A Comparison of Univariate Stochastic Volatility Models for U.S. Short Rates Using EMM Estimation*

Ying Gu and Eric Zivot

June 26, 2006

ABSTRACT:

In this paper, the efficient method of moments (EMM) estimation using a semi-nonparametric (SNP) auxiliary model is employed to determine the best fitting model for the volatility dynamics of the U.S. weekly three-month interest rate. A variety of volatility models are considered, including one-factor diffusion models, two-factor and three-factor stochastic volatility (SV) models, non-Gaussian diffusion models with Stable distributed errors, and a variety of Markov regime switching (RS) models. The advantage of using EMM estimation is that all of the proposed structural models can be evaluated with respect to a common auxiliary model. We find that a continuous-time two-factor SV model, a continuous-time three-factor SV model, and a discrete-time RS-in-volatility model with level effect can well explain the salient features of the short rate as summarized by the auxiliary model. We also show that either an SV model with a level effect or a RS model with a level effect, but not both, is needed for explaining the data. Our EMM estimates of the level effect are much lower than unity, but around 1/2 after incorporating the SV effect or the RS effect.

KEYWORDS:

U.S. short rate; stochastic volatility; Markov regime switching; EMM; model selection.

---

* We would like to thank Jiahui Wang, George Tauchen, Hao Zhou, John Nolan, Angelo Melino, Bruce Wang, seminar participants at University of Washington and conference participants at the 22nd Canadian Econometrics Study Group Conference, for helpful comments on this paper. All remaining errors are our own responsibility.

1. Introduction

The risk-free short-term interest rate is a key state variable in asset pricing models, term structure models and macroeconomic models. It is used to express the expected equilibrium returns on risky assets in terms of excess. It directly affects the short end of the term structure and thus has implications for the pricing of the full range of fixed income securities and derivatives. Further, the short rate is an important input for business cycle analysis through its impact on the cost of credit, its sensitivity to the stance of monetary policy, and to inflationary expectations.[1]

Originating from the Brownian motion representation of Merton (1973), an enormous amount of work has been directed towards modeling and estimating the dynamics of the short rate. The mean-reverting model in Vasicek (1977) allows the dynamics of interest rates to be stationary. The square-root model of Cox, Ingersoll, and Ross (1985) (CIR) guarantees positive interest rates and incorporates the "level effect," which allows volatility to increase with the level of the interest rate. Chan et al. (1992) (hereafter CKLS) compared a variety of single factor linear diffusion models for the short rate. They found that models that freely estimated the level effect outperformed other models, and that the level effect parameter estimate was significantly greater than unity. Due to the poor empirical performance of linear diffusion models, several authors have focused on the estimation of the functional form of the drift and volatility of the diffusion model. Conley et al. (1995) exploited the moment generating techniques of Hansen and Scheinkman (1995) to obtain nonparametric estimates of the drift; Ait-Sahalia (1996a) estimated the volatility function nonparametrically; and Stanton (1997) provided nonparametric discrete-time approximations to the drift and volatility functions.

The poor performance of one factor models led to the incorporation of an additional stochastic volatility (SV) factor in order to accommodate the strong conditional

[1] See Andersen (2005).

heteroskedasticity in short rates. Longstaff and Schwartz (1992) derived a two-factor general equilibrium model for the short rate, with its level and its conditional volatility as factors. They showed that a two-factor model improves upon a single factor model, and carries additional information about the term structure and leads to better pricing and hedging performance compared with a single factor model. Similarly, Brenner et al. (1996) and Koedijk et al. (1997) modeled the conditional volatility process of the short rate as a GARCH process and found that a model with both level and GARCH effects outperforms models that exclude one of them. Later, Anderson and Lund (1997) (hereafter AL) and Ball and Torous (1999) found that a two-factor model with level and SV factors outperforms the two-factor model with GARCH volatility.

In the specification of interest rate models, evidence has been documented for regime switching (RS) behavior in short rates. Garcia and Perron (1996) provided a three-regime model using the methodology of Hamilton (1989), which allows the drift and volatility of the ex-post real interest rate to switch over regimes. Their results suggested that both the drift and volatility are essentially different for the periods 1961-1973, 1973-1980 and 1980-1986. Cai (1994) presented a RS-ARCH model for the excess returns of the three-month T-bill over the thirty-day T-bill and reported two periods of high interest rate volatility: one is in 1974 (the energy crisis) and the other is between 1979 and 1982 (the "monetary experiment" of the Federal Reserve). Gray (1996) developed a generalized RS model based on a CIR process with regime dependence in both mean reversion and conditional volatility driven by a GARCH process. He found evidence of a high (low) volatility regime with high (low) mean reversion for one-month U.S. T-Bill yields. An additional high-volatility regime is found in 1987, corresponding to the stock market crash. Gray argued that the RS *and* GARCH effect as well as the diffusion terms are necessary for accommodating the dynamics of the short rates. Smith (2002) presented a model for the short rate based on the CKLS process but only allowed the unconditional volatility to switch between regimes. He empirically compared the RS models and SV models using a quasi-maximum likelihood estimation technique, and argued that either a RS or an SV effect, but not both, is needed to adequately describe the data. Ang and Bekaert (2002a) found that regime-switching models of interest rates replicate non-linear patterns in the drift and volatility functions of short rates found in non-parametric approaches.

As pointed out by many authors, the RS model is more than a mere device used to fit the data; it has important implications for business cycle analysis and yield curve dynamics due to the natural association between the notion of regimes that underlie the econometric model and the large economy-wide shocks that have strong and persistent influences on the behavior of interest rates. For example, Naik and Lee (1998) showed that the RS model generates an empirically more reasonable term structure of volatilities, fat tails, and persistence in volatility compared to those of the SV models. Ang and Bekaert (2001) argued that the two-regime classification of U.S. nominal short term rates corresponds reasonably well with business cycles. Lahiri et al. (2000) studied the comparative performance of a number of interest rate spreads as predictors of the German inflation and business cycle in the post-Bretton Woods era using a two-regime RS model. Bansal and Zhou (2002) developed a term structure model, compared the two-factor RS model with the benchmark CIR model and affine models with up to three factors, and argued that only the RS model can account for the well documented violations of the expectations hypothesis, the observed conditional volatility, and the conditional correlation across yields with regimes intimately related to business cycles. Dai, Singleton, and Yang (2004) developed a term structure model with priced factor and RS risks, provided closed-form solutions for zero-coupon bond prices, and argued that the shapes of the term structures of bond yield volatilities are very different across regimes.

In response to the non-Gaussian behavior of interest rates and asset returns, models have been developed that relax the assumption of conditionally normally distributed innovations to take into account of both volatility clustering and leptokurtosis in describing the financial series.. The GARCH model with Student-t distributed innovations was considered by Bollerslev (1987), and the GARCH model with the extended skewed Student-t distribution was utilized by Lambert and Laurent (2000). Other distributions have been examined, including the normal inverse Gaussian process by Barndorff-Nielsen (1997) and Andersson (2001), the variance-gamma process by

Madan and Seneta (1990), the generalized hyperbolic process of Eberlein, Keller and Prause (1998), and the CGMY process by Carr, Geman, Madan and Yor (2000). In general, estimates of conditional volatility using non-Gaussian distribution showed better results relative to estimates obtained assuming normality. For a review of these results see Peters (2001), and Verhoven and McAleer (2003).

While the SV model and its extensions have theoretical appeal, efficient estimation is not straightforward. Standard statistical methods, both classical and Bayesian, are usually not applicable either because it is not practicable to obtain the likelihood for the entire state vector or because the integration required to eliminate unobservable factors from the likelihood is infeasible. A variety of estimation procedures has been proposed to overcome these difficulties, including the generalized method of moments (GMM) used by Melino and Turnbull (1990), the quasi maximum likelihood (QML) approach followed by Harvey et al. (1994), the simulated maximum likelihood approaches used by Danielsson (1994) and Sandmann and Koopman (1998), the Markov-chain Monte Carlo (MCMC) procedures used by Jaquier et al. (1994) and Kim et al. (1998), and the efficient methods of moments (EMM) approach developed by Gallant and Tauchen (1996) and Gallant and Long (1997). [2]

Although there is a large literature on SV models for interest rates, there still remains substantial disagreement on the empirical performance of different model specifications. The main reason for these disagreements is the use of estimation techniques that make it difficult to compare competing models in a unified way. In this paper we follow the methodology of Gallant, Hsieh, and Tauchen (1997) and use the EMM to estimate and compare a comprehensive collection of univariate SV models for the short-term interest rate including one-factor diffusion models, two-factor and three-factor stochastic volatility (SV) models, non-Gaussian diffusion models with stable distributed errors, and a variety of Markov regime switching (RS) models. The use of EMM allows for a straightforward comparison of models, even if the models are non-nested.

[2] See Andersen et al. (1990) for performance comparisons, and Broto and Ruiz (2002) for a survey on asymptotic properties, finite sample experiments, limitations and advantages of various estimators. Shephard (2005) provides a general overview of the literature.

Our results favor the one-factor non-Gaussian diffusion model over the one-factor Gaussian diffusion model, and the multi-factor SV models and the RS models over the one-factor non-Gaussian diffusion model. We show that a two-factor SV model, a three-factor SV model, and a RS-in-volatility model that allows for a level effect adequately describe the salient features of the short rate process. Our results show that the EMM estimates of the level effect are much lower than unity in the accepted SV models and RS model. Specifically, in our two-factor SV and three-factor SV models, the level effects are estimated similarly to that found in other studies of two-factor models (e.g. AL). In addition, the level effect estimate obtained from our RS-in-volatility model is also found to be around 1/2. Finally, we provide the first EMM estimations for a series of forms of the RS models and offer a performance comparison between different RS models and between the RS models and SV models for fitting the U.S. short rates. Our EMM estimation results clearly indicate that either an SV effect or a RS effect, but not both, are needed for describing the data accurately.

