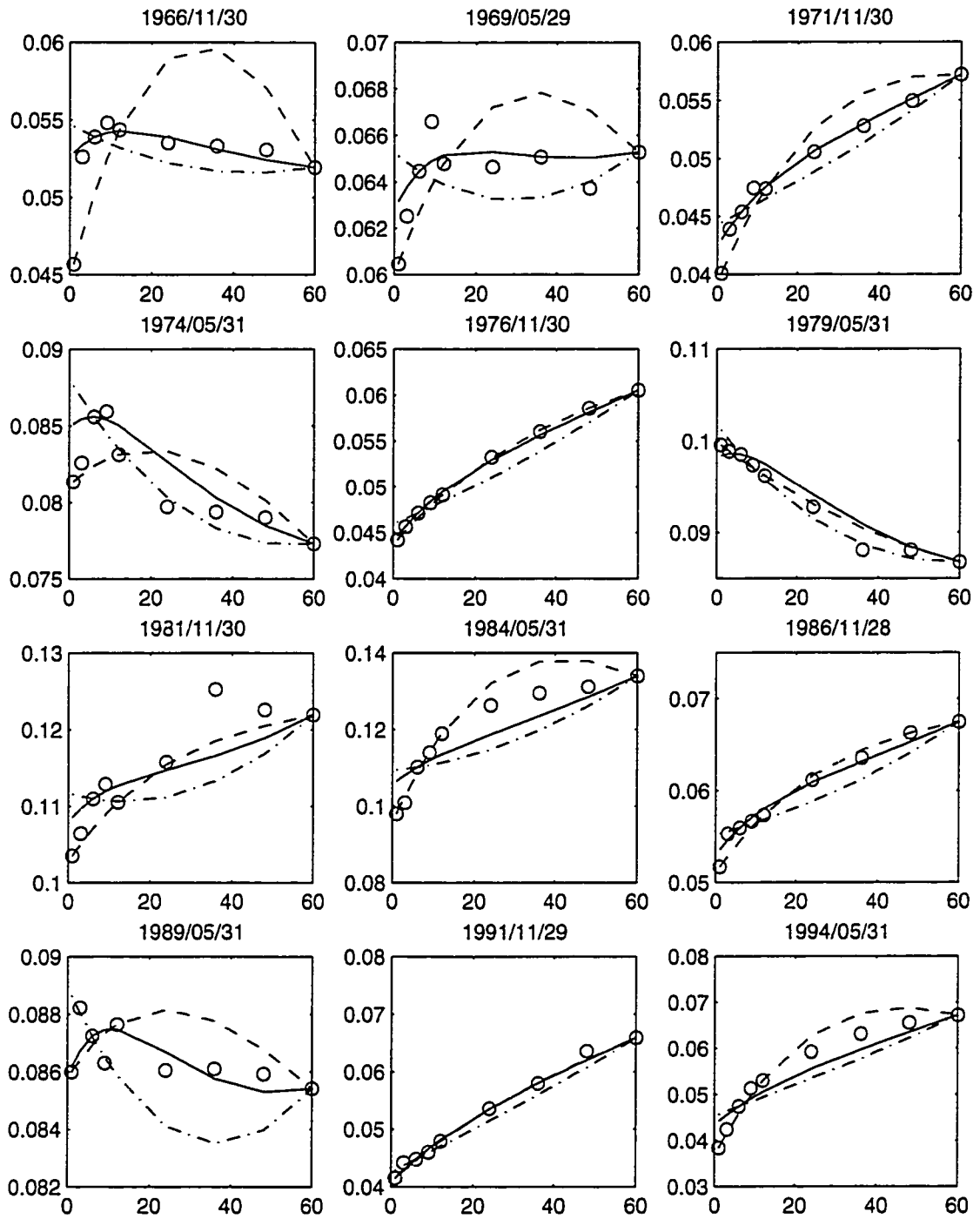
Figure 4.10: Point in Time Yield Curve
The symbols are respectively: "o o o" observed yield, "----" 2-Factor[CIR], "——" 2-Factor[RS], "- - -" 3-Factor[CIR].

138

Chapter 5

Estimating Stochastic Volatility Diffusions

[1]This essay exploits the distributional information contained in high-frequency intraday data in constructing a simple conditional moment estimator for stochastic volatility diffusions. The estimator is based on the analytical solutions of the first two conditional moments for the integrated volatility, which is effectively approximated by the quadratic variation of the process. The resulting GMM estimator is successfully implemented with high-frequency five-minute foreign exchange and equity index returns. Both simulation evidence and actual empirical results indicate that the method is very reliable and accurate. The computational speed of the procedure compares very favorably to other existing estimation methods in the literature.

5.1   Introduction

Continuous time methods and no-arbitrage arguments figure prominently in the theoretical asset pricing literature. However, some of the most influential contributions have been based upon fairly simple and restrictive assumptions concerning the process for the underlying state variable(s)——leading examples include the celebrated Black-Scholes option pricing formula, which assumes that the true process for the underlying asset follows a geometric Brownian motion; and the CIR model for the term structure of interest rates, which is derived under the assumption of a square-root process for the short rate. Meanwhile, the burgeon empirical literature on discrete time ARCH and stochastic volatility models (see Bollerslev, Engle, and Nelson (1994) and

---

[1]The main result of the this essay also appears in a paper distributed under the title "Estimating Stochastic Volatility Diffusions Using Conditional Moments of Integrated Volatility" (Bollerslev and Zhou 2000).

Ghysels, Harvey, and Renault (1996)), have called into question the empirical validity of a time invariant diffusion, or a single state variable, as a reasonable assumption for most speculative rate of return series. In response to this, several recent studies have utilized more realistic continuous time models, explicitly allowing for time varying volatility in the state variables. The Hull and White (1987) and Heston (1993) stochastic volatility option pricing formula, and the exponential-affine class of term structure models in Duffie and Kan (1996) and Dai and Singleton (2000), are all notable examples.

Aside from a few special cases, estimation of these continuous time stochastic volatility models are complicated by the lack of a closed form expression for the transition density function for the corresponding discretely sampled observations, and numerous competing estimation strategies have been proposed in the literature. An incomplete list of these different techniques includes the Markov Chain Monte Carlo (MCMC) methods advanced by Jacquier, Polson, and Rossi (1994), Eraker (1998), Kim, Shephard, and Chib (1998) and Elerian, Chib, and Shephard (1998); the simulated methods of moments approach in Duffie and Singleton (1993); the indirect inference procedure of Gourieroux et al. (1993) utilized by Engle and Lee (1997); the Efficient Methods of Moments (EMM) developed by Gallant and Tauchen (1996b) and Gallant and Long (1997) and implemented by Andersen et al. (1999c); the infinitesimal moment generator underlying the GMM procedure in Hansen and Scheinkman (1995) and Conley et al. (1997a); the non-parametric series expansions of the transition density function advocated by Aït-Sahalia (1996a) and Stanton (1997) and the related kernel estimator in Bandi and Phillips (1999); the approximation method to the likelihood function building on the Kolmogorov forward equations in Lo (1988) and Aït-Sahalia (1998); and the spectral GMM estimator utilizing the empirical characteristic function in Chacko and Viceira (1999) and Singleton (1999). While all of

140

these procedures yield consistent, and in most cases also asymptotically efficient, parameter estimates for the various model specifications, they are all computationally demanding and cumbersome to implement in practice.

This chapter proposes a new, much easier to compute, estimation procedure for stochastic volatility diffusions. The basic idea is straight forward. Instead of integrating out the latent volatility, as it is implicitly done in the estimation procedures in the extant literature, the strategy proposed here utilizes high-frequency data for explicitly measuring the realized volatility.

High-frequency, or tick-by-tick, data have recently become available for a host of different financial instruments. Following the work of Merton (1980) and Nelson (1992), such data could in principle be used to construct point-wise consistent filtering measurements for the instantaneous volatility. Unfortunately, the optimal filter weights depend in complicated ways on the particular model structure (Nelson and Foster 1994), and in practice the continuous record asymptotics underlying the theoretical arguments are corrupted by inherent discreteness, time-of-day effects, bid-ask spreads, and other market microstructure frictions.[2] Meanwhile, it is possible to construct *model-free* unbiased estimates of the integrated volatility over a fixed time interval, say one day, by simply summing the squared returns over the relevant time-period. Moreover, by the theory of quadratic variation, the sum of the squared inter-period returns afford increasingly more accurate ex-post volatility measurements as the length of the return-horizon decreases.[3] Motivated by this idea, Andersen, Bollerslev, Diebold, and Labys (1999b) and Andersen, Bollerslev, Diebold,

---

[2]In a related context, Brandt and Santa-Clara (1999) and Ledoit and Santa-Clara (1999) suggest using Black-Scholes implied volatilities for short-lived at-the-money options to estimate the instantaneous volatility. In practice, this implicitly assumes that the volatility is constant over the remaining life of the option, and that the volatility risk is not priced.

[3]Alternatively, it is possible to extract information about the forward integrated volatility from the high-low range of the discretely sampled data as in Gallant, Hsu, and Tauchen (1999). Also, Alizadeh, Brandt, and Diebold (1999) have recently proposed using the high-low range as a volatility proxy in a Gaussian quasi-maximum likelihood estimation procedure for a simple stochastic volatility model.

141

and Ebens (2000b) offer a detailed descriptive characterization of the salient distributional features of daily *realized* foreign exchange and individual stock return volatilities constructed from high-frequency five-minute returns.

Here, by matching the sample moments of the realized volatility to the population moments of the integrated volatility implied by a particular continuous-time model structure, a standard, and easy-to-compute, GMM-type estimator for the underlying model parameters is immediately applicable. For concreteness we restrict the analysis in the paper to the square-root, or affine, class of stochastic volatility models. This particular class of models arguably constitutes the leading case in the literature, but the method is general. In particular Barndorff-Nielsen and Shephard (1999) present analytical expressions for the moments of the integrated volatility for a general class of continuous time stochastic volatility models, in which the instantaneous variance is defined by the sum of multiple Ornstein-Uhlenbeck processes, each of which is driven by a homogeneous Levy process.

The rest of the paper is organized as follows. The next section demonstrates how the population moments for the integrated volatility may be derived from the moments for the point-in-time volatility. This section also briefly discusses the basic GMM setup employed in the estimation. The Monte Carlo simulations in Section 5.3 highlight, that the method works very well in empirically realistic finite sample settings, and that the efficiency of the parameter estimates compares favorably to that of a non-feasible QML procedure treating the instantaneous volatility as observable. The statistical inference concerning the true values of the individual parameters and the overall fit of the model are generally also very reliable. The only caveat is a negligible upward bias in the estimates of the variance-of-variance parameter. This is directly attributable to the measurement error in the quadratic variation as a proxy for the integrated volatility, and we show how a simple adjustment term in the

142

moment conditions is effectively able to eliminate this bias. Section 5.4 gives the empirical results from applying the new estimation procedure to a set of high-frequency five-minute foreign exchange rates and Japanese equity index returns. Section 5.5 concludes. Mathematical details regarding the derivation of the moment conditions for the integrated volatility are relegated to a technical appendix.

5.2   Estimating Stochastic Volatility Diffusion

The basic estimation strategy builds on the usual asymptotic theoiy of GMM assuming an increasing number of discretely sampled observations (Hansen 1982). However, the construction of the sample moments explicitly relies on the availability of high-frequency data, and the almost sure convergence of the quadratic variation to the integrated volatility of the process. I begin with a general discussion of the main idea, and then proceed to a concrete illustrative example.

5.2.1   Integrated Volatility and GMM Estimation

To set out the main idea, let p_t denote the time t logarithmic price for some asset. The generic continuous time stochastic volatility model may then be written as

$$
\begin{aligned}
dp_t &= \mu(p_t, V_t, t; \xi)dt + \upsilon(p_t, V_t, t; \xi)dB_t, \\
dV_t &= \kappa(p_t, V_t, t; \xi)dt + \sigma(p_t, V_t, t; \xi)dW_t,
\end{aligned}
\tag{5.1}
$$

where V_t denotes a vector of latent volatility factors, dB_t and dW_t denote compatible, possibly correlated, Brownian motions, and the drift and diffusions functions are assumed to be sufficiently regular to guarantee the existence of a unique strong solution (see, e.g., Karatzas and Shreve (1997)). Moreover, the parameters, ξ, are restricted to lie within some compact set, Ξ, containing the true parameters of the process, say ξ_0. Of course, the dependence of p_t on dW_t through both V_t and $\text{corr}(dB_t, dW_t)$ may

143

be redundant. Also, for concreteness, in the subsequent empirical analysis we will normalize the unit time interval to correspond to one day.

The exact form of the drift function, $\mu(p_t, V_t, t; \xi)$, is generally irrelevant for the consistent estimation of the parameters entering the diffusion functions. Meanwhile, the estimation of these parameters based on discretely sample observations for the p_t process are complicated by the V_t process being latent, and the lack of a closed form expression for the corresponding transition density. As noted in the introduction, this in turn has spurred the development of several alternative computationally demanding estimation procedures. However, by the theory of quadratic variation

$$\lim_{N \to \infty} \sum_{i=1}^{2^N} \left[p_{t + \frac{i}{2^N}(T-t)} - p_{t + \frac{i-1}{2^N}(T-t)} \right]^2 \xrightarrow{a.s.} \int_t^T v(p_s, V_s, s; \xi) ds \equiv \mathcal{V}_{t,T}, \qquad (5.2)$$

where $\mathcal{V}_{t,T}$ denotes for the integrated volatility from time t to T. Thus, while the point-in-time volatility, $v(p_t, V_t, t; \xi)$, is generally unobservable, by summing increasingly finer sampled squared high-frequency returns, it is possible to obtain increasingly more accurate estimates of the integrated volatility of the process. Importantly, in the limit the integrated volatility is effectively observable.[4]

Explicitly treating the integrated volatility as observable, in turn permits the implementation of a standard GMM type estimator for the underlying model parameters, by minimizing the weighted distance between the sample moments and the corresponding population moments of $\mathcal{V}_{t,T}$ implied by the particular model structure. Of course, in practice continuously sampled observations are not available, so that the integrated volatility is not actually observable. However, the same GMM estimation strategy may be formally justified under the additional assumption, that the number of observations employed in the construction of the sample moments converges to in-

---

[4]Andersen and Bollerslev (1998a) provide simulation evidence in support of this idea, and argue for the practical use of the quadratic variation as a meaningful measure of the ex-post *realized* volatility.