The remainder of the paper is organized as follows. Section 2 provides a description of the EMM methodology, procedure, and diagnostics. Section 3 presents the models for the short rate to be estimated and compared. Section 4 describes the data, and Section 5 reports the EMM estimation results and processes the diagnostic tests. Section 6 summarizes and concludes.

2. Methodology

To facilitate a consistent evaluation and estimation across non-nested models, we rely on the EMM estimation technique developed in Gallant and Tauchen (1996) and extended in Gallant and Long (1997). The basic procedure of EMM estimation, summarized in Figure 1.1, consists of two steps.[3] First, in the projection step, the empirical conditional density of the observed time series is estimated by a semi-nonparametric (SNP) series

[3]See Bansal and Zhou (2002).

expansion. This SNP expansion has a VAR-GARCH Gaussian density as its leading term, and departures from the Gaussian leading term are captured by a Hermite polynomial expansion. Second, in the estimation step, a GMM-type criterion function is constructed using the score functions from the log-likelihood of the SNP density as moments. The scores are evaluated using simulated data from a given structural model, and the criterion function is minimized with respect to the parameters underlying the structural model. A brief description of these steps, following Gallant and Tauchen (2001), is given below.

2.1. Projection Step

Gallant and Tauchen (2001) recommended the SNP model as the score generator for use with the EMM estimation. The advantage of the SNP model is that it can approximate virtually any smooth distribution, even a mixture distribution (as is the case with a model of regime shifts).

To describe the SNP model, let y_t denote the observed data, and let $x_{t-1} = \{y_{t-1}, ..., y_1\}$ denote the lagged observations representing the complete and relevant information set. A SNP model starts with a Gaussian vector autoregression (VAR) with L_u lags, and a GARCH (L_g, L_r) or ARCH (L_r) conditional variance specification. The innovation density is a Hermite density of degree K_z, having the form of a polynomial times the standard normal density.

The SNP conditional density, $f(y_t | x_{t-1}, \theta)$, with parameter vector θ, has the form:

$$f(y_t | x_{t-1}, \theta) \propto [P(z_t)]^2 N(y_t | \mu_x, \Sigma_x) \qquad (1.1)$$

where $z = R_x^{-1}(y - \mu_x)$ with $\Sigma_x = R_x R_x'$. $N(.)$ is a normal density of y with conditional mean μ_x and conditional variance Σ_x, where μ_x is estimated using a VAR specification, and Σ_x is estimated using an ARCH/GARCH specification, which parameterizes R_x.

To accommodate any remaining non-Gaussianity and time series structure in the innovation process, P is a Hermite polynomial with degree K_z in z ; to allow for additional conditional heterogeneity over that allowed by GARCH, the coefficients of the polynomial in the Hermite density are themselves polynomials of degree K_x in L_p lags of the data.[4] For example, if only x_{t-1} is allowed to impact the conditional distribution, the Hermite polynomial P is given by

$$P(z_t, x_{t-1}) = \sum_{i=0}^{K_z} a_i z_t^i$$

$$\text{where } a_i(x_{t-1}) = \sum_{j=0}^{K_x} a_{ij} x_{t-1}^j \qquad (1.2)$$

The order of the polynomial expansion, K_z, controls the extent to which the tails deviate from normality. If $K_z = 0$, the SNP reduces to the normal density. The order of the coefficients of the polynomial, K_x, determines the degree of the heterogeneity of the innovations $\{z_t\}$. When $K_x = 0$, z_t are homogeneous, that is, the conditional density is independent of the lagged observations, x_{t-1}. If $K_x > 0$, we effectively multiply the innovations by functions of x_{t-1}.

Because the number of terms in a polynomial expansion becomes exponentially large as the dimension increases, two additional tuning parameters are introduced: $I_z > 0$ implies that all interactions larger than $K_z - I_z$ are suppressed; similarly for $I_x > 0$. The tuning parameters that describe a SNP model are summarized by the vector $(L_u, L_g, L_r, L_p, K_z, I_z, K_x, I_x)$. Table 1.1 gives a taxonomy of common SNP models.[5]

For a given set of set of tuning parameters, the parameters θ of the SNP model are estimated by quasi-maximum likelihood (QML). The quasi-maximum likelihood estimator, $\tilde{\theta}_n$ satisfies the first-order conditions of the optimization problem,

---

[4] See Gallant and Tauchen (1996).
[5] See Gallant and Tauchen (1997).

$$m_n(\tilde{\theta}) = \frac{1}{n}\sum_{t=1}^{n}\frac{\partial}{\partial\theta}\ln f(y_t \mid x_{t-1}, \tilde{\theta}_n) = \frac{1}{n}\sum_{t=1}^{n} s_f(y_t \mid x_{t-1}, \tilde{\theta}_n) = 0 \qquad (1.3)$$

where $s_f(y_t \mid x_{t-1}, \tilde{\theta}_n) = \frac{\partial}{\partial\theta}\ln f(y_t \mid x_{t-1}, \tilde{\theta}_n)$ denotes the quasi-score function. The dimension of the auxiliary model, l_θ, is selected by following an upward model expansion path, using the Schwarz's Bayesian information criterion (BIC) $BIC = s_n(\tilde{\theta}) + (l_\theta / 2n)\ln(n)$, where $s_n(\tilde{\theta}) = -L_n(\tilde{\theta}, \{y_t\}_{t=1}^{n})$ is the negative maximized objective function. Implied by standard QML theory, even if the auxiliary model is misspecified, under suitable regularity, $\tilde{\theta}_n \xrightarrow{p} \theta_0$, where the limiting value, θ_0, is denoted the quasi-true value of θ.

The projection step provides a summary of the data, which will be used as the score generator for the next step of estimation. Gallant and Long (1997) show that a judicious selection of the auxiliary model, ensuring that it approximates the salient features of the observed data, will result in full asymptotic efficiency. Effectively, as the score generator approaches the true conditional density, the estimated covariance matrix for the structural parameter approaches that of maximum likelihood. This result embodies one of the main advantages of EMM. It prescribes a systematic approach to the derivation of efficient moment conditions for estimation in a general parametric setting.

2.2. Estimation Step

In the estimation step estimates of the parameters of a candidate structural parameter are obtained from a GMM-type estimation procedure using the fitted scores from the SNP model as the moment conditions. To do this, for a specific structural model represented by $P(y_t \mid x_{t-1,}\rho)$ with a given parameter vector ρ, a simulated series $\{\hat{y}_t\}_{t=1}^{N}$ is generated. Identification requires that the dimension of the quasi-score (the length of θ), l_θ, exceeds that of the structural parameter vector, l_ρ. An average over a long simulation from the true structural model, reevaluated at the fixed QML estimate,

$$m_N(\rho,\tilde{\theta}_n)=\frac{1}{N}\sum_{n=1}^{N}\frac{\partial}{\partial\theta}\ln f(\hat{y}_t(\rho)\,|\,\hat{x}_{t-1}(\rho),\tilde{\theta}_n) \tag{1.4}$$

would satisfy $m(\rho_0,\tilde{\theta}_n)=0$. In the usual case in which $l_\theta > l_\rho$, the structural parameters ρ are estimated by minimizing the EMM objective function

$$\hat{\rho}_n=\arg\min_{\rho}\left[m_N(\rho,\tilde{\theta}_n)'\tilde{I}_n^{-1}m_N(\rho,\tilde{\theta}_n)\right] \tag{1.5}$$

where $\tilde{I}_n$ denotes a consistent estimator of the asymptotic covariance matrix sample quasi-score vector. The estimate $\tilde{I}_n$ is obtained directly from the first step which avoids the need for computation of the weighting matrix during the second GMM-based estimation step. In addition, if the auxiliary model is expanded to the point where it accommodates all main systematic features of the data, likelihood theory implies that the quasi-scores constitute a (near) martingale difference sequence, and a convenient estimator of the quasi-information matrix is obtained from the outer product of the score:

$$\tilde{I}_n=\frac{1}{n}\sum_{t=1}^{n}\left[\frac{\partial}{\partial\theta}\ln f(y_t\,|\,x_{t-1}(\rho),\tilde{\theta}_n)\right]\left[\frac{\partial}{\partial\theta}\ln f(y_t\,|\,x_{t-1}(\rho),\tilde{\theta}_n)\right]' \tag{1.6}$$

Gallant and Tauchen (1996) show that, under suitable regularity conditions, the EMM estimator $\hat{\rho}_n$ is almost surely consistent and asymptotically normal. Moreover, the asymptotic variance-covariance matrix may be estimated consistently by

$$\mathrm{cov}(\hat{\rho}_n)=\frac{1}{n}\left[\frac{\partial m_N(\hat{\rho}_n,\tilde{\theta}_n)'}{\partial\rho}\tilde{I}_n^{-1}\frac{\partial m_N(\hat{\rho}_n,\tilde{\theta}_n)'}{\partial\rho'}\right]^{-1} \tag{1.7}$$

The usual GMM test of over-identifying restrictions may be used to test model adequacy. If the structural model is correctly specified, then the normalized EMM objective function satisfies

$$nm_N(\hat{\rho}_n,\tilde{\theta}_n)'\tilde{I}_n^{-1}m_N(\hat{\rho}_n,\tilde{\theta}_n)\sim\chi^2(l_\theta-l_\rho) \tag{1.8}$$

If the overidentification test rejects an underlying structural model, the individual elements of the score vector may provide useful information regarding the dimensions in which the structural model fails to accommodate the data. These model diagnostics are based on the standard *t*-statistics of the individual elements of the score vector,

$m_N(\hat{\rho}_n, \tilde{\theta}_n)$. Obtained by normalizing the score vector by its standard error, these *t*-statistics can be interpreted much as normalized regression residuals. Thus, large *t*-ratios reveal those characteristics that are not well approximated. Subject to the same risk as the interpretation of regression residual, the *t*-ratios are usually biased downward, and therefore conservative. Nonetheless, as with regression residuals, inspecting normalized elements of $m(\hat{\rho}_n, \tilde{\theta}_n)$ is usually the most informative diagnostic available.

Another advantage of using EMM estimation is the ability to rank non-nested structural models. Notice that the weight matrix in GMM used in constructing the specification test is identical across different model specifications. Consequently, the *p*-value based on the overidentification test can be directly compared across different structural models to identify the best structural model. It is well recognized in the literature that tests for the presence of regime shifts against an alternative require nonstandard approaches. Our approach of comparing all the considered models to a common nonparametric density allows us to rank order all the considered models according to the *p*-values implied by the EMM criterion function.

3. SV Models for the Short Rate

In this section we discuss a series of models and extensions to explain short-term interest rate dynamics. The first type of model is the generalized Gaussian diffusion model that is commonly used in building term structure models. To incorporate additional factors, we extend the one-factor diffusion model to the two-factor and three-factor SV models that has been proven to be more successful than the ARCH/GARCH model in modeling the dynamics of the second moment of many financial time series. The second type of model is the non-Gaussian diffusion model with Stable distributed innovations, which has recently become popular in the empirical finance literature. The third type of model allows for Markov RS behavior in the specification of the volatility dynamics, with the flexibility of simultaneously mixing the RS effect with the SV effect and the level effect of volatility. The Gaussian and non-Gaussian diffusion models are continuous-time models and the RS models are discrete-time models.

3.1. Gaussian Diffusion Models

A. One-Factor Gaussian Diffusion Model

Firstly, we consider the generalized diffusion model, presented by Chan et al. (1992), in which the instantaneous change in the short rate can been characterized as a stochastic differential equation (SDE) given by

$$dr_t = (\phi_0 - \phi_1 r_t)dt + \sigma r_t^{\gamma} dW_1 = k_r(\mu_r - r_t)dt + \sigma r_t^{\gamma} dW_1 \tag{1.9}$$

where $\{r_t\}$ is the short rate at time *t*, and dW_1 is a standard Wiener process. We call (1.9) the CKLS model. The key characteristic of the dynamics is that the conditional mean and variance of changes in the short rate depend on the level of the rate. Specifically, in this model, r_t mean-reverts towards the long-run level μ_r, with the speed of the reversion measured by k_r, and γ captures the so-called "level effect" in which of the level of rates influences the conditional volatility. By allowing γ to be estimated freely, many well known models can be nested with appropriated parameter restrictions within this generalized model.[6]

To empirically calibrate the general SDE (1.9), Chan et al (1992) employed the following discretization approximation

$$\Delta r_t = \phi_0 - \phi_1 r_t + \sigma r_t^{\gamma} z_{t+1} = k_r(\mu_r - r_t) + \sigma r_t^{\gamma} z_{t-1} \tag{1.10}$$

and estimated the model parameters with the generalized methods of moments (GMM) estimation technique of Hansen (1982). Using monthly data from 1964-1989, they found that the short rate was mean reverting, and reported a point estimate of 1.4999 for the level effect parameter γ which implies the volatility of short-term interest rates is explosive. With similar data, Smith (2002) estimated the CKLS model using a quasi-maximum likelihood methodology and reported a similar level effect estimate of 1.4515.

[6] See Chan et al (1992).

In order to obtain the maximum likelihood estimates and guarantee a compatible comparison with the SV and RS models, Smith used a two step procedure for estimating the models. In the first step, he used the ordinary least square (OLS) to obtain an estimate of the mean reversion parameters ϕ_0 and ϕ_1. In the second step, he formed the fitted residuals $\Delta r_t = \phi_0 - \phi_1 r_{t-1} + e_t$, and then estimated the remaining parameters from the transformation of the log of the squared residual. This estimation procedure is required to build up the likelihood functions for the SV and RS models.

Although these findings are instructive for understanding the short-term rate dynamics, they are not entirely satisfactory. First, Monte Carlo studies have questioned the efficiency of using GMM estimation in sense of the choice of the moment conditions and its finite sample performance. The two-step estimation procedure used in Smith (2002) suffers from the loss of the estimation efficiency as well. Lastly, evidence has been shown that the estimated parameters of the CKLS model are sensitive to the data frequency. In particular, the level effect parameter estimate from monthly data could be spuriously high and unstable; using more frequently sampled data leads to different results. In addition, as pointed out by Andersen and Lund (1997), the internal dynamics proposed in the discrete-time models, at estimated parameter values, are excessively erratic. This severely limits their usefulness for numerical or simulation-based estimation procedures. To avoid the previously mentioned difficulties in estimating models of the short rate, in this paper we rely on the EMM estimation using weekly data and estimate the continuous-time CKLS model directly rather than using a discretization approximation.

B. Two-Factor SV Model

We consider the following CKLS model extended to have stochastic volatility in the spirit of Taylor (1986, 1994):

$$\begin{cases} dr_t = (\phi_0 - \phi_1 r_t)dt + r_t^{\gamma}\sigma_t dz_t = k_r(\mu_r - r_t)dt + r_t^{\gamma}\sigma_t dW_1 \\ d\log(\sigma_t^2) = (\omega_0 + \omega_1 \log(\sigma_t^2))dt + \xi dW_2 \end{cases} \tag{1.11}$$

where dW_1 and dW_2 are mutually independent i.i.d. Wiener processes. For these dynamics, the log-volatility of short rate series is assumed to follow a mean reverting process as well as the series itself. Also, the conditional volatility is subject to random shocks, and the sensitivity to these shocks is measured by the parameter $\xi > 0$.

Maximum likelihood estimation is generally not feasible for estimating the SV models due to the presence of an unobserved volatility. One procedure available is the quasi-maximum likelihood procedure of Harvey, Ruiz, and Shephard (1994). This approach uses a transformation on the log of the squared residual in order to write the system in state-space form, and then applies the Kalman filter to recursively build up the likelihood function. Smith (2002) followed this two-step estimation procedure and reported an estimate of 1.44 for the level effect parameter using monthly data.

Andersen and Lund (hereafter AL) (1997) estimated (1.11) directly using the EMM estimation technique with a SNP auxiliary model that employs a Level-EGARCH leading term. Using weekly data over the 1954-1995 sample periods, they found the level effect parameter to be close to 0.5. While their model was rejected by the data at the 5 percent significant level, the incorporation of the unobservable volatility factor was shown to greatly enhance the model's ability to fit the data and the implied process was much less erratic than the process implied by the CKLS estimates[7]. Following AL, we estimate (1.11) using the EMM estimation with a longer span of weekly data.

C. Three-Factor SV Model

[7] A number of other estimation procedures have been implemented for the two-factor SV model, including the Bayesian technique of Jacquier, Polson, and Rossi (1994), the maximum likelihood procedure of Fridman and Harris (1998), and the maximum likelihood Monte Carlo method of Sndmann and Koopman (1998).

We consider the following continuous-time three-factor SV model for the short rate:

$$\begin{cases} dr_t = (\phi_0 - \phi_1 r_t)dt + \sigma_t r_t^{\gamma} dz_t = k_r(\mu_{r,t} - r_t)dt + \sigma_t r_t^{\gamma} dW_1 \\ d\log(\sigma_t^2) = (\omega_0 + \omega_1 \log(\sigma_t^2))dt + \xi dW_2 \\ d\mu_t = (\upsilon_0 + \upsilon_1 \mu_t)dt + \zeta dW_3 \end{cases} \qquad (1.12)$$

where dW_1, dW_2 and dW_3 are mutually independent i.i.d. Wiener processes. In (1.12), the log-volatility of short rate series and the long-run mean are assumed to follow mean-reverting process. The sensitivity of shocks to the log-volatility and to the long-run mean are measured by the non-negative parameters ξ and ζ, respectively.

The model (1.12) is an extension of the two-factor SV model (1.11) suggested by the AL. The introduction of a third factor associated with the reverting mean level may improve the data fitting through accommodating the time-varying drift behavior over the sample period. According to AL, time variation in the reverting mean could be interpreted as variation in an underlying inflation rate.

The three-factor model (1.12) is a particular form of a general class of affine multifactor models. Dai and Singleton (2002) discussed the general issues for the identification and admissibility conditions of affine diffusion models, which are characterized by linearity of the drift and variance functions. The investigation of other types of three-factor SV models or general affine diffusion models is left for further research.

3.2. Non-Gaussian Diffusion Model

The modern asset pricing theory and, more specific, the option pricing theory have been firmly built upon the Gaussian diffusion framework based on the beliefs that the financial data tends to become more Gaussian over longer timescales. The popularity of the SV approach is partially due to its consistency with the Gaussian assumption making possible an appropriate generalization of the Black-Scholes option pricing framework. However, empirically studies have shown that financial returns exhibit features that are incompatible with the assumption of Gaussian data. The leptokurtosis implied by the Gaussian diffusion and SV models tend to be far less than the sample kurtosis observed from many financial series, although the implied time-varying and persistent volatilities are consistent with the data.

One generalization developed to explain the observed leptokurtosis and skewness is the jump-diffusion model originally proposed by Merton (1976). This model consists of two parts: a continuous part modeled by a geometric Brownian motion, and a jump part with the logarithm of the jump sizes having a double exponential distribution and the jump times corresponding to the event times of a Poisson process. General properties of jump-diffusion models with independent identically distributed jump sizes have been extensively studied; for an excellent survey, see Duffie (2000).