144

finity at a slower rate than the almost sure convergence rate of $1/2^N$ for the quadratic variation. The validity of this assumption is obviously an empirical question.

The next section details the derivation of the first two population moments for a particular class of stochastic volatility models. For concreteness, I shall focus on the square-root volatility model, or the single factor affine diffusion, analyzed by Heston (1993) among others. But, the same basic approach employed in the next section could in principle be extended to any multifactor stochastic volatility processes for which the conditional mean and conditional variance of the point-in-time volatility have tractable analytical expressions.[5] This latter class is quite general, including the affine class of stochastic differential equations popularized by Duffie and Kan (1996), and Dai and Singleton (2000), as well as the quadratic stochastic volatility class of models (see, e.g., Kloeden and Platen (1992)).

5.2.2 Conditional Moments of Integrated Volatility

The square-root volatility model, or scalar affine diffusion, is succinctly defined by,

$$
\begin{aligned}
dp_t &= \mu_t dt + \sqrt{V_t} dB_t, \\
dV_t &= \kappa(\theta - V_t) dt + \sigma \sqrt{V_t} dW_t,
\end{aligned}
\tag{5.3}
$$

where V_t is a scalar latent volatility process. While this first-order parameterization is obviously somewhat restrictive, it is nonetheless rich enough to illustrate the general idea, and it has in fact been widely used in the literature. In this parameterization, θ determines the long-run (unconditional) mean, κ is the mean reversion parameter, and σ denotes the local variance (volatility-of-volatility) parameter. For the process to be well defined, the parameters must satisfy: $\theta > 0$ (non-negativity), $\kappa > 0$

---

[5]Although the procedure implemented here hinges on the matching of selected population and sample moments for the integrated volatility, in situations where analytical expressions for the population moments $V_{t,T}$ are not directly available, these could easily be evaluated by simulations, and the underlying parameters estimated by simulated methods of moments (Duffie and Singleton 1993).

145

(stationary in mean), and $\sigma^2 \leq 2\kappa\theta$ (stationary in volatility). Note that the drift of the asset returns, μ_t, can be any linear or nonlinear function of the state variables, p_t, V_t, or even other unobservable factor(s), without impeding the estimation of the stochastic volatility component.

In deriving the conditional moments for the integrated volatility, it is useful to distinguish between two different information sets—the continuous sigma-algebra $\mathcal{F}_t = \sigma\{V_s; s \leq t\}$, generated by the point-in-time volatility process, and the discrete sigma-algebra $\mathcal{G}_t = \sigma\{\mathcal{V}_{t-s-1,t-s}; s = 0,1,2,\cdots,\infty\}$, generated by the integrated volatility series. Obviously, the coarser filtration is nested in the finer filtration (i.e., $\mathcal{G}_t \subset \mathcal{F}_t$), and by the Law of Iterated Expectations, $E[E(\cdot|\mathcal{F}_t)|\mathcal{G}_t] = E(\cdot|\mathcal{G}_t)$.

Conditional Mean

In deriving the conditional mean for the integrated volatility, it is useful to start with the conditional mean of the point-in-time volatility. In particular, it follows from the result in Cox et al. (1985a) that,

$$E(V_T|\mathcal{F}_t) = \alpha_{T-t}V_t + \beta_{T-t}, \tag{5.4}$$

where α_{T-t} and β_{T-t} are functions of the structural parameters and the horizon of the forecast, $T - t$ (see Appendix B for details). The second step is to express the conditional mean of the integrated volatility as a (linear) function of the point-in-time volatility by interchanging the integration operators

$$E\left(\mathcal{V}_{t,T}|\mathcal{F}_t\right) = E\left(\left.\int_t^T V_s ds \right| \mathcal{F}_t\right) = a_{T-t}V_t + b_{T-t}, \tag{5.5}$$

where a_{T-t} and b_{T-t} denote other explicit functions of the drift parameters and the sampling interval.

Now, by iteratively substituting the above two results, the conditional mean of the integrated volatility for the one-day horizon, given the finer information set $\mathcal{F}_t$,

146

is readily expressed as (see Appendix B),

$$E\left(\mathcal{V}_{t+1,t+2}|\mathcal{F}_t\right) = \alpha E\left(\mathcal{V}_{t,t+1}|\mathcal{F}_t\right) + \beta,$$

where for notational simplicity we omit the subscript for the daily horizon; i.e. $\alpha \equiv \alpha_1$ and $\beta \equiv \beta_1$. Using the Law of Iterated Expectation, the above relationship can be conditioned on the coarser information set $\mathcal{G}_t$, yielding

$$E\left[E\left(\mathcal{V}_{t+1,t+2}|\mathcal{F}_t\right)|\mathcal{G}_t\right] = E\left(\mathcal{V}_{t+1,t+2}|\mathcal{G}_t\right) = \alpha E\left(\mathcal{V}_{t,t+1}|\mathcal{G}_t\right) + \beta. \qquad (5.6)$$

The first order moments for the multi-period integrated volatility may be derived by similar reasoning.

Conditional Second Moment

Analogous to the derivation of the conditional first moment above, it is convenient to start from the expression for the conditional variance for the point-in-time volatility. Again, following Cox et al. (1985a), we have

$$E(V_T^2|\mathcal{F}_t) = Var(V_T|\mathcal{F}_t) + [E(V_T|\mathcal{F}_t)]^2 = C_{T-t}V_t + D_{T-t} + [\alpha_{T-t}V_t + \beta_{T-t}]^2, \qquad (5.7)$$

where C_{T-t} and D_{T-t} are functionally dependent on the structural parameters and the sampling interval. Now by expressing the conditional variance of the integrated volatility as a linear function of the point-in-time volatility and by exploiting Itô's Lemma, it is possible to show that

$$Var(\mathcal{V}_{t,T}|\mathcal{F}_t) = A_{T-t}V_t + B_{T-t}, \qquad (5.8)$$

where A_{T-t} and B_{T-t} represent other functionals of the parameters (see Appendix B for detailed derivations).

Now combining the conditional variance formula in (5.8) and the conditional mean formula in (5.5), we can derive the second moment of the integrated volatility

147

conditional on the finer information set $\mathcal{F}_t$. In particular, for the one-day horizon this takes the form

$$E(\mathcal{V}_{t,t+1}^2|\mathcal{F}_t) = Var(\mathcal{V}_{t,t+1}|\mathcal{F}_t) + [E(\mathcal{V}_{t,t+1}|\mathcal{F}_t)]^2 = a^2 V_t^2 + (2ab + A)V_t + (b^2 + B)$$

where we have omitted the "daily" subscript "1" on a, b, A and B for notational convenience. Finally by repeatedly applying the Law of Iterated Expectation on different information sets and substituting expressions between integrated volatility and point-in-time volatility, it follows that

$$E[E(\mathcal{V}_{t+1,t+2}^2|\mathcal{F}_t)|\mathcal{G}_t] = E(\mathcal{V}_{t+1,t+2}^2|\mathcal{G}_t) = H E(\mathcal{V}_{t,t+1}^2|\mathcal{G}_t) + I E(\mathcal{V}_{t,t+1}|\mathcal{G}_t) + J, \quad (5.9)$$

where the functions H, I, and J are again defined in Appendix B. Corresponding moment conditions for the squared multi-period integrated volatility follow by analogous arguments.

Conditional Moment Restrictions

The analytical solutions for the conditional first and second moments in equations (5.6) and (5.9), immediately set the stage for the construction of a standard GMM type estimator. Of course, the efficiency of the resulting estimator defined from these equations will depend upon the particular choice of instruments (see Hansen (1985), Hansen, Heaton, and Ogaki (1988), and Gallant and Tauchen (1996b) for additional discussion and formal results along these lines). In the implementation pursued here, we simply augment the two basic moments with their own lag-one and lag-one squared counterparts, resulting in the following six moments,

$$f_t(\xi) \equiv \begin{bmatrix} E[\mathcal{V}_{t+1,t+2}|\mathcal{G}_t] - \mathcal{V}_{t+1,t+2} \\ E[\mathcal{V}_{t+1,t+2}^2|\mathcal{G}_t] - \mathcal{V}_{t+1,t+2}^2 \\ E[\mathcal{V}_{t+1,t+2}\mathcal{V}_{t-1,t}|\mathcal{G}_t] - \mathcal{V}_{t+1,t+2}\mathcal{V}_{t-1,t} \\ E[\mathcal{V}_{t+1,t+2}^2\mathcal{V}_{t-1,t}|\mathcal{G}_t] - \mathcal{V}_{t+1,t+2}^2\mathcal{V}_{t-1,t} \\ E[\mathcal{V}_{t+1,t+2}\mathcal{V}_{t-1,t}^2|\mathcal{G}_t] - \mathcal{V}_{t+1,t+2}\mathcal{V}_{t-1,t}^2 \\ E[\mathcal{V}_{t+1,t+2}^2\mathcal{V}_{t-1,t}^2|\mathcal{G}_t] - \mathcal{V}_{t+1,t+2}^2\mathcal{V}_{t-1,t}^2 \end{bmatrix} . \quad (5.10)$$

148

By construction $E[f_t(\xi_0)|\mathcal{G}_t] = 0$, and the corresponding GMM, or minimum chi-square, estimator is defined by $\hat{\xi}_T = \arg\min g_T(\xi)'W g_T(\xi)$, where $g_T(\xi)$ refers to the sample mean of the moment conditions, $g_T(\xi) \equiv 1/T \sum_{t=1}^{T-2} f_t(\xi)$, and W denotes the asymptotic covariance matrix of $g_T(\xi_0)$ (Hansen 1982). Under standard regularity conditions, the minimized value of the objective function multiplied by the sample size is distributed asymptotically as a chi-square distribution with three degrees of freedom, which allows for an omnibus test of the overidentifying restrictions. Moreover inference concerning the individual parameters is readily available from the standard formula for the asymptotic covariance matrix, $(\partial f_t(\xi)/\partial\xi' W \partial f_t(\xi)/\partial\xi)/T$.

The one-period lag in the moment conditions in equation (5.10) implies an MA(1) error structure. However, in order to avoid any finite sample problems with the sample analogue of W not being positive definite, in the simulations and the actual empirical estimates reported below, I used a heteroskedasticity and autocorrelation consistent robust covariance matrix estimator with a Bartlett-kernel and a lag length of five (Newey and West 1987).[6] The next section details the results from a Monte Carlo study designed to investigate the finite sample performance of this particular GMM estimator.

5.3 Monte Carlo Study

One important aspect in evaluating econometric methods for estimating continuous time process concerns their finite sample performance. With strong temporal dependence and/or conditional heteroskedasticity in the data generating process, asymptotically sound estimators have been shown to exhibit very slow convergence rates (see, e.g., Pritsker (1998)). This section qualifies the small sample efficiency of our GMM estimator, along with the resulting omnibus specification test, and Wald based

---

[6]I also experimented with other lag lengths. All of the results were very similar to the ones reported here, and are available upon request.

149

parameter inference.

5.3.1 Experimental Design

Here I present the results for three benchmark specifications. Scenario A ($\kappa = 0.03$, $\theta = 0.25$, $\sigma = 0.10$) features a highly persistent volatility process (nearly unit-root); Scenario B ($\kappa = 0.10$, $\theta = 0.25$, $\sigma = 0.10$) has a more stationary variance process; Scenario C ($\kappa = 0.10$, $\theta = 0.25$, $\sigma = 0.20$) has a higher variance-of-variance and is close to the non-stationary region ($\sigma^2 > 2\kappa\theta$).

In simulating the data, we utilize a first order Euler scheme with 82 artificial "five-minute" intervals per day, further partitioning each five-minute interval into 10 segments.[7] The quadratic variation formula (5.2) is employed to approximate the integrated volatility series. To check the standard "long-span" asymptotics, the econometric sample sizes are chosen as T $= 1,000$ and T $= 4,000$. Since the true "continuous time" record is known inside the simulations, we compare the GMM estimator using the "five-minute" quadratic variation with the corresponding non-feasible estimator based on the true integrated volatility. Lastly, we also compare the results for the GMM estimator with a QML estimator based on the "daily" point-in-time volatility assuming the process to be Gaussian.[8] Of cause, this latter estimator is not feasible in practice either. The results are summarized in Tables 5.1-5.3 and Figures 5.1 and 5.2.

---

[7]Most US financial markets are open between six-and-a-half to seven hours, corresponding to 78-84 five-minute intervals.

[8]This estimator is closely related to the ideas in Fisher and Gilles (1996), who propose a Quasi-Maximum Likelihood estimator for Affine diffusion process, using closed form solutions for the conditional mean and variance.

150

5.3.2 Parameter Estimate and Efficiency

First, the finite sample results not only corroborate the moment conditions derived earlier for the integrated volatility, but also indicate that the GMM estimator fairs well (if not better) than the other two non-feasible alternatives—using "unobserved" point-in-time volatility or the "continuous time" record. The root-mean-squared-errors (RMSEs) of the drift parameters, κ and θ, decrease roughly at the rate of $\sqrt{4}$ as the sample size increases from 1,000 to 4,000 "days". Meanwhile, the mean-reversion parameter κ is upward biased, and the long-run mean parameter θ exhibits a small downward bias.

Second, the accuracy of the local variance parameter estimates is affected by both the long-span asymptotics and the fill-in asymptotics. Although the RMSE of σ does decrease when the sample size goes from 1,000 to 4,000, the rate is not always $\sqrt{4}$. Also, while the drift parameter estimates are almost unaffected by the fill-in asymptotics, the RMSE of σ clearly diminishes when the sampling frequency increases from "five-minute" to the "continuous time" limit. This confirms the theoretical arguments that the diffusion parameter can be estimated exactly with continuous sampling (Merton 1980, Lo 1988, Nelson 1992).

Third, when the process is close to a unit-root (Scenario A), the variance parameter seems to converge at a faster rate than $\sqrt{T}$ (Table 5.1 Panel A and Figure 5.1). Also when the variance-of-variance parameter is large (Scenario C in Figure 5.1), the finite sample biases of the drift parameter estimates are larger than for the more stationary case (Scenario B in Figure 5.1). Basically the GMM estimator is not able to distinguish between a very persistent yet stationary process and a non-stationary, near unit-root process in "small" samples.

Lastly, the GMM estimates of the local variance parameter, σ, are systematically upward biased in all three scenarios. Interestingly, this bias completely vanishes when

151

the true integrated volatility is used in place of the "five-minute" quadratic variation. While the measurement error from using the quadratic variation to approximate the integrated volatility process is averaged out in the first moment condition, the second moment condition depends non-linearly on the measurement error. We will investigate this issue further in Section 5.3.4.