Another generalization is to consider diffusion models assuming non-Gaussian distributions to capture the departures from the Gaussian diffusion model. Following this direction, a variety of non-Gaussian distributions has been considered in discrete-time models.[8] However, these models suffer from the lack of "stability"; i.e., the distribution of the increments do not depend on the time intervals, which is a desirable property for asset returns particularly in the context of portfolio analysis and risk management as stressed by Mandelbrot (1963). In fact, the stable law[9] is the only possible weak limit of properly normalized sums of i.i.d. random variables and only for stable distributed returns do we have the property that linear combinations of different return series follow again a stable distribution.

Motivated by the nice properties of the stable law and stability under-addition, we consider the following continuous-time non-Gaussian CKLS model with stable Lèvy increments

[8] See a review of, among others, Peters (2001) and Verhoven and McAleer (2003).
[9] The sum of a number of random variables with power-law tail distributions having infinite variance will tend to a stable Lèvy distribution as the number of variables grows, also referred to as the generalized central theorem.

$$dr_t = (\phi_0 - \phi_1 r_t)dt + \sigma r_t^{\gamma} dW_1 = k_r(\mu_r - r_t)dt + \sigma r_t^{\gamma} dL_t \quad (1.13)$$

The key characteristics of this model are essentially the same as those of Gaussian CKLS model (mean-reverting process for the drift dynamics and the incorporated level effect for the variance dynamics of the short rate), except that L_t is a stable Lèvy process.

Stable Lèvy processes are stochastic processes with independent and stationary increments. A stochastic process L_t is a stable Lèvy process if and only if: (1) it has independent increments; that is, for $0 < a < b < c < d$, $L_d - L_c$ and $L_b - L_a$ are independent; (2) it has stationary increments; that is, the distribution of $L_{t+s} - L_t$ does not depend on t; (3) it is stochastically continuous (4) with probability one it has right-continuous paths with finite left-limits; and (5) $L_0 = 0$ almost surely. The Brownian motion is a special example of Lèvy processes, one which is with stationary, independent increments having a Gaussian distribution; here we consider the standard stable process, which is a Lèvy process with stationary, independent increments having a standard stable distribution, $\mathrm{S}(\alpha, \beta, 0, 1)$. A stable distribution is characterized by four parameters: $(\alpha, \beta, c, \delta)$. The exponent, α, confined to the interval $0 < \alpha \le 2$, is known as the shape variable, which influences the total probability contained in the extreme tails, or the shape of the distribution. The smaller the value of α, the thicker the tails of the distribution. In particular when $\alpha = 2$ we get the normal distribution. The parameter β measures asymmetry of the distribution. If $\beta = 0$, the distribution is symmetric about the location parameter δ; if $\beta = 1$, the distribution is totally skewed to the right and similarly it is totally to the left when $\beta = -1$. The scale parameter c narrows or broadens the distribution about δ in proportion to c. A standard stable distribution has $\delta = 0$ and $c = 1$. Note a stable distribution given by $(2, 0, c, \delta)$ is exactly a normal distribution with mean δ and variance $2c^2$.

There is not much published literature on empirical volatility modeling using stable distributions. The non-existence of moments of second or higher order is a major drawback of the use of the stable distribution from an empirical point of view. Also, with the exception of a few cases, the probability density function is not known in closed form; therefore, one has to use their characteristic functions instead. On the other hand, one can use stable distributions to save the CLT argument, based on which a similar asset pricing framework to the current Gaussian one could be established; it also can easily accommodate heavy tails and skewness of financial series, which is a much desired property in empirical finance. For these reasons, the use of stable processes has recently become substantially more popular in the modeling of stochastic volatility (Liu and Brosen (1995)), portfolio theory (Olotarev (1986), Mittnik and Rachev (1991), Cheng and Rachev (1995)), asset pricing theory (Connor (1984), Gamrowski and Rachev (1994,1995)), option pricing (Rachev and Samorodnitsky (1993), Janicki and Weron (1994), Bouleau and Lepingle (1994), Matacz (2004)), and other financial phenomena.[10]

In our estimation of (1.13) using EMM, we fix the characteristic parameters α and β of the stable distribution and freely estimate the remaining parameters. The choices for α and β are *ad hoc* and it would be desirable to estimate these parameters freely.[11]

3.3. RS Models

The diffusion models discussed in the previous subsections are single-regime models in that they have a single structure for the conditional mean and variance. For example, the CKLS model for the short rate is assumed to be mean reverting to the same long-run mean, with the same speed of reversion and the same level effect throughout the sample. A more flexible extension is to relax the assumption of a single regime in favor of a two-state Markov RS specification. Many authors have proposed RS models for fitting the dynamics of the short-term interest rate (see, Hamilton (1998), Garcia and Perron (1996), Gray (1996) and Ang and Bakeart (2001), Liechty and Roberts (2001)), for the impact on the entire yield curve using dynamic term structure models (see, Naik and Lee (1997), Boudoukh et al. (1999), Evans (2001) and Bansal and Zhou (2003), Dai, Singleton and

[10] See Marinelli and Rachev (2002).

[11] Garcia, Renault and Veredas (2004) discussed the estimation of the parameters of a Stable distribution using the indirect inference methods relative to other prevalent methods based on the characteristic function and the empirical quantiles of the Stable distribution.

Yang (2004)), and for the bond pricing in the RS context (see, Landén (2000) and Wu and Zeng (2003)). While many theoretical and empirical works show strong evidence for regime switching in interest rates, the specification issue of the RS model for the conditional mean and variance dynamics of the interest rates has not been extensively explored in the literature. Considering that our interest in this paper is to model the volatility dynamics for the short rate, we assume a simple specification in which the conditional mean parameters are regime independent. Furthermore, in our specification of RS models we use the discrete-time approximation to the continuous-time diffusion used in CKLS (1992), which is consistent with the rationale that large regime switching behavior only occurs infrequently over time. EMM estimation enables us to compare the RS models with the continuous-time models based on the EMM objective function *p*-value.

Given the assumption of the single-regime conditional mean dynamics, we propose four RS models to describe the volatility dynamics. The first model is a simplified regime switching-in-volatility model (RS-in-σ model hereafter) based on a discretized OU process, given by

$$\Delta r_t = \phi_0 - \phi_1 r_{t-1} + \sigma_i z_t \qquad i = 1,2 \qquad (1.14)$$

This model assumes the same speed of mean reversion ϕ_1 to a common long-run mean (ϕ_0/ϕ_1), but allows different shocks within each regime to accommodate time-varying volatility. The switching states are governed by a first-order Markov process. The time invariant transition probabilities from regime j to regime i are defined as $p_{ij} = \Pr(S_t = j \mid S_{t-1} = i)$ with the restriction $\sum_{i=1}^{2} p_{ij} = 1$. For the case of two states, the matrix of transition probabilities is given by [12]

$$P = \begin{pmatrix} P_1 & 1-P_1 \\ 1-P_2 & P_2 \end{pmatrix} \qquad (1.15)$$

[12] For future research, the RS models can be generalized to have a greater number of states or the regime switching probabilities can be made a function of the level of interest rates. The latter case allows for the possibility that a switch to the high-volatility regime may be more likely when interest rates are high according to Gray (1996).

Due to the success of the two-factor SV model over the one-factor diffusion model, the second RS model we consider is an extension of the $RS-in-\sigma$ model where the conditional variance is driven by a SV process. This model ($RS-in-\sigma+SV$ model hereafter) is given by

$$\begin{cases} \Delta r_t = \phi_0 - \phi_1 r_{t-1} + \sigma_t z_t \\ \Delta \log(\sigma_t^2) = \omega_{0,i} + \omega_1 \log(\sigma_{t-1}^2) + \xi u_t \end{cases} \quad \text{with } i = 1,2 \quad (1.16)$$

The conditional variance of (3.3.3) has a regime independent random shock but regime-dependent reverting mean. Thus, the $RS-in-\sigma+SV$ model nests the simple OU process, OU-SV process, and $RS-in-\sigma$ model as special cases.

Different from the above two RS models built on the OU process, the following two RS models are based on the generalized CKLS process. Incorporating both the RS-in-volatility effect and the level effect, the third RS model is called the $RS-in-\sigma+Level$ model and is given by

$$\Delta r_t = \phi_0 - \phi_1 r_{t-1} + \sigma_i r_{t-1}{}^{\gamma} z_t \quad \text{with } i = 1,2 \quad (1.17)$$

The $RS-in-\sigma+Level$ model incorporates the sensitivity of volatility to the current level of short rate, measured by γ, to accommodate additional time-varying behavior and conditional heteroskedasticity, although the level effect parameter is kept the same across the regimes.

The fourth RS model is an extension of the $RS-in-\sigma+Level$ model, which we call the $RS-in-\sigma+Level+SV$ model, is given by

$$\begin{cases} \Delta r_t = \phi_0 - \phi_1 r_{t-1} + \sigma_t r_{t-1}{}^{\gamma} z_t \\ \Delta \log(\sigma_t^2) = \omega_{0,i} + \omega_1 \log(\sigma_{t-1}^2) + \xi u_t \end{cases} \quad \text{with } i = 1,2 \quad (1.18)$$

In addition to the characteristics of the $RS-in-\sigma+Level$ model, the conditional log-volatility process is driven by a SV process, with regime-dependent mean reversion $(\omega_{0,i})$ but regime-independent random shocks. The $RS-in-\sigma+Level+SV$ model nests as special cases discrete-time versions of the CKLS model, the two-factor SV model, and the three other RS models.

The first two RS models based on a simple OU process are motivated by the work of Gray (1996). He used a generalized RS framework where all conditional mean parameters (ϕ_0 and ϕ_1) and conditional variance parameters (σ) are allowed to switch across the two regimes. He considered a different extension of the $RS-in-\sigma$ model where the conditional variance is driven by a GARCH process rather than an SV process. Using weekly data on the 30-day T-bill rate, he argued that both the RS effect and the GARCH effect are important to adequately fit the data. He also constructed a likelihood ratio test to compare his $RS-in-\sigma$ model with his $RS-in-\sigma+GARCH$ model.