In terms of relative efficiency, the GMM estimator using the "five-minute" realized volatility actually performs better than the non-feasible QML estimator using the unobservable point-in-time volatility for the drift parameters in all three scenarios, and better for the variance parameter in all but the stationary scenario. The middle rows in Table 5.1 suggest that the RMSEs of the GMM estimator using the true integrated volatility process are much smaller than those of the QML estimator. However, going to the "continuous time" limit does not necessarily improve the efficiencies of the GMM drift parameters, but it does increase the convergence rate of the diffusion parameter.

5.3.3 Statistical Inference

In practice, inference concerning the individual model parameters and the overall specification of the model will have to be based on the standard GMM type test statistics discussed in Section 5.2.2. In this regard, the t-statistics for the drift parameters in Figure 5.1 clearly indicate that the GMM estimator based on the "five-minute" quadratic variation is close to normal for both 1,000 and 4,000 "daily" sample sizes analyzed here. Meanwhile for the diffusion parameter, the use of "five-minute" realized volatility in the GMM estimation gives rise to a systematic upward bias in the t-statistics. This is consistent with the earlier explanation of the non-dissipating measurement error in the second moment condition.[9]

[9]The corresponding t-tests for the non-feasible QML estimator based on the point-in-time volatility are generally much more distorted, while the t-tests for the GMM estimates using the true integrated volatility are all extremely close to normal. These results are available upon request.

152

Turning to Figure 5.2 and the GMM tests of overidentifying restrictions, it follows that except for the near unit-root case (Scenario A in Figure 5.2), the test performs very well. Moreover, the slight over-rejection and under-rejection biases largely vanishes as the sample size increases from 1,000 to 4,000.[10]

5.3.4 Measurement Error Adjustment

By construction the quadratic variation based on the simulated "five-minute" returns provides an unbiased estimate for the true integrated volatility. At the same time, the squared quadratic variation for any fixed sampling interval yields a biased estimate of the true squared integrated volatility. Consequently, while the linear expectations operator washes out the measurement errors in the first conditional moment and the corresponding two augmented moments in equation (5.10), the three moment conditions involving the squared quadratic variation will entail a non-zero measurement error.[11] Although the exact form of the measurement error is not known, it follows by the almost sure convergence of the quadratic variation, that the expectation of the squared error term is bounded by the local maximum of the continuous local martingale process (Karatzas and Shreve 1997, Protter 1992). In order to conveniently approximate this term, we simply included an additive nuisance parameter, γ, in each of the three second order moment conditions, replacing the squared "five-minute" quadratic variation, $V^2_{t+1,t+2}$, by $V^2_{t+1,t+2} + \gamma$.

Not surprisingly, from the results reported in Table 5.4 and Figure 5.3, the parameter estimates for the two drift parameters and the corresponding t-statistics are

---

[10] Overrejection biases of GMM omnibus tests are widely reported in the literature (Andersen and Sørenson 1996, Hansen et al. 1996), whereas underrejection biases often occur when lag instruments are used to form the moment conditions (Tauchen 1986).

[11] Andersen and Bollerslev (1998a) provide some limited simulation evidence on the size of this measurement error as a function of the sampling frequency. Andersen, Bollerslev, Diebold, and Labys (2000c) and Bai, Russell, and Tiao (1999) discuss practical considerations related to the inherent market microstructure frictions and the choice of the sampling frequency with actual high-frequency data.

153

largely unaffected by the estimation of this additional nuisance parameter. More important, the pervasive finite sample biases for the local variance parameter estimates have completely disappeared.[12] Moreover, the rejection frequencies for the GMM specification test for the overidentifying restrictions appear marginally closer to their nominal sizes. Thus, all in all, the inclusion of a simple additive correction term for the squared quadratic variation has effectively eliminated the only notable statistical bias in the procedure. Of course, it is possible that more advanced measurement error adjustment procedures could result in further improvements, especially for more complicated models. However, for the square-root volatility diffusion in equation (5.3), the GMM estimation procedure proposed here works very well in realistic fixed-interval finite sample settings.

5.4 Empirical Illustration

This section provides an empirical illustration of the new estimation procedure using actual high-frequency date. For ex positional purposes, we will focus on the estimation results for the simple scalar affine diffusion analyzed in the previous two sections. To illustrate the applicability of the procedure across different markets and institutional arrangements, we present the results for two separate data sets: spot foreign exchange rates, and Japanese equity index returns. In line with the simulations in the preceding section, we partition the trading day for each of the markets into five-minute intervals, incorporating an additive nuisance parameter to correct the inherent measurement error in the resulting five-minute quadratic variation measures.

---

[12]Meanwhile, the RMSEs for σ have increased somewhat.

154

5.4.1 Data Description

The data for the foreign exchange market were obtained from Olsen&Associates in Zürich, Switzerland, and consists of continuously recorded five-minute returns for the Deutsche Mark/U.S. Dollar (DM/$) and Japanese Yen/U.S. Dollar (Yen/$) spot exchange rates. The sample for the exchange rates spans the period from December 1, 1986 through December 1, 1996. After removing missing data, weekends, fixed holidays, and other calendar effects, as detailed in Andersen et al. (1999b), we are left with a total of 2,445 trading days, each of which consists of 288 five-minute returns over the 24-hour trading cycle.

The intraday data for the Nikkei 225 composite stock market index were provided by Nihon Keizei Shimbun Inc. The five-minute returns for the Nikkei 225 covers the period from January 2, 1994 through December 31, 1997. Excluding days on which the Japanese equity market was closed results in a total of 984 trading days. The Tokyo Stock Exchange opens at 9:00 a.m., closes for lunch from 11:00 to 12:30, and closes for the day at 15:00 p.m.. Omitting the first five-minute interval of the day associated with the special Itayose batched trading process at the opening, leaves us with 53 five-minute returns per day. In contrast to the very actively traded foreign exchange rates, the five-minute returns for the Nikkei 225 cash index is plagued by important non-synchronous trading effects (see, e.g., Lo and MacKinlay (1990) and Chan, Chan, and Karolyi (1991), for a discussion of non-synchronous trading effects in equity index returns). While the resulting autocorrelation in the high-frequency returns does not formally affect the continuous record asymptotics underlying the GMM estimator, any mean dependencies in the discretely sampled returns will systematically bias the quadratic variation as an estimate for the true latent integrated volatility. In order to minimize this bias we pre-whitened the returns by a first order autoregressive model, treating the residuals from this model as the actual five-minute

155

return series.[13] For a more detailed discussion of the pertinent institutional arrangements and the pre-whitened five-minute Nikkei 225 returns, we refer to Andersen, Bollerslev, and Cai (2000a).

Next we transform the three five-minute return series into daily time series of integrated volatilities, as approximated by the quadratic variations in equation (5.2). Table 5.5 provides the standard set of summary statistics for each series. The means of the integrated volatility for the two exchange rates imply an annualized standard deviation of approximately 11.5 percent, whereas the annualized volatility for the Japanese stock market equals 14.6 percent.[14] The standard deviations of the integrated volatilities are close to the mean for all three markets. The higher order moments indicate extremely heavy tails and, most notably in the case of the Yen/$ spot exchange rate, important skewness to the right. These distributional features are confirmed by visual inspection of the time series plots in Figure 5.5. Each of the panels also reveals a high degree of serial correlation in the integrated volatility series. The next subsection presents the estimation results from the stochastic volatility model in equation (5.3) explicitly designed to capture this volatility clustering effect.

5.4.2 Estimation Results

Before proceeding to the actual estimation results, we caution that the scalar volatility diffusion in equation (5.3) is too simplistic to fully account for the complex dynamic dependencies in the high-frequency return series. In particular, there are sound theoretical reasons to expect there to be at least two factors affecting the exchange rates (Bansal 1997). Also, a number of recent studies have argued for

---

[13]The estimated AR(1) coefficient for the raw Nikkei 225 five-minute returns equals 0.1429. Details concerning the model estimates based on the un-adjusted five-minute Nikkei 225 returns are available on request. The resulting daily integrated volatility series is slightly smoother, and the parameter estimates are marginally lower in level, persistence, and variance.

[14]The annualized standard deviation is obtained by multiplying the mean of the daily integrated volatility by 250 and taking the square-root.

the empirical relevance of including multiple factors and/or jump components when modeling equity index returns (Chacko and Viceira 1999, Andersen, Benzoni, and Lund 1999a, Chernov et al. 1999). Moreover, the model in equation (5.3) does not incorporate the strong periodic dependencies in the volatility within the day documented in several recent studies (Andersen and Bollerslev 1997).[15] In spite of these deficiencies, we feel that the square-root stochastic volatility model is rich enough to offer a first meaningful empirical illustration of the applicability of the new estimation procedure.

The parameter estimates for the three series are reported in Table 5.4.[16] With the exception of the slightly higher values for σ, the estimates are almost identical to the ones reported here. As expected the estimates for the long-run means, or θ, are all fairly close to the sample means for the three integrated volatility series reported in Table 5.5. Also, not surprisingly, the estimates of the variance-of-variance parameter, or σ, have the largest standard errors among all of the parameters. Meanwhile, the estimated mean reversion parameters, κ, are on the high side relative to the values reported in the extant literature using more complicated discrete time ARCH and stochastic volatility type formulations. Even though the GMM omnibus test only rejects the model for the DM/$ exchange rate, the one-factor model is obviously an oversimplification of the true dynamic dependencies for all three markets. However, from an overall perspective, the estimation results in Table 5.6 are generally in line with the simulation evidence reported in the previous section, and clearly suggest that the new estimation procedure could effectively be employed in the empirical estimation of more complicated continuous time diffusions.

---

[15]Interestingly, the forecasting result in Andersen and Bollerslev (1998b) and Andersen et al. (2000a) suggest that the influences of the intraday periodicities are effectively eliminated in the daily integrated volatility measures utilized in the GMM estimation.

[16]Details regarding the parameter estimates without the additive measurement error term are available upon request.

157

5.5   Concluding Remarks

Exploiting closed form analytic expressions for the conditional moments of integrated volatility coupled with highly accurate empirical quadratic variation measures constructed from high-frequency data, we proposed a new class of GMM-type estimators for stochastic volatility diffusions. In contrast to other computationally demanding estimation procedures routinely used in the literature, such as the simulation based EMM and MCMC methods, the GMM estimator developed here is very easy to implement, requiring only the solution to a standard non-linear optimization problem. The Monte Carlo evidence shows that the procedure results in highly accurate parameter estimates and reliable statistical inferences in realistic finite samples. In implementing the new estimator with actual five-minute rates of return, my results confirm prior evidence in the literature concerning the existence of strong volatility clustering at the inter-daily level.

It would be interesting to extend the estimator developed here to more complicated continuous time jump-diffusion and multi-factor diffusion processes. More ambitious empirical applications might also entail the estimation of multivariate diffusions, which in turn would require vector versions of the integrated volatility and quadratic variation measures exploited here. Another interesting extension, would be to use the distributional features of the integrated volatility in pricing financial options, although this would necessitate additional assumptions about the price of volatility risk. I leave further work along these lines for future research.

5.6   Tables and Figures

158

Table 5.1: Monte Carlo Experiment——Panel A

The table reports the simulation results for the GMM and QML procedures discussed in the main text applied in estimating the stochastic volatility diffusion in equation (5.3). The total number of Monte Carlo replications is 1,000

True Value	Mean		Median		RMSE	
GMM with Quadratic Variation from High-Frequency Return						
	T = 1000	T = 4000	T = 1000	T = 4000	T = 1000	T = 4000
$\kappa = 0.03$	0.0352	0.0313	0.0340	0.0310	0.0130	0.0054
$\theta = 0.25$	0.2430	0.2487	0.2355	0.2460	0.0523	0.0258
$\sigma = 0.10$	0.1016	0.1030	0.1018	0.1030	0.0080	0.0050
GMM with Integrated Volatility						
	T = 1000	T = 4000	T = 1000	T = 4000	T = 1000	T = 4000
$\kappa = 0.03$	0.0382	0.0323	0.0374	0.0319	0.0139	0.0055
$\theta = 0.25$	0.2338	0.2456	0.2273	0.2437	0.0521	0.0257
$\sigma = 0.10$	0.0992	0.0999	0.0992	0.0998	0.0044	0.0020
QML with Point-in-Time Volatility						
	T = 1000	T = 4000	T = 1000	T = 4000	T = 1000	T = 4000
$\kappa = 0.03$	0.0446	0.0360	0.0434	0.0361	0.0195	0.0095
$\theta = 0.25$	0.2327	0.2441	0.2271	0.2410	0.0537	0.0290
$\sigma = 0.10$	0.1012	0.1014	0.0999	0.1011	0.0095	0.0052

159

Table 5.2: Monte Carlo Experiment——Panel B

The table reports the simulation results for the GMM and QML procedures discussed in the main text applied in estimating the stochastic volatility diffusion in equation (5.3). The total number of Monte Carlo replications is 1,000

True Value	Mean		Median		RMSE	
GMM with Quadratic Variation from High-Frequency Return						
	T = 1000	T = 4000	T = 1000	T = 4000	T = 1000	T = 4000
$\kappa = 0.10$	0.1057	0.1023	0.1048	0.1016	0.0214	0.0100
$\theta = 0.25$	0.2478	0.2491	0.2474	0.2489	0.0158	0.0078
$\sigma = 0.10$	0.1059	0.1073	0.1061	0.1072	0.0093	0.0082
GMM with Integrated Volatility						
	T = 1000	T = 4000	T = 1000	T = 4000	T = 1000	T = 4000
$\kappa = 0.10$	0.1102	0.1032	0.1090	0.1027	0.0214	0.0091
$\theta = 0.25$	0.2460	0.2486	0.2459	0.2483	0.0163	0.0078
$\sigma = 0.10$	0.0994	0.1000	0.0995	0.0998	0.0042	0.0020
QML with Point-in-Time Volatility						
	T = 1000	T = 4000	T = 1000	T = 4000	T = 1000	T = 4000
$\kappa = 0.10$	0.1136	0.1040	0.1134	0.1048	0.0259	0.0138
$\theta = 0.25$	0.2497	0.2517	0.2480	0.2510	0.0196	0.0097
$\sigma = 0.10$	0.0967	0.0956	0.0967	0.0958	0.0059	0.0054

160

Table 5.3: Monte Carlo Experiment——Panel C

The table reports the simulation results for the GMM and QML procedures discussed in the main text applied in estimating the stochastic volatility diffusion in equation (5.3). The total number of Monte Carlo replications is 1,000

True Value	Mean		Median		RMSE	
GMM with Quadratic Variation from High-Frequency Return						
	T = 1000	T = 4000	T = 1000	T = 4000	T = 1000	T = 4000
$\kappa = 0.10$	0.1113	0.1035	0.1091	0.1035	0.0253	0.0111
$\theta = 0.25$	0.2389	0.2468	0.2364	0.2463	0.0326	0.0158
$\sigma = 0.20$	0.2031	0.2051	0.2030	0.2049	0.0122	0.0078
GMM with Integrated Volatility						
	T = 1000	T = 4000	T = 1000	T = 4000	T = 1000	T = 4000
$\kappa = 0.10$	0.1153	0.1048	0.1131	0.1047	0.0270	0.0114
$\theta = 0.25$	0.2346	0.2455	0.2319	0.2449	0.0341	0.0160
$\sigma = 0.20$	0.1984	0.1997	0.1982	0.1995	0.0097	0.0046
QML with Point-in-Time Volatility						
	T = 1000	T = 4000	T = 1000	T = 4000	T = 1000	T = 4000
$\kappa = 0.10$	0.1257	0.1093	0.1242	0.1107	0.0390	0.0208
$\theta = 0.25$	0.2459	0.2537	0.2432	0.2520	0.0336	0.0199
$\sigma = 0.20$	0.1977	0.1960	0.1966	0.1958	0.0135	0.0084

161

Table 5.4: Monte Carlo Experiment with Measurement Error Correction
The table reports the GMM estimation results obtained by including an additive measurement error correction term, γ, in the moment conditions involving the squared integrated volatility. The RMSE column for γ gives the sample standard deviation across the 1,000 Monte Carlo replications.

True Value	Mean		Median		RMSE	
	Scenario A: GMM with Quadratic Variation					
	T = 1000	T = 4000	T = 1000	T = 4000	T = 1000	T = 4000
$\kappa = 0.03$	0.0364	0.0317	0.0354	0.0315	0.0138	0.0056
$\theta = 0.25$	0.2456	0.2491	0.2384	0.2464	0.0520	0.0257
$\sigma = 0.10$	0.0909	0.0994	0.0905	0.0983	0.0230	0.0127
γ	0.0007	0.0004	0.0006	0.0004	0.0008	0.0005
	Scenario B: GMM with Quadratic Variation					
	T = 1000	T = 4000	T = 1000	T = 4000	T = 1000	T = 4000
$\kappa = 0.10$	0.1067	0.1027	0.1061	0.1023	0.0219	0.0104
$\theta = 0.25$	0.2489	0.2494	0.2484	0.2492	0.0157	0.0078
$\sigma = 0.10$	0.0990	0.1049	0.0986	0.1046	0.0214	0.0121
γ	0.0007	0.0004	0.0006	0.0003	0.0009	0.0005
	Scenario C: GMM with Quadratic Variation					
	T = 1000	T = 4000	T = 1000	T = 4000	T = 1000	T = 4000
$\kappa = 0.10$	0.1133	0.1042	0.1109	0.1043	0.0274	0.0119
$\theta = 0.25$	0.2435	0.2481	0.2400	0.2473	0.0314	0.0157
$\sigma = 0.20$	0.1893	0.1999	0.1884	0.1987	0.0303	0.0162
γ	0.0017	0.0010	0.0015	0.0009	0.0019	0.0013

162

Table 5.5: Summary Statistics for Daily Integrated Volatility
The daily integrated volatilities are approximated by the quadratic variations constructed from five-minute returns.

Statistics	DM/$ Rate	Yen/$ Rate	Nikkei 225
Mean	0.5290	0.5383	0.8511
Std. Dev.	0.4839	0.5217	0.7757
Skewness ·	3.7083	5.5713	3.0203
Kurtosis	24.0505	66.6545	18.1780
Minimum	0.0517	0.0280	0.0309
5% Quant.	0.1384	0.1382	0.1494
25% Quant.	0.2542	0.2533	0.3681
Medium	0.3990	0.4008	0.6479
75% Quant.	0.6252	0.6317	1.0782
95% Quant.	1.3450	1.3598	2.2491
Maximum	5.2453	10.0971	7.5651
Num. of Obs.	2445	2445	984

Table 5.6: GMM Estimation of Stochastic Volatility Model
The GMM estimator and the specification test are described in Section 5.2. The daily integrated volatilities are approximated by the quadratic variations from five-minute returns. The variance-covariance matrix is estimated using a Newey-West weighting scheme with a lag-length of five.

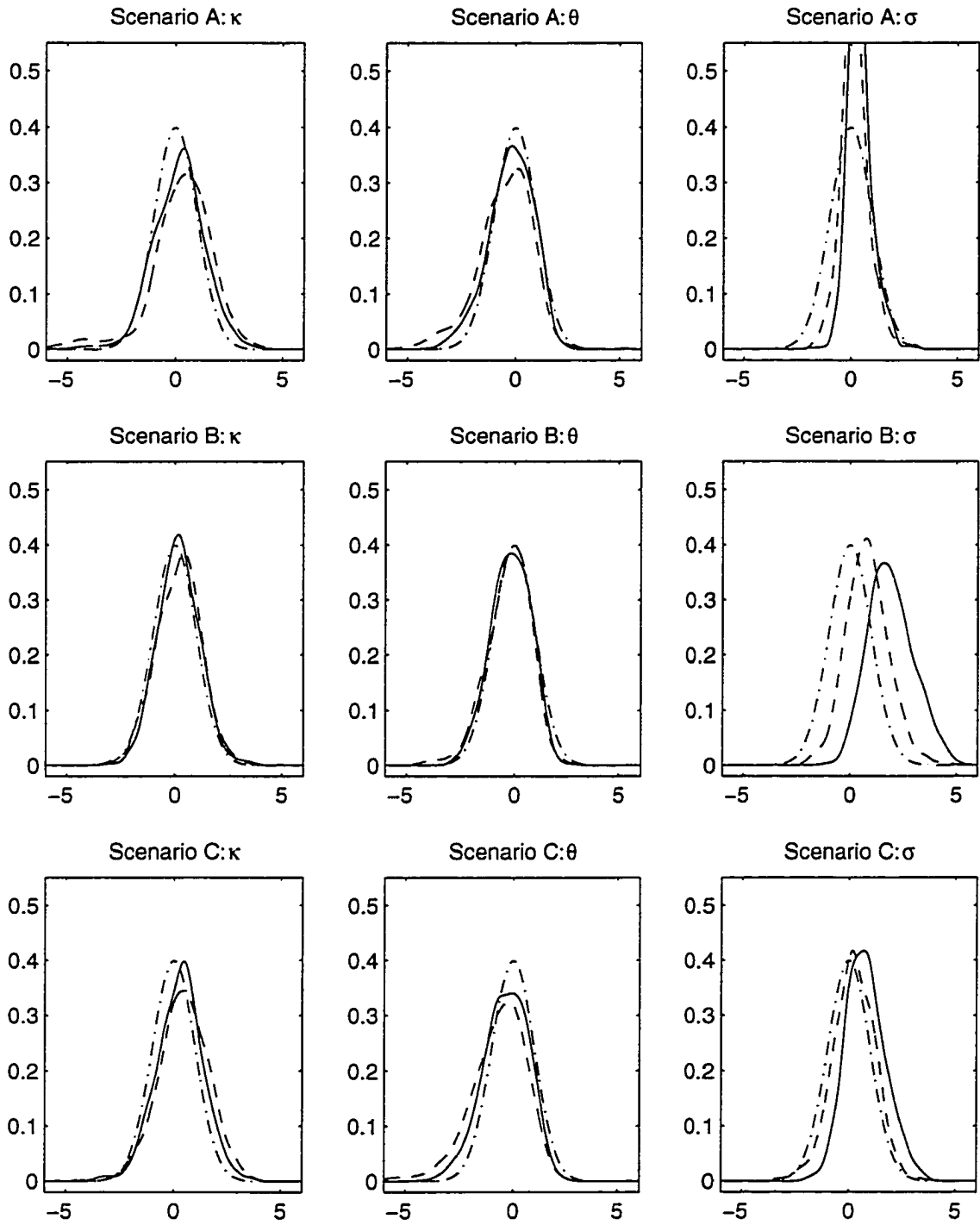| Parameter	DM/$ Rate	Yen/$ Rate	Nikkei 225
Mean Reversion κ	0.1464	0.2472	0.1236
(Standard Error)	(0.0387)	(0.0463)	(0.0492)
Long-run Mean θ	0.5172	0.5190	0.8312
(Standard Error)	(0.0342)	(0.0240)	(0.0950)
Local Variance σ	0.5789	0.4242	0.1909
(Standard Error)	(0.0580)	(0.1804)	(0.3992)
GMM Specification Test			
Chi-Square (2)	12.1476	3.6182	0.8040
p-Value	0.0023	0.1638	0.6690

163

Figure 5.1: T-test Distributions
"- - -" t-statistics with 1000 observations; "——" t-statistics with 4000 observations;
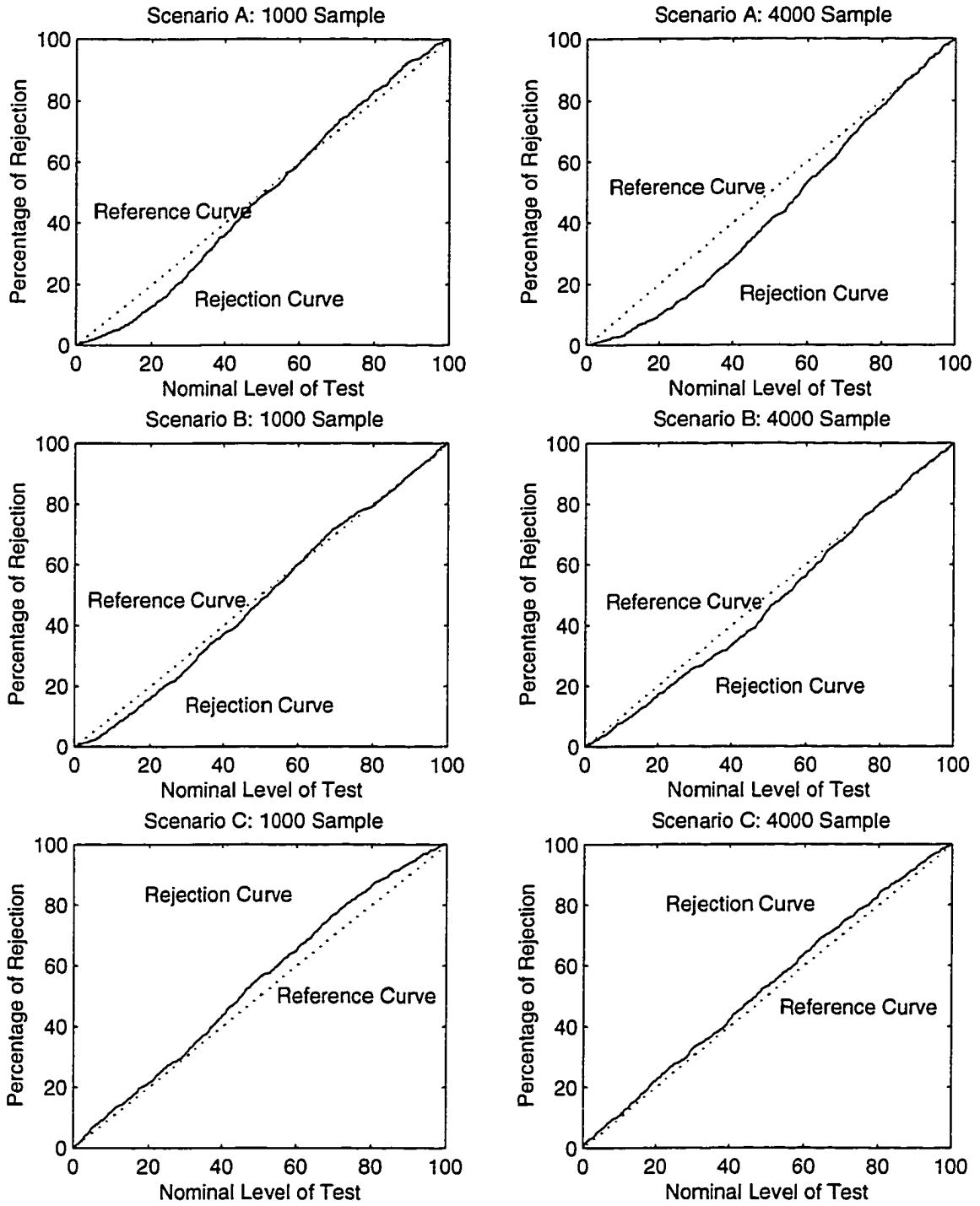"-.-.-" Normal (0,1) reference density.