The last two RS models, based on the CKLS model, are motivated by Smith (2002). He employed a two-step procedure in order to overcome the difficulty of estimating the RS model using the quasi-maximum likelihood approach of Harvey, Ruiz, and Shephard (1994). Smith showed that the level effect parameter is spuriously high in the single-regime models, and is reduced to around unity in his RS models. He also argued that either the SV effect or the RS effect, but not both, are needed for describing the data accurately. We note that So, Lam, and Li (1998) developed a similar model as our $RS-in-\sigma+Level+SV$ model and estimated it using the Bayesian technique of Jacquier, Polson, and Rossi (1994).

For RS models, EMM estimation has advantages over the QML Kalman filter procedure and other estimation techniques. With EMM, we can estimate all the unknown parameters simultaneously to ensure that no important information has been lost in the process, which cannot be guaranteed by the two-step procedure of Smith (2002). Another problem that may relate to the efficiency loss of Smith (2002) is that the simulated conditional volatility process based on his parameter estimates (especially the positive volatility persistency parameter) is a highly explosive process. In addition, the usual test statistics cannot be applied to test the existence of the second regime since parameters associated with the second state are unidentified under the null of one regime. Most of the past works obtained the evidence for the existence of the additional regime from the enormous increase in the likelihood value when moving from a single-regime model to a two-regime model or carefully applied the LRT to compare the regime-switching models. With EMM, all the comparable models could be easily ranked according to the simple measurement of the p-values implied by the EMM criterion function. After the one-to-one model comparison, we expect to have a systematic answer for questions such as (1) whether the simple $RS-in-\sigma$ model could mimic the performance of complicated non-Gaussian diffusion models, (2) whether $RS-in-\sigma+Level$ model could save the efforts of adding one stochastic factor as implied by SV models, and (3) which effect or effects among the three, the level effect, the SV effect, and the RS effect, are needed to adequately fit the data of US short rates.

4. Data

Our empirical work uses weekly (Wednesday) observations of the annualized yield on the 3-month U.S. T-bill over the period January 1954 to September 2004, forming 2648 observations. The data was constructed from a daily series available from the Federal Reserve Bank, where the rates are calculated as unweighted averages of closing bid rates quoted by at lease five dealers in the secondary market, and the rates are posted on a bank discount basis, but converted into continuously compounded yields prior to analysis. We analyze weekly rates over daily rates to avoid missing data, possible holiday and weekday effects, and other potential problems associated with market microstructure effects. Wednesday data are used because of the least number of missing observations for this weekday. When a Wednesday rate is missing, we use the Tuesday rate; when a Tuesday rate is missing, use the Thursday rate. The data preparation procedure follows Andersen and Long (1997).

The raw data plotted in Figure 1.2, and descriptive statistics are given in Table 1.2. The basic stylized facts concerning the short-rate are: near nonstationary behavior (slow mean reversion), large changes and small changes are clustered together (ARCH effect), the volatility of rates increases with the level of rates (level effect), and positive skewness and excess kurtosis[13] (non Gaussian distribution). The non Gaussian behavior of the short rate is clearly shown in the qq-plot in Figure 1.4 and in the statistics summary in Table 1.2, and the slow mean reversion and ARCH effect are illustrated in the autocorrelation plots in Figure 1.3.

The data period of our sample, 1954 to 2004, represents the longest weekly set of observations on the 3-month T-bill rate, which is important for evaluating models that purport to explain mean and volatility dynamics. Also, our sample contains seven major recessions and six major expansions, which provides economic motivation for incorporating regime shifts into the models. Some important events that may cause strong shifts in the behavior of interest rates dynamics include: the Vietnam War from 1961 to 1975, the simultaneous occurrence of recession and inflation in the early 1970s, the 1973 energy crisis due to the onset of an oil embargo by OPEC until 1975, the "Monetary experiment" conducted by the Federal Reserve during 1979-82 when its policy shifted away from targeting federal fund rate, the largest stock market crash on October 19, 1987, the Gulf War which started in August 1990, and the longest peacetime economic expansion in U.S. history beginning in March 1991.[14] The period from 1996 to 2004, which was not covered by many previous analyses of the short rate, poses an especially tough challenge for standard asset pricing models. This period started with an unprecedented period of long economic growth and a bull stock market run, which was interrupted by the September 11, 2001 terrorist attack, and was followed by a downturn of the stock market, and finally ended with the "War on Terrorism" campaign with the invasion of Iraq on March 2003.[15]

5. Empirical Results:

5.1. Estimation of the SNP Auxiliary Model

[13] Kurtosis of the Gaussian distribution is three; excess kurtosis for a non-Gaussian distribution is the different between its kurtosis and three.
[14] See Choi (2004).
[15] See Bansal, Tauchen and Zhou (2003).

The first step in EMM estimation is to project the observed data onto an auxiliary model that captures all of the relevant characteristics of the data. We use the semi-nonparametric (SNP) conditional density model described in Gallant and Tauchen (2001) as our auxiliary model. The selection of an appropriate auxiliary model is essential for the success of EMM estimation, especially for interest rate data as stressed by Andersen and Lund (1997) and Gallant and Tauchen (2004). The empirical literature on EMM estimation of the short-rate, however, has not explored the relevance of this issue in a systematic manner.[16]

We follow Gallant and Tauchen (2001) and use a specific-to-general model selection procedure based on minimizing a Bayesian information criterion (BIC). In particular, the SNP tuning parameters $(L_u, L_g, L_r, L_p, K_z, I_z = 0, K_x, I_x = 0)$ are selected by moving upward along a model expansion path where small values of BIC are preferred. The expansion paths we follow are illustrated in Table 1.3. First, the autoregressive order L_u is determined. The expansion path with ARCH leading terms is to expand L_r, then to expand K_z, and finally expand K_x. For GARCH leading terms, the strategy is to put $L_r = L_g = 1$ first, then expand K_z and K_x. The expansion paths we follow are not exhaustive across models and it sometimes happens that the best set of the tuning parameters lies elsewhere within the expansion path. Therefore, we also explore some other paths which slightly deviate from the ones specified in Table 1.3.

The best fitting SNP models for the 3-month T-bill rate in terms of BIC, characterized by the set of tuning parameters, $(L_u, L_g, L_r, L_p, K_z, I_z = 0, K_x, I_x = 0)$, are reported in Table 1.4. Following the upward BIC protocol and exploring beyond the expansion path a bit, the preferred auxiliary model is the SNP 11117000 model. The SNP 11117000 model is a GARCH (1,1) model with a nonparametric error density represented as a seven-degree Hermite polynomial expansion of the normal density where the Hermite coefficients are state independent. The model is similar to the semi-parametric GARCH of Engle and

[16] See Brandt and Chapman (2002).

Gonzalez-Rivera (1991). Table 1.5 gives the parameter estimates. The estimated AR coefficient is 0.999 which implies a very slow mean reversion and near nonstationary behavior. The sum of the ARCH and GARCH terms implies highly persistent conditional volatility.[17] The large positive coefficient on the 4^{th} order Hermite term and the positive coefficient on the seventh order Hermite term capture the fat tails and positive skewness in the demeaned short-rate series. Our preferred SNP model for the short rate is similar to the SNP models used by other authors as shown in Table 1.6.

As stressed by Gallant and Tauchen (2001), if the fitted SNP model is to be used as the score generator in conjunction with EMM it is important to check the dynamic stability of the model. For complicated SNP models, a simple way to check dynamic stability is to generate long simulations from the fitted model and observe if these simulations become explosive. For non-explosive models, the simulations should capture all the salient feathers of the observed data. The simulated series based on the fitted SNP models are shown in Figure 1.5. From the plots, it can be observed that the 11117000 SNP model mimics the observed data fairly well, although it produces simulations with negative interest rates. The simulation from the 11118000 SNP model is also plotted in Figure 1.5, and it appears mildly explosive.

Residual diagnostic checks on the fitted model are conducted to verify that it is adequate and appropriate. Panel (A) in Figure 1.6 gives the estimated conditional volatilities from the 11117000 model, and these capture the observed volatility patterns in the observed data. Panel (B) shows the estimated conditional density, which is more peaked in the center with heavy tails relative to the Gaussian distribution. The qq-plots for the simulated series from the fitted SNP 11117000 model and its first order change are shown in Panel (C) and Panel (D), both of which capture the patterns of the real data series. The standardized residuals, shown in Panel (E), mostly resemble a white noise process. However, there are some large outliers present. The autocorrelation plots of the residuals and squared residuals in Panel (F) reveal no significant autocorrelation and

[17] Because of the absolute value formulation in the GARCH specification, the sum of the ARCH and GARCH coefficients do not have to be less than one for the model to be stationary.

indicate that the fitted SNP model adequately captures the conditional dynamics in the mean and volatility.

5.2. EMM Estimation Results

In this subsection we report the EMM estimation results for a number of structural models for interest rates described in Section 3.2. The single regime structural models we consider are: the one-factor CKLS model with Gaussian errors (CKLS-N), the two and three-factor SV model (SV2, SV3), the non-Gaussian stable diffusion model with shape variable α and skewness variable β (CKLS-S(α, β)). The Markov regime switching (RS) models we consider are: the RS-in-σ and RS-in-$\sigma + SV$ model based on a simple OU process, and the RS-in-$\sigma + Level$ and RS-in-$\sigma + Level + SV$ model based on the generalized CKLS model.