Figure 5.2: GMM Specification Test of Overidentifying Restrictions.
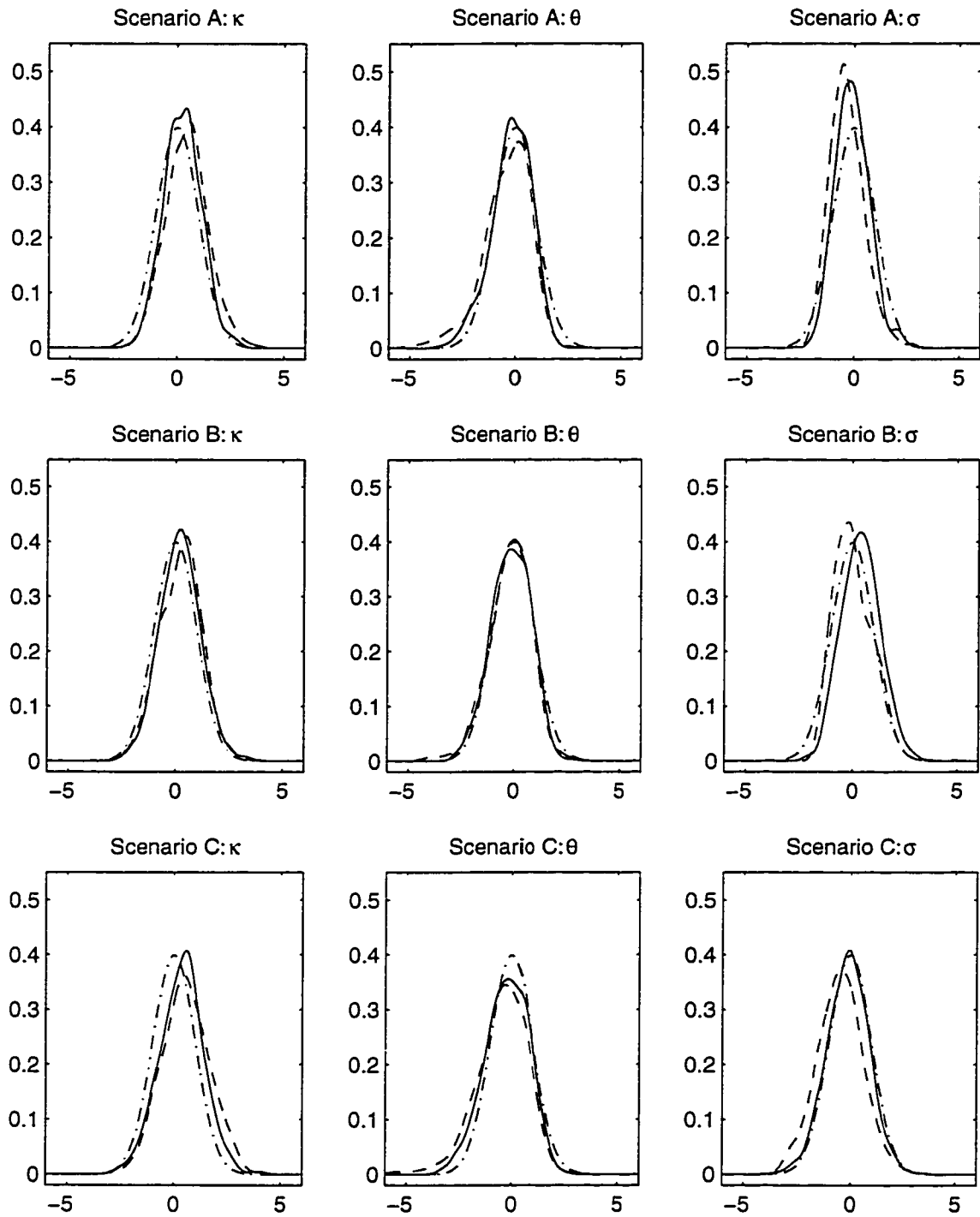
165

Figure 5.3: T-test Distributions with Measurement Error Correction
"- - -" t-statistics with 1000 observations; "——" t-statistics with 4000 observations; "-.-.-" Normal (0,1) reference density.
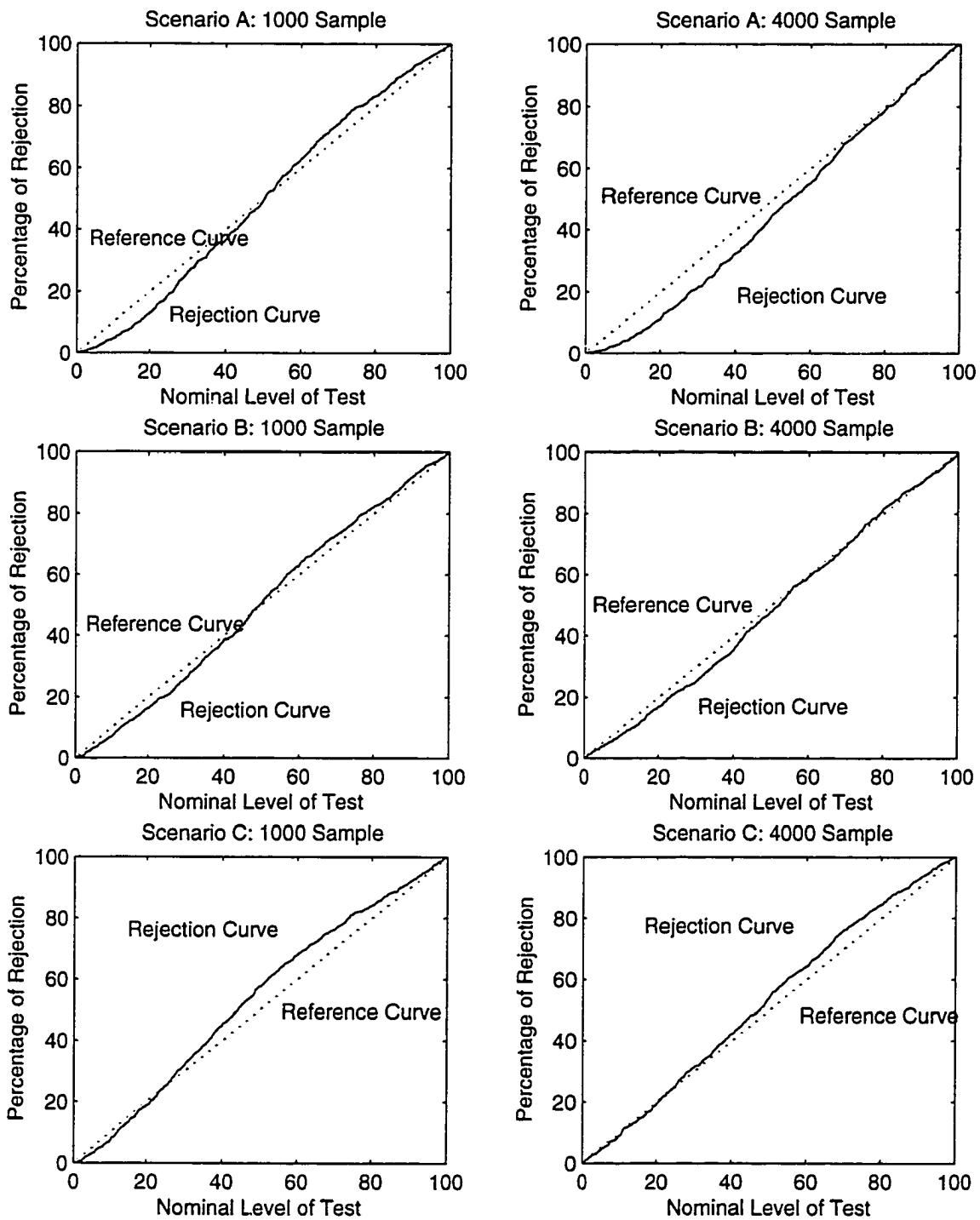
166

Figure 5.4: GMM Specification Test of Overidentifying Restrictions with Measurement Error Correction
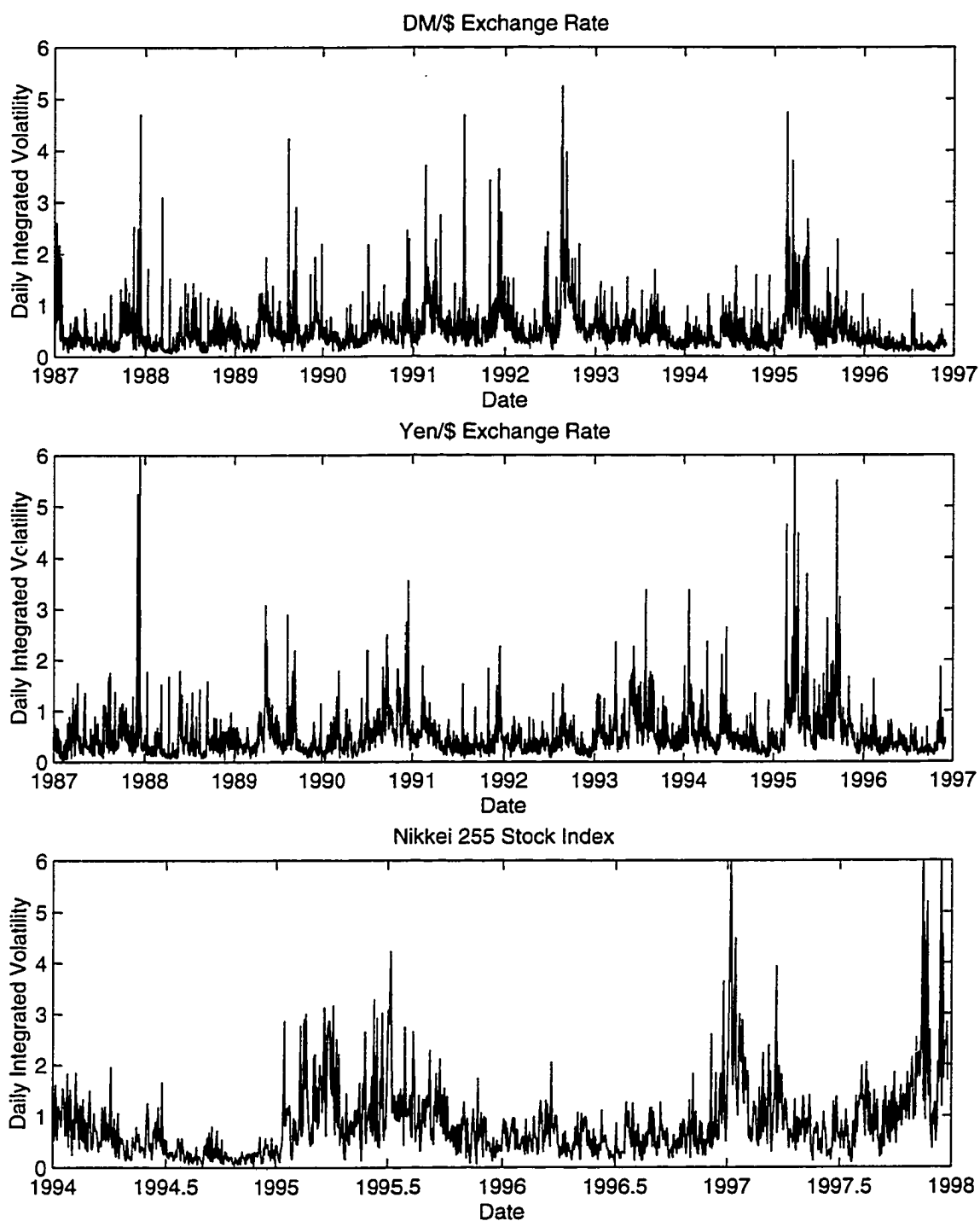
167

Figure 5.5: Daily Integrated Volatility on Financial Markets

168

Appendix A

Technical Derivations in Chapter 3

A.1   Martingale Pricing Result

Consider a simple homogeneous economy with free lending and borrowing at a "risk free" interest rate. Assume that the representative agent has a strictly increasing concave utility function $U(\cdot)$ and seeks to maximize his life-time expected utility by optimally choosing between consumption, saving, and portfolio investment, subject to the wealth constraint. Further assume that the market is complete and competitive, no transaction costs or short-sale restrictions. Under technical differentiable and measurable conditions (Cox et al. 1985b), there exists a unique equilibrium in such an exchange economy or a production economy with constant return to scale. The optimal solution is a bundle of stochastic consumption, investments, interest rates and asset returns. In equilibrium, the interest rate and the risky asset return must adjust simultaneously with the investment portfolio, until all the expected returns weighted by the marginal utility equal the risk-free rate.

A.1.1   Sufficient Condition for No-Arbitrage

The market price $P(t)$ of a risky asset without dividend must satisfy the equilibrium first order condition

$$P(t) = E_t \left[\frac{U'(s)}{U'(t)} P(s) \right], \text{ for any } s \geq t.$$

169

Economic equilibrium is the sufficient condition for no-arbitrage (Duffie 1996), which implies the existence of a *state price deflater* $\pi(t)$ such that

$$P(t) = E_t\left[\frac{\pi(s)}{\pi(t)}P(s)\right], \text{ for any } s \geq t. \tag{A.1}$$

This condition is fully justified by assuming a strictly increasing concave utility function; hence a positive stochastic discount factor exists up to a constant scale. Now our task is simply to characterize the discount factor $\pi(t)$, without having to fully specify the underlying economic structure. In order to do so, the consumption-saving decision must be separable from the portfolio investment decision. Typically the utility function is assumed to be a constant relative risk aversion (CRRA), which guarantees that the optimal investment decision is independent of wealth and is only dependent on its own return relative to the risk-free rate.

A.1.2 Necessary Condition for No-Arbitrage

The necessary condition for no-arbitrage is the existence of market-price-of-risk processes for both diffusion and jump risks. Intuitively, if any risk in a jump-diffusion economy is appropriately priced, there would exists no chance to make a profit from a zero-investment portfolio. Therefore, the stochastic discount factor must take the following form (Merton 1976, Bates 1996, Das 1998):

$$\frac{d\pi(t)}{\pi(t)} = -r(t)dt - \Lambda_W(t)dW(t) - \Lambda_J(t)[dN(\rho(\cdot)t) - \rho(\cdot)dt], \tag{A.2}$$

where $\Lambda_W(t)$ is the unit risk premium of diffusion and $\Lambda_J(t)$ the unit risk premium of jump. The *compensated Poisson process* $[dN(\rho(\cdot)t) - \rho(\cdot)dt]$ is a martingale (Karatzas and Shreve 1997). The functional forms of the market prices of risks $\Lambda_W(t)$ and $\Lambda_J(t)$, although not unique, must be proportional to the instantaneous standard deviation of the diffusion and jump risks. Once the pure diffusion risk and the net jump risk are

170

properly priced, the expected return of the pricing kernel is nothing but the risk-free rate

$$E_t \left[\frac{d\pi(t)}{\pi(t)} \right] = -r(t)dt.$$

A.1.3 A Characterization of Bond Price

To ease the burden of notations, we suppress the time subscript t in those intermediate steps which do not cause confusions. Also, the jump-rate ρ and jump-size J can be state-dependent, without further notice. The price $P(t)$ of a discount bond, as a function of the stochastic interest rate, also assumes a jump-diffusion process

$$\frac{dP(t)}{P(t)} = \mu_P(t)dt + \sigma_P(t)dW(t) + J_P(t)dN(\rho(\cdot)t), \qquad (A.3)$$

where $\mu_P(t)$, $\sigma_P(t)$, and $J_P(t)$ are respectively the drift, diffusion, and jump-size functions, which are dependent on the underlying short-rate process and the functional form of the bond price.

Define a new process $M(t) = \pi(t)P(t)$, which is a martingale due to the equilibrium first-order condition (A.1); hence

$$M(t) = E_t[M(s)], \text{ for any } s \geq t.$$

Since $M(t)$ is defined as a martingale, it must have zero drift $\mu_M(t) = 0$. Applying the Generalized Itô's Lemma (Kushner 1967, Ikeda and Watanabe 1981, Merton 1971, Lo 1988), one can show that

$$\mu_P(t) = r(t) + \Lambda_W(t)\sigma_P(t) + \rho(\cdot)(\Lambda_J(t) - 1)E[J_P(t)], \qquad (A.4)$$

i.e., the instantaneous bond return should be equal to the risk-free rate plus risk premiums. The expectation operator is taken with respect to the left continuous filtration $\mathcal{F}_t^-$. The total diffusion risk premium of the bond return is the product

171

of the unit price of risk $\Lambda_W(t)$ with the total diffusion risk $\sigma_P(t)$. The total jump risk premium of the bond return is the product of trend-adjusted unit price of jump risk is $(\Lambda_J(t) - 1)$ with the total expected jump impact $\rho(\cdot)E[J_P(t)]$. This result is a simple extension of Vasicek (1977) and Cox et al. (1985a).

A.2 Characterization of Conditional Moment

The analytical solution to conditional moments are given in Section 3.3.1 as

$$E_t(R_T) = e^{(T-t)A}R_t + e^{TA}A^{-1}(e^{-tA} - e^{-TA})g,$$

which is fully characterized by the fundamental matrix A and the forcing function g.

Model 1: Square-Root Diffusion.

$$A = \begin{bmatrix} -\kappa & 0 & 0 & 0 \\ 2\kappa\theta + \sigma^2 & -2\kappa & 0 & 0 \\ 0 & 3\kappa\theta + 3\sigma^2 & -3\kappa & 0 \\ 0 & 0 & 4\kappa\theta + 6\sigma^2 & -4\kappa \end{bmatrix}, \quad g = \begin{bmatrix} \kappa\theta \\ 0 \\ 0 \\ 0 \end{bmatrix}.$$

Model 2: Square-Root Diffusion with Independent Jump Term.

$$A = \begin{bmatrix} -\kappa & 0 & 0 & 0 \\ 2\kappa\theta + \sigma^2 + 2\rho E_J(J) & -2\kappa & 0 & 0 \\ 3\rho E_J(J^2) & 3\kappa\theta + 3\sigma^2 + 3\rho E_J(J) & -3\kappa & 0 \\ 4\rho E_J(J^3) & 6\rho E_J(J^2) & 4\kappa\theta + 6\sigma^2 + 4\rho E_J(J) & -4\kappa \end{bmatrix},$$

$$g = \begin{bmatrix} \rho E_J(J) + \kappa\theta \\ \rho E_J(J^2) \\ \rho E_J(J^3) \\ \rho E_J(J^4) \end{bmatrix}.$$

Model 3: Square-Root Diffusion with State-Dependent Jump-Rate

$$A = \begin{bmatrix} -\kappa + \rho_1 E_J(J) & 0 \\ 2\kappa\theta + \sigma^2 + 2\rho_0 E_J(J) + \rho_1 E_J(J^2) & -2\kappa + 2\rho_1 E_J(J) \\ 3\rho_0 E_J(J^2) + \rho_1 E_J(J^3) & 3\kappa\theta + 3\sigma^2 + 3\rho_0 E_J(J) + 3\rho_1 E_J(J^2) \\ 4\rho_0 E_J(J^3) + \rho_1 E_J(J^4) & 6\rho_0 E_J(J^2) + 4\rho_1 E_J(J^3) \end{bmatrix}$$

172

$$\begin{bmatrix} 0 & 0 \\ 0 & 0 \\ -3\kappa + 3\rho_1 E_J(J) & 0 \\ 4\kappa\theta + 6\sigma^2 + 4\rho_0 E_J(J) + 6\rho_1 E_J(J^2) & -4\kappa + 4\rho_1 E_J(J) \end{bmatrix},$$

$$g = \begin{bmatrix} \rho_0 E_J(J) + \kappa\theta \\ \rho_0 E_J(J^2) \\ \rho_0 E_J(J^3) \\ \rho_0 E_J(J^4) \end{bmatrix}.$$

Model 4: Square-Root Diffusion with State-Dependent Jump-Size.

$$A = \begin{bmatrix} -\kappa - \rho & 0 & 0 & 0 \\ 2\kappa\theta + \sigma^2 & -2\kappa - \rho & 0 & 0 \\ 0 & 3\kappa\theta + 3\sigma^2 & -3\kappa - \rho & 0 \\ 0 & 0 & 4\kappa\theta + 6\sigma^2 & -4\kappa - \rho \end{bmatrix}, \quad g = \begin{bmatrix} \kappa\theta + \rho J_0 \\ \rho J_0^2 \\ \rho J_0^3 \\ \rho J_0^4 \end{bmatrix}.$$

173

Appendix B

Technical Derivations in Chapter 5

B.1 Conditional Moments of Integrated Volatilities

B.1.1 Conditional Mean of Integrated Volatility

Because of the linear drift specification of the stochastic volatility, the conditional mean of the integrated volatility can be shown as a linear function of the point-in-time volatility

$$
\begin{aligned}
E\left(\mathcal{V}_{t,T}|\mathcal{F}_t\right) &= E\left(\int_t^T V_s ds \,\middle|\, \mathcal{F}_t\right) \\
&= \int_t^T E(V_s|\mathcal{F}_t)ds \\
&= \int_t^T \left[V_t e^{-\kappa(s-t)} + \theta\left(1 - e^{-\kappa(s-t)}\right) \right] ds \\
&= V_t \frac{1}{\kappa}\left(1 - e^{-\kappa(T-t)}\right) + \theta(T-t) - \frac{\theta}{\kappa}\left(1 - e^{-\kappa(T-t)}\right) \\
&= a_{T-t} V_t + b_{T-t},
\end{aligned}
\tag{B.1}
$$

where a_{T-t} and b_{T-t} are functions of the drift parameters and the time difference $(T - t)$. For notational simplicity we denote the parameters for the daily horizon, or $T - t = 1$, by $a \equiv \frac{1}{\kappa}(1 - e^{-\kappa})$ and $b \equiv \theta - \frac{\theta}{\kappa}(1 - e^{-\kappa})$. The above derivation explicitly uses the conditional mean of the point-in-time volatility

$$
E(V_T|\mathcal{F}_t) = V_t e^{-\kappa(T-t)} + \theta\left(1 - e^{-\kappa(T-t)}\right) = \alpha_{T-t} V_t + \beta_{T-t},
\tag{B.2}
$$

where α_{T-t} and β_{T-t} are also functions of the drift parameters and the time difference $(T - t)$. Again for $T - t = 1$, we define $\alpha \equiv e^{-\kappa}$ and $\beta \equiv \theta(1 - e^{-\kappa})$.

174

Focusing on the one-day horizon and using $E\left(\mathcal{V}_{t+1,t+2}|\mathcal{F}_{t+1}\right) = aV_{t+1} + b$ and $E(V_{t+1}|\mathcal{F}_t) = \alpha V_t + \beta$, it follows that

$$
\begin{aligned}
E\left[E\left(\mathcal{V}_{t+1,t+2}|\mathcal{F}_{t+1}\right)|\mathcal{F}_t\right] &= aE(V_{t+1}|\mathcal{F}_t) + b \\
&= a(\alpha V_t + \beta) + b \\
&= \alpha[E\left(\mathcal{V}_{t,t+1}|\mathcal{F}_t\right) - b] + a\beta + b,
\end{aligned}
$$

which simplifies as

$$
E\left(\mathcal{V}_{t+1,t+2}|\mathcal{F}_t\right) = \alpha E\left(\mathcal{V}_{t,t+1}|\mathcal{F}_t\right) + \beta.
$$

Finally, by the Law of Iterated Expectations,

$$
E\left[E\left(\mathcal{V}_{t+1,t+2}|\mathcal{F}_t\right)|\mathcal{G}_t\right] = E\left(\mathcal{V}_{t+1,t+2}|\mathcal{G}_t\right) = \alpha E\left(\mathcal{V}_{t,t+1}|\mathcal{G}_t\right) + \beta. \tag{B.3}
$$

B.1.2 Conditional Variance of Integrated Volatility

By definition $\mathcal{V}_{t,T} = \int_t^T V_s ds$, and from equation (B.1) $E\left(\mathcal{V}_{t,T}|\mathcal{F}_t\right) = a_{T-t}V_t + b_{T-t}$. The stochastic differential equation (SDE) for $E\left(\mathcal{V}_{t,T}|\mathcal{F}_t\right)$ may therefore be generated as a function of V_t by applying Itô's formula to the affine diffusion in equation (5.3),[1]

$$
dE\left(\mathcal{V}_{t,T}|\mathcal{F}_t\right) = [a_{T-t}\kappa(\theta - V_t) + \frac{\partial a_{T-t}}{\partial t}V_t + \frac{\partial b_{T-t}}{\partial t}]dt + a_{T-t}\sigma\sqrt{V_t}dW_t,
$$

which may be further simplified as

$$
dE\left(\mathcal{V}_{t,T}|\mathcal{F}_t\right) = -V_t dt + a_{T-t}\sigma\sqrt{V_t}dW_t. \tag{B.4}
$$

Now fix the upper limit T, and let the lower limit t be time varying. The Itô integral implied by SDE (B.4) then takes the form

$$
E\left(\mathcal{V}_{T,T}|\mathcal{F}_T\right) = E\left(\mathcal{V}_{t,T}|\mathcal{F}_t\right) + \int_t^T (-V_s)ds + \int_t^T a_{T-s}\sigma\sqrt{V_s}dW_s.
$$

---

[1]The simple version of Itô's Lemma for a smooth function $f(V_t, t, T) \in C^2$ of a diffusion process V_t states that

$$
df(V_t, t, T) = [f_V(V_t, t, T)\mu(V_t, t) + f_t(V_t, t, T) + \frac{1}{2}f_{VV}(V_t, t, T)v^2(V_t, t)]dt + f_V(V_t, t, T)v(V_t, t)dW_t,
$$

where $\mu(V_t, t)$ and $v(V_t, t)$ are the drift and diffusion functions defining the V_t process.

175

However, $E\left(\mathcal{V}_{T,T}|\mathcal{F}_T\right) = 0$, which implies that

$$\mathcal{V}_{t,T} - E\left(\mathcal{V}_{t,T}|\mathcal{F}_t\right) = \int_t^T a_{T-s}\sigma\sqrt{V_s}dW_s.$$

Using standard arguments in stochastic calculus, it follows from the substitution of equation (B.2) that

$$
\begin{aligned}
Var(\mathcal{V}_{t,T}|\mathcal{F}_t) &= E[(\mathcal{V}_{t,T} - E\left(\mathcal{V}_{t,T}|\mathcal{F}_t\right))^2|\mathcal{F}_t] \\
&= E\left\{\left[\int_t^T a_{T-s}\sigma\sqrt{V_s}dW_s\right]^2\bigg|\mathcal{F}_t\right\} \\
&= \int_t^T a_{T-s}^2\sigma^2 E(V_s|\mathcal{F}_t)ds \\
&= \int_t^T a_{T-s}^2\sigma^2[\alpha_{s-t}V_t + \beta_{s-t}]ds \\
&= A_{T-t}V_t + B_{T-t},
\end{aligned}
\tag{B.5}
$$

where

$$A_{T-t} = \frac{\sigma^2}{\kappa^2}\left[\frac{1}{\kappa} - 2e^{-\kappa(T-t)}(T-t) - \frac{1}{\kappa}e^{-2\kappa(T-t)}\right],$$

$$
\begin{aligned}
B_{T-t} &= \frac{\sigma^2}{\kappa^2}\left[\theta(T-t)\left(1+2e^{-\kappa(T-t)}\right) - \frac{3\theta}{\kappa}\left(1-e^{-\kappa(T-t)}\right) + \frac{\theta}{2\kappa}\left(1-e^{-\kappa(T-t)}\right)^2\right] \\
&= \frac{\sigma^2}{\kappa^2}\left[\theta(T-t)\left(1+2e^{-\kappa(T-t)}\right) + \frac{\theta}{2\kappa}\left(e^{-\kappa(T-t)}+5\right)\left(e^{-\kappa(T-t)}-1\right)\right].
\end{aligned}
$$

In particular, the conditional variance of the integrated volatility is a linear function of the point-in-time volatility. It follows also from Cox et al. (1985a) and equation (B.2) above that,

$$
\begin{aligned}
E(V_T^2|\mathcal{F}_t) &= Var(V_T|\mathcal{F}_t) + [E(V_T|\mathcal{F}_t)]^2 \\
&= V_t\frac{\sigma^2}{\kappa}\left(e^{-\kappa(T-t)} - e^{-2\kappa(T-t)}\right) + \frac{\sigma^2\theta}{2\kappa}\left(1-e^{-\kappa(T-t)}\right)^2 + [\alpha_{T-t}V_t + \beta_{T-t}]^2
\end{aligned}
$$

176

$$
\begin{aligned}
&= C_{T-t}V_t + D_{T-t} + \alpha_{T-t}^2 V_t^2 + \beta_{T-t}^2 + 2\alpha_{T-t}\beta_{T-t}V_t \\
&= \alpha_{T-t}^2 V_t^2 + [C_{T-t} + 2\alpha_{T-t}\beta_{T-t}]V_t + [D_{T-t} + \beta_{T-t}^2]
\end{aligned}
\tag{B.6}
$$

where

$$
C_{T-t} = \frac{\sigma^2}{\kappa}\left(e^{-\kappa(T-t)} - e^{-2\kappa(T-t)}\right),
$$

$$
D_{T-t} = \frac{\sigma^2\theta}{2\kappa}\left(1 - e^{-\kappa(T-t)}\right)^2.
$$

Focusing on the one-day horizon and using the conditional variance formula (B.5), $Var(\mathcal{V}_{t,t+1}|\mathcal{F}_t) = AV_t + B$, and the corresponding one-day conditional mean formula (B.1), $E(\mathcal{V}_{t,t+1}|\mathcal{F}_t) = aV_t + b$, implies that

$$
E(\mathcal{V}_{t,t+1}^2|\mathcal{F}_t) = Var(\mathcal{V}_{t,t+1}|\mathcal{F}_t) + [E(\mathcal{V}_{t,t+1}|\mathcal{F}_t)]^2 = a^2 V_t^2 + (2ab+A)V_t + (b^2+B) \tag{B.7}
$$