The EMM estimation procedure requires the simulation of a long sample from the underlying structural models. For both the discrete-time and continuous-time models, the EMM objective function is formed using a default simulation size of 75,000, where we have discarded the first 5,000 observations. Restarts of the optimizer at random perturbation of the initial value values are employed for EMM to avoid local optima. For continuous-time diffusion models, the simulations are generated by the Euler scheme using 25 subintervals per week.[18] Tables 1.7-1.9 contain the results for short rate from estimating each model outlined above.[19] In the following sections, we present in detail the one-by-one model performance and comparison.

5.2.1. Gaussian Diffusion Models

Table 1.7a reports the EMM estimation results for the Gaussian diffusion models: the one-factor Gaussian CKLS model, the two-factor SV model, and the three-factor SV

[18] Further details regarding the implementation refer to the appendix in AL (1997).
[19] The estimation is conducted using the S-PLUS implementation of Gallant and Tauchen's EMM FORTRAN code available in S+FinMetrics 2.0 and described in Zivot and Wang (2006).

model. The small p-value based on the χ^2 distribution associated with the EMM objective function values, leads to a strong rejection of the one-factor Gaussian CKLS model. On the other hand, the two-factor SV model and the three-factor SV model are not rejected at the 10% level; the former is in contrast to what Andersen and Lund (AL) (1997) found. Our results indicate that the introduction of an additional stochastic volatility factor is important for explaining observed interest rate behavior.

Our estimation results suggest the following insights about the dynamics of the short rate. The signs of all the parameter estimates for the mean dynamics are consistent with the GMM estimates of the one factor CKLS model reported in Chan et al (1992) based on monthly data. All of the models indicate that short rates revert ($-\phi_1 < 0$) to a positive long-run mean ($\phi_0 / \phi_1 > 0$), with a very slow rate of mean reversion.[20] Based on our estimates of the two-factor SV model, the implied estimated measure of log-volatility persistence, $\exp(-\omega_1 / 52)$, is about 0.9893 at the weekly level, and the discrete-time autoregressive coefficient in the mean dynamics, $\exp(-\phi_1 / 52)$, is about 0.994. These estimates are comparable with those reported in AL. Moreover, we find that the conditional volatility of rates is sensitive to the level of the rates; that is, the elasticity of volatility measured by γ is significantly in excess of zero.

The incorporation of data after 1989 in the estimation changes the implied dynamics of the short rate substantially in many aspects from previous estimates. For example, our two-factor SV model implies a lower long-run mean, measured by ϕ_0 / ϕ_1, of 2.85% and a faster speed of mean reversion than the results found by AL using a similar model. This difference may be partially explained by the fact that the Federal Reserve started to decrease the Federal Funds rate dramatically after 2001 in order to boost the economy after the "9/11" recession. More striking is the change in the estimate of the level effect, measured by γ. The level effect estimate in our one-factor CKLS model is 0.3 which is

[20] Notice that the reason that the long-run reverting mean for the short rates implied by the estimation of our one-factor CKLS models differs substantially from the GMM estimation is due to the fact that we use percentage interest rates for the analysis, rather than decimal interest rates as Chan et al (1992).

substantially lower than the GMM estimate, $\gamma = 1.49$, reported in CKLS (1992). AL showed that the level effect is weakened if a second volatility factor is incorporated. Our results show that the evidence for a strong level effect is significantly weakened without an additional SV factor. The estimate of the level effect in our two-factor SV model is 0.67 which is a bit larger than what AL found in their SV2 model.

Our SV3 model involves the introduction of a third factor associated with the mean level as suggested by AL who suspected that a time-varying long-run reverting mean as well as a time-varying conditional volatility is needed to accommodate the data.[21] We are not surprised to see that the three-factor SV model is favored over the two-factor SV model and the result improves significantly by adding this mean related factor. Implied from the SV3 model, the short rate process is reverting to a time-varying unconditional mean, which itself is also a mean reverting process with reverting trend measured by υ_0 / υ_1, of 2.50%. This estimate is close to the reverting mean implied from the previous SV2 model. Moreover, the corresponding level effect estimate is lowered to round 1/2, which is slightly smaller than that of the two-factor SV model.

Part of Table 1.8a displays the diagnostics for all Gaussian diffusion models, based on the informative standard *t*-ratios of the individual elements of the score vector. These *t*-statistics can be interpreted conveniently as normalized regression residuals. Therefore, large *t*-ratios reveal those characteristics that are not well approximated. It appears that the one-factor CKLS model encounters difficulties to accommodate the scores; the large *t*-ratios on individual score elements associated with the second to sixth Hermite polynomial elements show that the it fails to capture certain aspects of volatility clustering that exists in the data as summarized by the 11117000 auxiliary model. On the other hand, for the accepted two-factor and three-factor SV models, all adjusted *t*-ratios are well below 2.0.

[21] See also Gallant and Tauchen (2002), in which they proposed a two-factor SV model with a mean factor for the Microsoft stock returns.

5.2.2. Non-Gaussian Diffusion Model

Table 1.7b reports the EMM estimation results for the non-Gaussian stable diffusion models with fixed combinations of the shape and skewness parameters. The small *p*-values for the EMM objective function lead to rejections for all the non-Gaussian diffusion models at the 5 percent significance level. The score diagnostics provided in Table 1.8 provide some explanations for the failure of these models. For example, the CKLS model with stable $(\alpha = 1.95, 0)$ errors fails to capture certain aspects of volatility clustering associated with the third and fifth order Hermite polynomial elements. The best fitting model, with a *p*-value of 0.045, is the CKLS-S($\alpha = 1.9$, $\beta = 0.1$) model. All of the score *t*-ratios for this model are smaller than two.

The parameter estimates from the best fitting CKLS-S($\alpha = 1.9$, $\beta = 0.1$) model are similar to the one-factor Gaussian diffusion model, implying a strong mean reversion in the short-term rates and a slightly larger level effect that is less than unity. Overall, the stable diffusion models with $\alpha \geq 1.9$ have higher χ^2 values than the Gaussian model which shows that allowing for heavier tails for the innovation density improves the model fit. In addition, the best fitting CKLS-Stable model has a positive skewness parameter $\beta = 0.1$, implying that fat-tailed *and* positive skewed innovations are important for explaining the data.

The one-factor CKLS-S($\alpha = 1.9$, $\beta = 0.1$) model that allows for fat tails and positive skewness can accommodate many of the complex features of the interest rate series. This model can accommodate outliers much more easily than the Gaussian model, and its fit is similar to the three-factor continuous-time SV model. However, there are some drawbacks associated with the stable diffusion models. In the estimation, we do not freely estimate the shape parameters of the stable Lèvy process. We instead specify several reasonable combinations of the shape and skewness parameters along a rough grid.[22] Estimating the shape parameters freely may make an even better use of the stable process. Still, many relevant issues associated with using the stable distribution need to be explored in a systematic manner. For example, the non-existence of moments of second or higher order is a potential problem from an empirical point of view. Also, when using simulation-based estimation techniques, the value of the shape variable is found to be closely related to the size of the simulation, which introduces difficulties in model comparisons.

5.2.3. RS Models

The first two columns of Table 1.7c reports EMM estimates for the RS models that do not incorporate a level effect: the RS-in-σ model, and the RS-in-$\sigma + SV$ model. For these RS models, the p-values on the EMM objective function are higher than 5% but lower than 10%, providing mild evidence in support of the models. The fitted models imply strong mean reversion in the short rate. They indicate that the short rates are reverting to a positive long-run mean of 9.08% and 6.75%, respectively, which are substantially higher than the long-run mean estimates implied by the single-regime models.[23] The estimates of the regime dependent volatility parameters reveal that regime 1 is a high-volatility regime and regime 2 is a low-volatility regime. The two estimated regime switching probabilities, P_1 and P_2, exceed 0.90 and are similar to estimates reported by other authors. Notice that while the transition probability of staying in the low-volatility regime, P_1, are estimated similarly to those in previous empirical works, the estimates of P_2 (0.91 and 0.94) are slightly lower than what has been shown, implying less persistence of staying in the high-volatility regime for the dynamics of the short rate.

Without implementing the level effect, the RS-in-σ model only allows the conditional volatility to switch across two regimes; that is, any conditional heteroskedasticity can

[22] This strategy has been used in Gallant et al. (1997) for their discrete-time SV model with Student-t errors. Further research on how the SNP model will encompass the Stable distributed errors is of strong interest.
[23] The substantially large difference of the long-run mean estimates in the single-regime models and RS models may suggest a regime-switching mean scenario, which could be left for future research.

only be driven by switches of conditional volatility between two regimes. For such a simple model, the RS-in-σ model does a good job of modeling the volatility dynamics of short rates. It fits much better than the one factor Gaussian CKLS model, and slightly better than the one-factor non-Gaussian Stable CKLS model. The flexibility of incorporating two different levels of volatility is the main reason for the success of this simple RS model relative to many single-regime models. As argued by Gray (1996), the single-regime models treat volatility as being constant at an average level, in which case volatility estimates are uniformly too high during periods of low volatility and uniformly too low during periods of high volatility. Hence, the models fail to describe well the data in either regime.

Contrary to our expectations, the RS-in-$\sigma + SV$ model does not explain the dynamic behavior of short rates appreciably better than the RS-in-σ model even though it allows for an additional source of conditional heteroskedasticity driven by the volatility persistence beyond the regime switched conditional volatility. This is in contrast what Gray (1996) found with his RS-GARCH model. Using a likelihood ratio (LR) test to compare his RS-in-σ model with his RS-in-$\sigma + GARCH$ model, Gray (1996) showed that both the RS effect and the GARCH effect are important. Our EMM estimation results imply that it is not necessary to incorporate the more complex RS-in-$\sigma + SV$ model.