Leading the arguments by one period and applying the Law of Iterated Expectation, immediately yields

$$
E[E(\mathcal{V}_{t+1,t+2}^2|\mathcal{F}_{t+1})|\mathcal{F}_t] = a^2 E(V_{t+1}^2|\mathcal{F}_t) + (2ab+A)E(V_{t+1}|\mathcal{F}_t) + (b^2+B).
$$

Now substitute $E(V_{t+1}|\mathcal{F}_t)$ by (B.2) and $E(V_{t+1}^2|\mathcal{F}_t)$ by (B.6), and reversely substitute out V_t^2 by (B.7) and V_t by (B.1), it is clear that

$$
\begin{aligned}
E(\mathcal{V}_{t+1,t+2}^2|\mathcal{F}_t) &= a^2[\alpha^2 V_t^2 + (C + 2\alpha\beta)V_t + (D + \beta^2)] \\
&\quad + (2ab+A)(\alpha V_t + \beta) + (b^2+B) \\
&= \alpha^2 a^2 V_t^2 + [a^2(C + 2\alpha\beta) + \alpha(2ab+A)]V_t \\
&\quad + [a^2(D + \beta^2) + \beta(2ab+A) + (b^2+B)] \\
&= \alpha^2[E(\mathcal{V}_{t,t+1}^2|\mathcal{F}_t) - (2ab+A)V_t - (b^2+B)] \\
&\quad + [a^2(C + 2\alpha\beta) + \alpha(2ab+A)]V_t \\
&\quad + [a^2(D + \beta^2) + \beta(2ab+A) + (b^2+B)]
\end{aligned}
$$

177

$$= \alpha^2 E(\mathcal{V}_{t,t+1}^2|\mathcal{F}_t)$$

$$+[a^2(C + 2\alpha\beta) + (\alpha - \alpha^2)(2ab + A)]\frac{1}{a}[E(\mathcal{V}_{t,t+1}|\mathcal{F}_t) - b]$$

$$+[a^2(D + \beta^2) + \beta(2ab + A) + (1 - \alpha^2)(b^2 + B)]$$

$$= \alpha^2 E(\mathcal{V}_{t,t+1}^2|\mathcal{F}_t)$$

$$+\frac{1}{a}[a^2(C + 2\alpha\beta) + (\alpha - \alpha^2)(2ab + A)]E(\mathcal{V}_{t,t+1}|\mathcal{F}_t)$$

$$-\frac{b}{a}[a^2(C + 2\alpha\beta) + (\alpha - \alpha^2)(2ab + A)]$$

$$+[a^2(D + \beta^2) + \beta(2ab + A) + (1 - \alpha^2)(b^2 + B)] \qquad (B.8)$$

Lastly, applying the Law of Iterated Expectations to (B.8) and changing the information set, we have

$$E[E(\mathcal{V}_{t+1,t+2}^2|\mathcal{F}_t)|\mathcal{G}_t] = E(\mathcal{V}_{t+1,t+2}^2|\mathcal{G}_t)$$

$$= \alpha^2 E(\mathcal{V}_{t,t+1}^2|\mathcal{G}_t)$$

$$+\frac{1}{a}[a^2(C + 2\alpha\beta) + (\alpha - \alpha^2)(2ab + A)]E(\mathcal{V}_{t,t+1}|\mathcal{G}_t)$$

$$-\frac{b}{a}[a^2(C + 2\alpha\beta) + (\alpha - \alpha^2)(2ab + A)]$$

$$+[a^2(D + \beta^2) + \beta(2ab + A) + (1 - \alpha^2)(b^2 + B)]$$

$$= H E(\mathcal{V}_{t,t+1}^2|\mathcal{G}_t) + I E(\mathcal{V}_{t,t+1}|\mathcal{G}_t) + J, \qquad (B.9)$$

where $H = \alpha^2$, $I = 1/a[a^2(C + 2\alpha\beta) + (\alpha - \alpha^2)(2ab + A)]$, and $J = -b/a[a^2(C + 2\alpha\beta) + (\alpha - \alpha^2)(2ab + A)] + [a^2(D + \beta^2) + \beta(2ab + A) + (1 - \alpha^2)(b^2 + B)]$.

178

Bibliography

Aït-Sahalia, Yacine (1996a), "Nonparametric Pricing of Interest Rate Derivatives," *Econometrica*, vol. 64, 527–560.

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

Aït-Sahalia, Yacine (1998), "Maximum Likelihood Estimation of Discretely Sampled Diffusions: A Closed-Form Approach," *Working Paper*, Department of Economics, Princeton University.

Alizadeh, Sassan, Michael W. Brandt, and Francis X. Diebold (1999), "Range-Based Estimation of Stochastic Volatility Models," *Working Paper*, Department of Finance, NYU.

Andersen, Torben G., Luca Benzoni, and Jesper Lund (1999a), "Estimating Jump-Diffusions for Equity Returns," *Working Paper*, Kellogg Graduate School of Management, Northwestern University.

Andersen, Torben G. and Tim Bollerslev (1997), "Intraday Periodicity and Volatility Persistence in Financial Markets," *Journal of Empirical Finance*, vol. 4, 115–158.

Andersen, Torben G. and Tim Bollerslev (1998a), "Answering the Skeptics: Yes, Standard Volatility Models Do Provide Accurate Forcasts," *International Economic Review*, vol. 39, 885–905.

Andersen, Torben G. and Tim Bollerslev (1998b), "DM-Dollar Volatility: Intraday Activity Patterns, Macroeconomic Announcements, and Longer-Run Dependencies," *Journal of Finance*, vol. 53, 219–265.

Andersen, Torben G., Tim Bollerslev, and Jun Cai (2000a), "Intraday and Interday Volatility in the Nikkei 225 Index," *Journal of International Financial Markets, Institutions & Money*, forthcoming.

Andersen, Torben G., Tim Bollerslev, Francis X. Diebold, and Heiko Ebens (2000b), "The Distribution of Stock Return Volatility," *Working Paper*, Department of Economics, Duke University.

Andersen, Torben G., Tim Bollerslev, Francis X. Diebold, and Paul Labys (1999b), "The Distribution of Exchange Rate Volatility," *NBER Working Paper*, No.

179

6961.

Andersen, Torben G., Tim Bollerslev, Francis X. Diebold, and Paul Labys (2000c), "Microstructure Bias and Volatility Signatures," *Work in Progress*, Department of Economics, Duke University.

Andersen, Torben G., Hyung-Jin Chung, and Bent E. Sørenson (1999c), "Efficient Method of Moments Estimation of a Stochastic Volatility Model: A Monte Carlo Study," *Journal of Econometrics*, vol. 91, 61–87.

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

Andersen, Torben G. and Jesper Lund (1998), "Stochastic Volatility and Mean Drift in the Short Rate Diffusion: Source of Steepness, Level, and Curvature of the Yield Curve," *Working Paper*, Kellogg Graduate School of Management, Northwestern University.

Andersen, Torben G. and Bent E. Sørenson (1996), "GMM estimation of a stochastic volatility model: a monte Carlo study," *Journal of Business and Economic Statistics*, vol. 14, 328–352.

Ang, Andrew and Geert Bekaert (1998), "Regime Switches in Interest Rates," *Working Paper 6508*, NBER, Cambridge, MA.

Backus, David and Stanley Zin (1994), "Reverse Engineering the Yield Curve," *Working Paper 4676*, NBER, Cambridge, MA.

Bai, Xuezheng, Jeffrey R. Russell, and George C. Tiao (1999), "Beyond Merton's Utopia: Effects of Non-Normality and Dependence on the Precision of Variance Estimates Using High-Frequency Financial Data," *Working Paper*, Graduate School of Business, University of Chicago.

Bandi, Federico M. and Peter C. B. Phillips (1999), "Econometric Estimation of Diffusion Models," *Working Paper*, Department of Economics, Yale University.

Bansal, Ravi (1997), "An Exploration of the Forward Premium Puzzle in Currency Market," *Review of Financial Studies*, vol. 10, 369–403.

Bansal, Ravi and Wilbur John Coleman (1996), "A Monetary Explanation of the Equity Premium, Term Premium, and Risk-Free Rate Puzzles," *Journal of Political Economy*, vol. 104, 1135–1171.

Bansal, Ravi, A. Ronald Gallant, Robert Hussey, and George Tauchen (1995), "Non-parametric Estimation of Structural Models for High-Frequency Currency Market Data," *Journal of Econometrics*, vol. 66, 251–287.

Bansal, Ravi, David A. Hsieh, and Y. Charles Shen (1998), "Value-at-Risk in Non-Nested Affine Term Structure Models," *Working Paper*, Department of Economics, Duke University.

Bansal, Ravi and Hao Zhou (1999), "Term Structure of Interest Rates with Regime Shifts," *Working Paper*, Fuqua School of Business and Department of Economics, Duke University.

Barndorff-Nielsen, Ole and Neil Shephard (1999), "Non-Gaussian OU Based Models and Some of Their Uses in Financial Economics," *Working Paper*, Nuffield College, Oxford University.

Bates, David S. (1996), "Jumps and Stochastic Volatility: Exchange Rate Process Implicit in Deutsche Mark Options," *The Review of Financial Studies*, vol. 9, 69–107.

Bollerslev, Tim, Robert F. Engle, and Daniel B. Nelson (1994), "ARCH Models," in "Handbook of Econometrics," (edited by Engle, Robert F. and Daniel L. McFadden), vol. IV, Amsterdam: North-Holland.

Bollerslev, Tim and Jeffery Wooldridge (1992), "Quasi-Maximum Likelihood Estimators and Inference in Dynamic Models with Time-Varying Covariances," *Econometric Review*, vol. 11, 143–172.

Bollerslev, Tim and Hao Zhou (2000), "Estimating Stochastic Volatility Diffusions Using Conditional Moments of Integrated Volatility," *Working Paper*, Department of Economics, Duke University.

Brandt, Michael W. and Pedro Santa-Clara (1999), "Simulated Likelihood Estimation of Multivariate Diffusions with an Application to Interest Rates and Exchange Rates with Stochastic Volatility," *Working Paper*, Wharton School, University of Pennsylvania.

Brown, Stephen J. and Philip H. Dybvig (1986), "The Empirical Implications of the Cox, Ingersoll, Ross Theory of the Term Structure of Interest Rates," *Journal of Finance*, vol. 41, 617–630.

Burnside, Craig and Martin Eichenbaum (1996), "Small-Sample Properties of GMM-Based Wald Tests," *Journal of Business and Economic Statistics*, vol. 14, 294–

308.

Cai, Jun (1994), "A Markov Model of Switching-Regime ARCH," *Journal of Business and Economic Statistics*, vol. 12, 309–316.

Campbell, John Y., Andrew W. Lo, and A. Craig MacKinlay (1997), *The Econometrics of Financial Markets*, Princeton University Press.

Campbell, John Y. and Robert J. Shiller (1991), "Yield Spreads and Interest Rate Movements: A Bird's Eye View," *Review of Economic Studies*, vol. 58, 495–514.

Chacko, George and Luis Viceira (1999), "Spectral GMM Estimation of Continuous-Time Processes," *Working Paper*, Graduate School of Business, Harvard University.

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

Chan, Kalok, K. C. Chan, and G. Andrew Karolyi (1991), "Intraday Volatility in the Stock Index and Stock Index Futures Markets," *Review of Financial Studies*, vol. 4, 1161–1187.

Chen, Lin (1996), "Stochastic Mean and Stochastic Volatility - A Three-Factor Model of the Term Structure of Interest Rates and Its Applications in Derivatives Pricing and Risk Management," *Financial Markets, Institution and Instruments*, vol. 5, 1–88.

Chen, Ren-Raw and Louis Scott (1993), "Maximum Likelihood Estimation for a Multifactor Equilibrium Model of the Term Structure of Interest Rates," *Journal of Fixed Income*, vol. 3, 14–31.

Chernov, Mikhail, A. Ronald Gallant, Eric Ghysels, and George Tauchen (1999), "A New Class of Stochastic Volatility Models with Jumps: Theory and Estimation," *Working Paper*.

Chumacero, Rómulo A. (1997), "Finite Sample Properties of the Efficient Method of Moments," *Studies in Nonlinear Dynamics and Econometrics*, vol. 2, 35–51.

Conley, Tim, Lars Peter Hansen, Erzo Luttmer, and Jose Scheinkman (1997a), "Short Term Interest Rates as Subordinated Diffusions," *Review of Financial Studies*, vol. 10, 525–578.

182

Conley, Timothy G., Lars Peter Hansen, and Wen-Fang Liu (1997b), "Bootstrapping the Long Run," *Macroeconomic Dynamics*, vol. 1, 279–311.

Cox, John C., Jonathan E. Ingersoll, and Stephen A. Ross (1985a), "A Theory of the Term Structure of Interest Rates," *Econometrica*, vol. 53, 385–407.

Cox, John C., Jonathan E. Ingersoll, and Stephen A. Ross (1985b), "An Intertemporal General Equilibrium Model of Asset Prices," *Econometrica*, vol. 53, 363–384.

Dai, Qiang and Kenneth J. Singleton (2000), "Specification Analysis of Affine Term Structure Models," *Journal of Finance*, forthcoming.

Das, Sanjiv Ranjan (1998), "Poisson-Gaussian Process and the Bond Markets," *Working Paper*, Harvard Business School.

Devroye, Luc (1986), *Non-Uniform Random Variate Geteration*, Spinger-Verlag.

Drost, Feike C., Theo E. Nijman, and Bas J. M. Werker (1998), "Estimation and Testing in Models Containing Both Jumps and Conditional Heteroskedasticity," *Journal of Business and Economic Statistics*.

Duffie, Darrell (1996), *Dynamic Asset Pricing Theory*, Princeton University Press, 2nd ed.

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

Duffie, Darrell, Jun Pan, and Kenneth Singleton (1999), "Transform Analysis and Asset Pricing for Affine Jump-Diffusions," *Working Paper*, graduate School of Business, Stanford University.

Duffie, Darrell and Kenneth Singleton (1993), "Simulated Moments Estimation of Markov Models of Asset Prices," *Econometrica*, vol. 61, 929–952.

Duffie, Darrell and Kenneth Singleton (1997), "An Econometric Model of the Term Structure of Interest-Rate Swap Yields," *Journal of Finance*, vol. 52, 1287–1321.

Elerian, Ola, Siddhartha Chib, and Neil Shephard (1998), "Likelihood Inference for Discretely Observed Non-Linear Diffusions," *Working Paper*, Nuffield College, Oxford University.