The third column of Table 1.7c reports EMM estimates of the RS-in-$\sigma + Level$ model. This model, with an EMM objective function *p*-value of 0.29, fits much better than the RS models that do not incorporate the level effect. The RS-in-$\sigma + Level$ model can be described as a generalized CKLS model in which the conditional volatility switches between two very persistent regimes. Incorporating both a level effect and a RS effect, the RS-in-$\sigma + Level$ model provides the best performance in terms of fitting the volatility of short rates; it fits even better than the three-factor SV model, in which both the level effect and SV effect are implemented in the underlying structural model. It appears that the flexibility of having two volatility regimes and having the level effect picking up the remaining information is the main reason for the relative success of the RS-in-$\sigma + Level$ model over the single-regime models and the previous RS-in-σ models.

Compared with the ML estimates in Smith (2002), our EMM results are quite different in several respects. The estimated process is reverting to a lower long-term mean with a fast speed and smaller regime-dependent variances. Although the transition probability P_1 is similar, the estimate of P_2 at 0.89 is much lower than what has been shown by Smith (2002). In terms of the estimate for the level effect, the estimated conditional volatility in the RS-in-$\sigma + Level$ model is sensitive to the level of the short rates; that is, the level effect parameter is significantly different from zero. However, the magnitude of the estimated level effect, much lower than that reported in Smith (2002) at 0.92, is very similar as these in our multi-factor SV models. It appears that the combination of the level effect with either a RS factor or a SV factor does not influence the importance of the level effect.

The last column of Table 1.7c shows results for the RS-in-$\sigma + Level + SV$ model. Characterized by combining all three effects of the level, RS, and SV effects within one model, the RS-in-$\sigma + Level + SV$ model is rejected by the EMM objective function at the significant level 5%. The score diagnostic *t*-ratios in Table 1.8b show that the score elements associated with the first, second, fourth Hermite polynomial elements and ARCH and GARCH coefficients are larger than two, which suggests that the RS-in-$\sigma + Level + SV$ model has trouble capturing the associated features as summarized by the 11117000 auxiliary model. A noticeable result for this most complex model specification is that the level effect has been almost squeezed out by the SV effect and the RS effect; its estimate is not significant from zero. Comparing the EMM results for the RS-in-$\sigma + Level + SV$ model with the other models, provides a way for addressing an important issue; that is, whether or not we need to include both RS and SV in the process of fitting the dynamics of the short rates. The answers from Gray (1996), Smith (2002), and many others are somewhat ambiguous due to the fact that traditional hypothesis testing procedures for evaluating the existence of Markov switching are nonstandard. For example, using the LR test, the statistical significance of the second regime cannot be tested using chi-square critical values because the parameters associated with the second regime are not identified under the null of a single regime. Although some extended tests have been developed for solving such kinds of difficulties, EMM provides a rather easy procedure to answer the issue by simply comparing the corresponding *p*-values for different non-nested model specifications. From our estimation, it indicates that either a RS with level effect or an SV with level effect, but not both, are needed to adequately fit the data series of the short rate.

6. Conclusion:

In this paper we develop a framework for evaluating and comparing the empirical fit of a number of discrete-time and continuous-time models for the US short rate. The models we consider include Gaussian diffusion models, non-Gaussian diffusion models with stable process, and different types of Markov RS models. A comprehensive model comparison is provided by utilizing the EMM estimation, which allows for ranking the non-nested model specifications. For the continuous-time models, we confirm the results from the existing empirical literature that the one-factor Gaussian diffusion model constitutes a poor candidate model for the short rate process. We find that a one-factor stable diffusion model shows stronger explanatory power to that of the one-factor Gaussian model, and that the multi-factor SV models (a two-factor SV model and a three-factor SV model) shows much better fitting performances. For the discrete-time RS models, we find that the simple RS-in-σ model, which allows the conditional variance to switch between regimes, describes the data surprisingly well. We also find that there are no fitting improvements of the extended RS-in-$\sigma + SV$ model over the RS-in-σ model, and of the extended RS-in-$\sigma + Level + SV$ model over the RS-in-$\sigma + Level$ model. These results suggest that either an SV effect or a RS effect, but not both, are needed for describing the data accurately. This point is consistent with the argument of Smith (2002), although his conclusion is much more informal and ambiguous. In summary, our multi-factor SV models and the RS-in-$\sigma + Level$ model provide the overall best fits for the short rate process. The success of our two-factor SV model is opposite to the general belief exiting in the literature that two factors are not enough to accommodate the complex process of short rates. Figure 1.7 displays representative simulated paths from three preferred models. Relative to the actual interest rate series in Figure 1.2, the three simulation series are capable of generating some extreme volatile periods as the monetary experiment experience, and share qualitative features with the actual interest rate data.

We also provide insights on the measurement of one of the important features of the US short rates, the level effect. Our finding shows that the level effect is similarly estimated a bit higher than 1/2 in the preferred multi-factor SV models and the RS-in-$\sigma + Level$ model, which is consistent with the finding in AL (1997). Although the corresponding estimate obtained from the RS-in-$\sigma + Level + SV$ model is significantly weakened, the estimated parameter is not significantly different from zero. Our estimations imply that the estimated level effect is relatively robust to the sample used for estimation; it may be spuriously low or high for misspecified models that fail to capture the time-varying and heteroskedastic behavior of the short rates.

REFERENCES:

Ahn, D.-H., Dittmar, R.F., Gallant, A.R., and Gao, B., 2003. “Purebred or hybrid? Reproducing the volatility in term structure dynamics”, Journal of Econometrics 116, 147-180.

Ahn, D.-H., Dittmar, R.F., Gallant, A.R., and Gallant, A.R., 2002, "Quadratic Term Structure Models: Theory and Evidence," Review of Financial Studies 15(1), 243-88.

Andersen, T., L. Benzoni and J. Lund (2002) "An Empirical Investigation of Continuous-Time Models for Equity Returns," Journal of Finance 57, 1239-1284.

Andersen, T.G. and Sorensen, B.E., 1997. “Estimating Continuous-time Stochastic Volatility Models of the Short-term Interest Rate”, Journal of Business and Economic Statistics, 14, 328-352.

Andersen, T.G., Chung, Hyung-Jin, and Sorensen, B.E., 1999. “Efficient Method of Moments Estimation of a Stochastic Volatility Model: A Monte Carlo Study”, Journal of Econometrics, 91, 61-87.

Andersen, T.G., and Bollerslev, T., 2005. “A Framework for Exploring the Macroeconomic Determinants of Systematic Risk”, Working Paper, Annual Meeting of American Economics Association.

Ang, A., Bekaert, G., 2002. “Regime Switching in Interest Rates”, Journal of Business and Economic Statistics 20(2), 163-182.

Bansal, R. and Zhou, H., 2002. “Term structure of interest rates with regime shifts”, The Jounal of Finance LVII(5) , 1997-2043.

Barndorff-Nielsen, O.E., 1997. “Normal Inverse Gaussian Distributions and Stochastic Volatility Modeling”, Scand Journal of Statistics 24, 1-13.

Bertoin, J., 1996. “Lèvy Processes”, Cambridge University Press, Melbourne, NY.

Bliss, R.R. and Smith, D.C. 1998. “The Elasticity of Interest Rate Volatility: Chan, Karolyi, Longstaff, and Sanders Revisited”, Journal of Risk 1(1), 21-246.

Bollerslev, T., 1986. “Generalized Autoregressive Conditional Heteroskedasticity”, Journal of Econometrics 31, 307-327.

Boudoukh, Jacob et al., 1999. “Regime Shifts and Bond Returns”, working paper, New York University.

Brandt, M. and Chapman, D., 2002. “Comparing Multifactor Models of the Term Structure”, working paper, Duke University.

Broto, C. and Ruiz, E., 2002. “Estimation Methods for Stochastic Volatility Models: A Survey”, Journal of Economic Surveys 18(5), 613-649.

Carr, P.,Geman, H., Madan, D., and Yor, M., 2000. “The Fine Structure of Asset Returns: An Empirical Investigation”, Journal of Business 75(2), 305-332.

Cai, J., 1994. “A Markov Model of Switching-Regime ARCH”. Journal of Business and Economic Statistics 12, 309-316.

Chan, K.C., Karolyi, G.A., Longstaff, F., and Sanders, A. 1992. “The Volatility of Short-term Interest Rates: An empirical Comparison of Alternative Models of the Term Structure of Interest Rates”, Journal of Finance 47, 1209-1227.

Chernov, M. and Ghysels, E., 2000. "A Study towards a Unified Approach to the Joint Estimation of Objective and Risk Neutral Measures for the Purpose of Options Valuation," Journal of Financial Economics 56, 407-458.

Christiansen, C., 2005. “Level-ARCH Short Rate Models with Regime Switching: Bivariate Modeling of US and European Short Rates”, Working Paper, Centre for Analytical Finance (CAF).

Cox, J.C., Ingersoll, J.E., and Ross, S.A., 1985. “A Theory of the Term Structure of Interest Rates”, Econometrica 53, 385-407.

Dai, Q. and Singleton, K.J., 2003. “Term Structure Dynamics in Theory and Reality", Review of Financial Studies.

Dai, Q. and Singleton, K.J., 2000. "Specification Analysis of Affine Term Structure Models," Journal of Finance 55, 1943-1978.

Dai, Q., Singleton, K.J., and Yang, W., 2004. “Regime Shifts in a Dynamic Term Structure Model of U.S. Treasury Bond Yields,” Working paper, New York University.

Danielsson, J., 1994. “Stochastic Volatility in Asset Prices: Estimation with Simulated Maximum Likelihood”, Journal of Econometrics 61, 375–400

Driffill, J., T. Kenc, M. Sola, and F. Spagnolo (2004): “An Empirical Examination of Term Structure Models with Regime Shifts,” Discussion paper, Centre for Economic Policy Research.