Engle, Robert F. and Gloria González-Rivera (1991), "Semiparametric ARCH Models," *Journal of Business and Economic Statistics*, vol. 9, 345–359.

Engle, Robert F. and Gary G. J. Lee (1997), "Estimating Diffusion Models of Stochastic Volatility," in "Modeling Stock Market Volatility: Bridging the Gap to Continuous Time," (edited by Rossi, Peter E.), Academic Press, New York.

Eraker, Bjorn (1998), "MCMC Analysis of Diffusion Models with Application to Finance," *Working Paper*, Graduate School of Business, University of Chicago.

Evans, Martin D. D. (1998), "Regime Shifts, Risk and the Term Structure," *Working Paper*.

Fama, Eugene F. (1984), "The Term Premium in Bond Returns," *Journal of Financial Economics*, vol. 13, 529–564.

Fama, Eugene F. and Robert T. Bliss (1987), "The Information in Long-Maturity Forward Rates," *American Economic Review*, vol. 77, 680–692.

Feller, Wiliam (1971), *An Introduction to Probability Theory and Its Applications*, vol. 2, John Wiley & Sons, Inc., Princeton University, 2nd ed.

Feller, William (1951), "Two Singular Diffusion Problems," *Annals of Mathematics*, vol. 54, 173–182.

Fisher, Mark and Christian Gilles (1996), "Estimating Exponential Affine Models of the Term Structure," *Working Paper*, Finance and Economics Discussion Series, Federal Reserve Board.

Froyen, Richard T. (1996), *Macroeconomics*, Prentice-Hall, Inc.

Gallant, A. Ronald (1987), *Nonlinear Statistical Models*, Wiley, New York.

Gallant, A. Ronald, David Hsieh, and George Tauchen (1997), "Estimation of Stochastic Volatility Models with Diagnostics," *Journal of Econometrics*, vol. 81, 159–192.

Gallant, A. Ronald, Chien-Te Hsu, and George Tauchen (1999), "Using Daily Range Data to Calibrate Volatility Diffusions and Extract the Forward Integrated Variance," *Review of Economics and Statistics*, vol. 81, 617–631.

Gallant, A. Ronald and Jonathan R. Long (1997), "Estimating Stochastic Differential Equations Efficiently by Minimum Chi-Square," *Biometrika*, vol. 84.

184

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

Gallant, A. Ronald and George Tauchen (1992), "A Nonparametric Approach to Nonlinear Time Series Analysis: Estimation and Simulation," in "New Directions in Time Series Analysis, Part II.", (edited by Brillinger, David, Peter Caines, John Geweke, Emanuel Parzen, Murray Rosenblatt, and Murad S. Taqqu), pages 71–92, Springer-Verlag, New York.

Gallant, A. Ronald and George Tauchen (1996a), *User's Guide for EMM: A Program for Efficient Method of Moments Estimation*, 1st ed.

Gallant, A. Ronald and George Tauchen (1996b), "Which Moment to Match?" *Econometric Theory*, vol. 12, 657–681.

Gallant, A. Ronald and George Tauchen (1997), *User's Guide for SNP: A Program for Nonparametric Time Series Analysis*, 8th ed.

Gallant, A. Ronald and George Tauchen (1998a), "The Relative Efficiency of Method of Moments Estimators," *Working Paper*, Department of Economics, Duke University.

Gallant, A. Ronald and George Tauchen (1998b), "Reprojecting Partially Observed Systems with Application to Interest Rate Diffusions," *Journal of the American Statistical Association*, vol. 93, 10–24.

Garcia, René (1992), "Asymptotic Null Distribution of the Likelihood Ratio Test in Markov Switching Models," *Working Paper*, université de Montréal.

Garcia, René and Pierre Perron (1996), "An Analysis of the Real Interest Rate Under Regime Shifts," *Review of Economics and Statistics*, vol. 78, 111–125.

Ghysels, Eric, Andrew Harvey, and Eric Renault (1996), "Stochastic Volatility," in "Handbook of Statistics Vol 14., Statistical Method in Finance," (edited by Maddala, G. S.), Amsterdam: North-Holland.

Gibbons, Michael R. and Krishna Ramaswamy (1993), "A Test of the Cox, Ingersoll, and Ross Model of the Term Structure," *Review of Fianacial Studies*, vol. 6, 619–658.

Gill, Philip E., Walter Murray, Michael A. Saunders, and Margaret H. Wright (1991),

185

"User's Guide for NPSOL (Version 4.06): A Fortran Package for Nonlinear Programming," Tech. rep., Stanford University.

Gourieroux, Christian, Alain Monfort, and Eric Renault (1993), "Indirect Inference," *Journal of Applied Econometrics*, vol. 8, s85–s118.

Gray, Stephen F. (1996), "Modeling the Conditional Distribution of Interest Rates as A Regime-Switching Process," *Journal of Financial Economics*, vol. 42, 27–62.

Hamilton, James D. (1988), "Rational Expectations Econometric Analysis of Changes in Regimes: An Investigation of the Term Structure of Interest Rates," *Journal of Economic Dynamics and Control*, vol. 12, 385–423.

Hamilton, James D. (1989), "A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle," *Econometrica*, vol. 57, 357–384.

Hansen, Bruce E. (1992), "The Likelihood Ratio Test under Non-Standard Conditions: Testing the Markov Switching Model of GNP," *Journal of Applied Econometrics*, vol. 7, s61–s82.

Hansen, Lars Peter (1982), "Large Sample Properties of Generalized Method of Moments Estimators," *Econometrica*, vol. 50, 1029–1054.

Hansen, Lars Peter (1985), "A Method for Calculating Bounds on the Asymptotic Covariance Matrices of Generalized Method of Moments Estimators," *Journal of Econometrics*, vol. 30, 203–238.

Hansen, Lars Peter, John Heaton, and Amir Yaron (1996), "Finite-Sample Properties of Some Alternative GMM Estimators," *Journal of Business and Economic Statistics*, vol. 14, 262–280.

Hansen, Lars Peter, John C. Heaton, and Masao Ogaki (1988), "Efficiency Bounds Implied by Multiperiod Conditional Moment Restrictions," *Journal of the American Statistical Association*, vol. 83, 863–871.

Hansen, Lars Peter and Ravi Jagannathan (1997), "Assessing Specification Errors in Stochastic Discount Factor Models," *Journal of Finance*, vol. 52, 557–590.

Hansen, Lars Peter and Jose Alexandre Scheinkman (1995), "Back to the Future: Generalized Moment Implications for Continuous Time Markov Process," *Econometrica*, vol. 63, 767–804.

Heston, Steven (1993), "A Closed-Form Solution for Options with Stochastic Volatility with Applications to Bond and Currency Options," *Review of Finanacial Studies*, vol. 6, 327–343.

Hull, John and Alan White (1987), "The pricing of Options on Assets with Stochastic Volatility," *Journal of Finance*, vol. 42, 381–340.

Ikeda, Nobuyuki and Shinzo Watanabe (1981), *Stochastic Differential Equations and Diffusion Process*, Kodansha Ltd.

Ingersoll, Jonathan E. (1987), *Theory of Financial Decision Making*, Rowman & Littlefield.

Ingram, Beth F. and B. S. Lee (1991), "Simulation Estimation of Time Series Models," *Journal of Econometrics*, vol. 47, 197–205.

Jacquier, Eric, Nicholas G. Polson, and Peter E. Rossi (1994), "Bayesian Analysis of Stochastic Volatility Models," *Journal of Business and Economic Statistics*, vol. 12, 371–389.

Johnson, Norman L. and Samuel Kotz (1970), *Distributions in Statistics: Continuous Univariate Distributions*, vol. 2, John Wiley & Sons.

Karatzas, Ioannis and Steven E. Shreve (1997), *Brownian Motion and Stochastic Calculus*, Springer.

Kim, S., Neil Shephard, and Siddhartha Chib (1998), "Stochastic Volatility: Likelihood Inference and Comparison with ARCH Models," *Review of Economic Studies*, vol. 65, 361–394.

Kloeden, Perter E. and Eckhard Platen (1992), *Numerical Solution of Stochastic Differential Equations*, Applications of Mathematics, Springer-Verlag.

Kushner, Harold J. (1967), *Stochastic Stability and Control*, Academic Press.

Kushner, Harold J. and Paul G. Dupuis (1992), *Numerical Methods for Stochastic Control Problems in Continuous Time*, Springer-Verlag.

Ledoit, Olivier and Pedro Santa-Clara (1999), "Relative Pricing of Options with Stochastic Volatility," *Working Paper*, Anderson Graduate School of Management, UCLA.

187

Lo, Andrew W. (1988), "Maximum Likelihood Estimation of Generalized Itô Process with Discretely Sampled Data," *Econometric Theory*, vol. 4, 231–247.

Lo, Andrew W. and A. Craig MacKinlay (1990), "An Econometric Analysis of Non-synchronous Trading," *Journal of Econometrics*, vol. 45, 181–212.

Longstaff, Francis A. and Eduardo S. Schwartz (1992), "Interest Rate Volatility and the Term Structure: A Two-Factor General Equilibrium Model," *Journal of Finance*, vol. 47, 1259–1282.

Merton, Robert C. (1971), "Optimum Consumption and Portfolio Rules in a Continuous Time Model," *Journal of Economic Theory*, vol. 3, 373–413.

Merton, Robert C. (1976), "Option Pricing When Underlying Stock Returns Are Discontinuous," *Journal of Financial Economics*, vol. 3, 125–144.

Merton, Robert C. (1980), "On Estimating the Expected Return on the Market," *Journal of Financial Economics*, vol. 8, 323–361.

Merton, Robert C. (1990), *Continuous-Time Finance*, Blackwell.

Naik, Vasant and Moon Hoe Lee (1997), "Yield Curve Dynamics with Discrete Shifts in Economic Regimes: Theory and Estimation," *Working Paper*, Faculty of Commerce, University of British Columbia.

Nelson, Daniel B. (1992), "Filtering and Forecasting with Misspecified ARCH Models I: Getting the Right Variance with the Wrong Model," *Journal of Econometrics*, vol. 52, 61–90.

Nelson, Daniel B. and Dean P. Foster (1994), "Asymptotic Filtering Theory for Univariate ARCH Models," *Econometrica*, vol. 62, 1–41.

Newey, Whitney K. (1985), "Maximum Likelihood Specification Testing and Conditional Moment Tests," *Econometrica*, vol. 53, 1047–1070.

Newey, Whitney K. and Douglas G. Steigerwald (1997), "Asymptotic Bias for Quasi-Maximum-Likelihood Estimators in Conditional Heteroskedasticity Models," *Econometrica*, vol. 65, 587–599.

Newey, Whitney K. and Kenneth D. West (1987), "A Simple Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica*, vol. 55, 703–708.

Oliver, F. W. J. (1972), *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*, John Wiley & Sons.

Pearson, Neil D. and Tong-Sheng Sun (1994), "Exploiting the Conditional Density in Estimating the Term Structure: An Application to the Cox, Ingersoll, and Ross Model," *Journal of Finance*, vol. 49, 1279–1304.

Press, William H., Saul A Teukolsky, William T. Vetterling, and Brian P. Flanney (1996), *Numerical Recipes in Fortran 77: the Art of Scientific Computing*, Cambridge University Press, Cambridge.

Pritsker, Matt G. (1998), "Nonparametric Density Estimation and Tests of Continuous Time Interest Rate Models," *Review of Financial Studies*, vol. 11, 449–487.

Protter, Philip (1992), *Stochastic Integration and Differential Equations: A New Approach*, Springer-Verlag.

Singleton, Kenneth (1999), "Estimation of Affine Asset Pricing Models Using the Empirical Characteristic Function," *Working Paper*, Graduate School of Business, Stanford University.

Stanton, Richard (1997), "A Nonparametric Model of Term Structure Dynamics and the Market Price of Interest Rate Risk," *The Journal of Finance*, vol. 52, 1973–2002.

Sun, Tong-Sheng (1992), "Real and Nominal Interest Rate: A discrete-Time Model and Its Continuous-Time Limit," *The Review of Financial Studies*, vol. 5, 581–611.

Tauchen, George (1985), "Diagnostic Testing and Evaluation of Maximum Likelihood Models," *Journal of Econometrics*, vol. 30, 415–443.

Tauchen, George (1986), "Statistical properties of generalized methods-of-moments estimators of structural parameters obtained from financial market data," *Journal of Business and Economic Statistics*, vol. 4, 397–416.

Tauchen, George (1996), "New Minimum Chi-Square Methods in Empirical Finance," in "Advances in Econometrics, Seventh World Congress," (edited by Kreps, D. and K. Wallis), Cambridge University Press, Cambridge UK.

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

189

White, Halbert (1994), *Estimation, Inference, and Specification Analysis*, Cambridge University Press, University of California, San Diego.

Wooldridge, Jeffrey M. (1994), *Handbook of Econometrics*, vol. 4, chap. 45, Estimation and Inference for Dependent Processes, pages 2642–2738, Elsevier Science B. V.

Zheng, Xiaodong and Wei-Yin Loh (1995), "Consistent Variable Selection in Linear Models," *Journal of the American Statistical Association*, vol. 90, 1029–1054.

Zhou, Hao (1999a), "Finite Sample Properties of Maximum Likelihood Estimation and Efficient Method of Moments for a Partially Observed Square-Root Process," *Working Paper*, Department of Economics, Duke University.

Zhou, Hao (1999b), "Jump-Diffusion Term Structure and Itô Conditional Moment Generator," *Working Paper*, Department of Economics, Duke University.

Biography

Hao Zhou was born in Shanghai, China, on January 31, 1967. He attended Peking University starting in 1985, earning a B.A. degree in economics in 1989 and an M.A. degree in economics in 1993. Zhou entered the Ph.D. program in economics at Duke University in 1994 and graduated in 2000. His publications include two papers, "Rural-Urban Disparity and Sectoral Labor Allocation in China" (with Dennis Yang) and "Reform the Financial System and Lift the Control of Interest Rates to Facilitate Long Run Economic Growth" (with Justin Lin), as well as a book, titled "Agriculture Research Priorities: A Demand and Supply Analysis of Grain Technology in China" (with Justing Lin and Minggao Shen). Zhou was an honors student in college, received an Excellent Student Award in postgraduate study, and was given a best paper award and a best book award by Peking University. The Graduate School of Duke University and the Department of Economics have awarded him graduate scholarships and fellowships, along with three travel grants for conference presentations.