Duffee, G., 1993. “On the Relation between the Level and Volatility of Short-Term Interest Rates: A Comment on Chan, Karolyi, Longstaff and Sanders”, Working Paper, Federal Researve Board Washington D.C.

Duffie, D., Pan, J., and Singleton, K., 2000. “Transform Analysis and Asset Pricing for Affine Jump Diffusions”, Econmetrica 68, 1343-1376.

Dahlquist, M. and Gray, S.F., 2000. “Regime-Switching and Interest Rates in the European Monetary System”, Journal of International Economics 50, 399-419.

Engle, R.F., 1982. “Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of United Kingdom Inflation”, Econometrica 50, 987-1007.

Engle, R.F. and Gonzalez-Rivera, G., 1991. “Semiparametric ARCH Models”, Journal of Business and Economic Statistics 9(4), 345-359.

Evans, M., 2001. “Real Risk, Inflation Risk, and the Term Structure", working paper, Georgetown University.

Fama, E.F., 1965. “The Behavior of Stock Market Prices”, Journal of Business 38, 34-105.

Gallant, A. R, Hsieh D.A., and Tauchen, G., 1997. “Estimation of Stochastic Volatility Models with Diagnostics,” Journal of Econometrics, 81(1), 159-192.

Gallant, R.A. and Tauchen, G., 2001. “Efficient method of moments”, Manuscript, University of North Carolina.

Gallant, R.A. and Tauchen, G., 2001. “SNP: A program for nonparametric time series analysis, Version 8.8, User’s Guide”, Manuscript, University of North Carolina.

Gallant, R.A. and Tauchen, G., 2002. “EMM: A program for efficient method of moments estimation, Version 1.6, User’s Guide”, Manuscript, University of North Carolina.

Garcia, P. and Perron, P., 1996. “An Analysis of the Real Interest Rate Under Regime Shifts”, The Review of Economic and Statistics 78(1), 111-125.

Garcia, R., Renault, E. and Veredas, D., 2004. “Estimation of Stable Distributions by Indirect Inference”, Working paper.

Gray, S.F., 1996. “Modeling the conditional distribution of interest rate as a regime-switching process”, Journal of Financial Economics 42, 27-62.

Hamilton, J.D., 1994. "Time Series Analysis”, Princeton University Press, Princeton.

Hamilton, J.D. and Susmel, R., 1994. “Autoregressive Conditional Heteroskedasticity and Changes in Regime”, Journal of Econometrics 64, 307-333.

Harvey, A., Ruiz, E., and Shephard, N, 1994. "Multivariate Stochastic Variance Models”, Review of Economic Studies, Blackwell Publishing, 61(2), 247-264.

Harvey, A. and Shephard, N., 1996. "Estimation of an Asymmetric Stochastic Volatility Model for Asset Returns”, Journal of Business and Economic Statistics 14, 429-434.

Iorio, F.D. and Calzolari, G. 2005. "Discontinuities in Indirect Estimation: An Application to EAR Models”, Computational Statistics and Data Analysis, forthcoming.

Jacquier, E., Polson, N.G., and Rossi, P.E., 1994. "Bayesian Analysis of Stochastic Volatility Models", Journal of Business and Economic Statistics 12(4), 413-17.

Jondeau E. and Rockinger M., 2003. “Conditional volatility, Skewness, and Kurtosis: Existence, Persistence, and Comovements”, Journal of Economic Dynamics & Control, 27, 1699-1737.

Kim, Chang-Jin, 1994, “Dynamic Linear Models with Markov-Switching”, Journal of Econometrics 60, 1-22.

Kim, Chang-Jin, and Nelson, C.R., 1999, “State-Space Models with Regime Switching: Classical and Gibbs-Sampling Approaches with Applications”, MIT Press.

Lahiri, K. and Wang, J.G., 1996. "Interest Rate Spreads as Predictors of Business Cycles", Statistical Methods in Finance: Handbook of Statistics (14), Edited by G.S.Maddala and C.R.Rao, North Holland, 297-315.

Lahiri, K., Ivanova, D. and Seitz, F., 2000. "Interest Rate Spreads as Predictors of German Inflation and Business Cycle", International Journal of Forecasting, 39-58.

Lambert, P. and Laurent S. 2000. “Modeling financial time series using GARCH-type models with a skewed student distribution for the innovations”, Discussion Paper 0125, Institute de Statistique, Universite Catholique de Louvain.

Landen, C., 2000. “Bond Pricing in a Hidden Markov Model of the Short Rate", Finance and Stochastics 4, 371-389.

Liechty, J. C., and Roberts, G.O., 2001. “Markov Chain Monte Carlo Methods for Switching Diffusion Models,” Biometrika, 88(2), 299-315.

Liesenfeld, R. and Richard, J.F. 2003. “Univariate and multivariate stochastic volatility models: estimation and diagnostics”, Journal of Empirical Finance, 10, 505-531.

Liu, S-Mi. and Brorsen, B.W. 1995. “Maximum Likelihood Estimation of a GARCH-Stable Model”, Journal of Applied Econometrics, 10, 275-285.

Liu, M., 2000. "Modeling Long Memory in Stock Market Volatility," Journal of Econometrics 99(1): 139-71.

Lo, A.W., 1988. “Maximum Likelihood Estimation of Generalized Ito Processes with Discretely Sampled Data,” Econometric Theory 4, 231-247.

Longstaff, F.A. and Schwartz, E.S., 1992. “Interest Rate Volatility and the Term Structure: A Two-Factor General Equilibrium Model”, Journal of Finance 47(4), 1259-1282.

Matacz, A., 2004. “Financial Modeling and Option Theory with the Truncated Lèvy Process”, working paper.

Marinelli C. and Rachev, S., 2002. “Some Applications of Stable Models in Finance”, working paper.

Merton, R., 1976. “Option Pricing When Underlying Stock Returns are Discontinuous”, Journal of Financial Economics 3, 125-144.

Naik, V. and Lee, M. H., 1997. “Yield Curve Dynamics with Discrete Shifts in Economic Regimes: Theory and Estimation", Working paper, University of British Columbia.

Sandmann, G. and Koopman, S.J., 1998, "Estimation of Stochastic Volatility Models via Monte Carlo Maximum Likelihood", Journal of Econometrics, 87(2), 271-301.

Smith, D.R., 2002. “Markov-Switching and Stochastic Volatility Diffusion Models of Short-Term Interest Rates”, Journal of Business and Economic Statistics 20(2) 183-97.

So, M.K., Lam, K. and Li, W.K., 1998. “A Stochastic Volatility Model with Markov Switching”, Journal of Business and Economic Statistics 7, 177-192.

Taylor, S.J., 1986. “Modeling Financial Time Series”, Wiley, New York.

Taylor, S.J., 1994. “Modeling Stochastic Volatility”, Mathematical Finance 4, 183-204.

Wu, S. and Zeng, Y., 2003. “Regime-switching Risk in the Term Structure of Interest Rates", working paper, University of Kansas.

Zivot, E., and Wang, J., 2005. “Modeling Financial Time Series with S-PLUS”, Springer-Verlag Publishing, forthcoming.

Appendix: Implementation of SNP Selection

To enhance the searching efficiency, we utilize the following strategies for determining the most appropriate SNP model. The expository discussion is in Gallant and Tauchen (2001).

(1) As a general rule with financial data, we always move K_z from 0 up to 4. Due to the fat-tailed error densities relative to the Gaussian for financial data, the polynomials has to increase the mass around zero, depress the mass on either side of zero and then increase the mass in the tails by going to infinity on the left and right side. Not linear, quadratic, cubic, but only the quartic polynomial is needed in order to reach the above goal easily and successfully. (2) We put an upper bound of 8 for K_z in order to improve the stability of computation, because the polynomials fit little wiggles when $K_z > 8$. (3) We also put an upper bound of 8 for L_r when fitting the SNP density as VAR-ARGH leading terms. (4) The spline transformation is recommended to use, which is essential for extremely persistent data such as interest rates. (5) In processing a specific starting parameter set, we perturb each active parameter as

$$\rho_i \rightarrow (1 + u \times tweak)\rho_i$$

where u is uniform (-1,1), then iterate from these values for 10 iterations, and repeat this process for many trials. Lastly, it iterates from the best parameter values of these 10 trials until convergence. Therefore, bad starting values leading to local optima are not a concern. This random restart strategy yielded satisfactory fits, sometime improving the estimations substantially; we also utilize this strategy in the estimation step.

Figure 1.1: Procedures of EMM Methodology

EMM procedure consists of two steps (1) the projection step, which is accomplished by projecting the data onto the SNP model, and (2) the estimation step, in which structural parameters are extracted from the summary of the data by minimizing the chi-squared criterion.

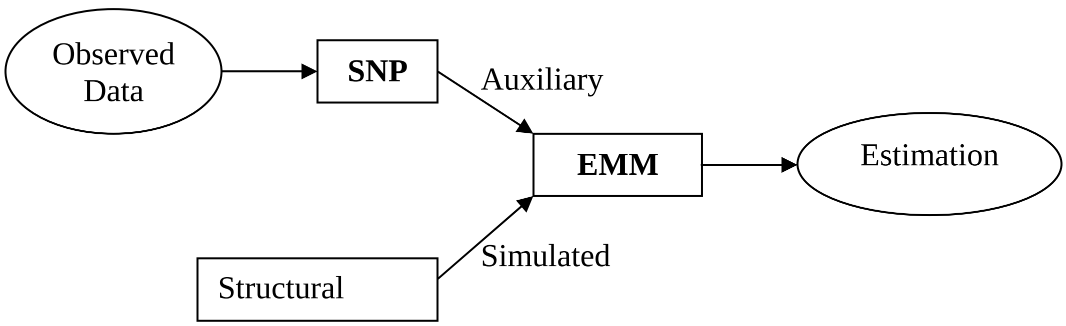
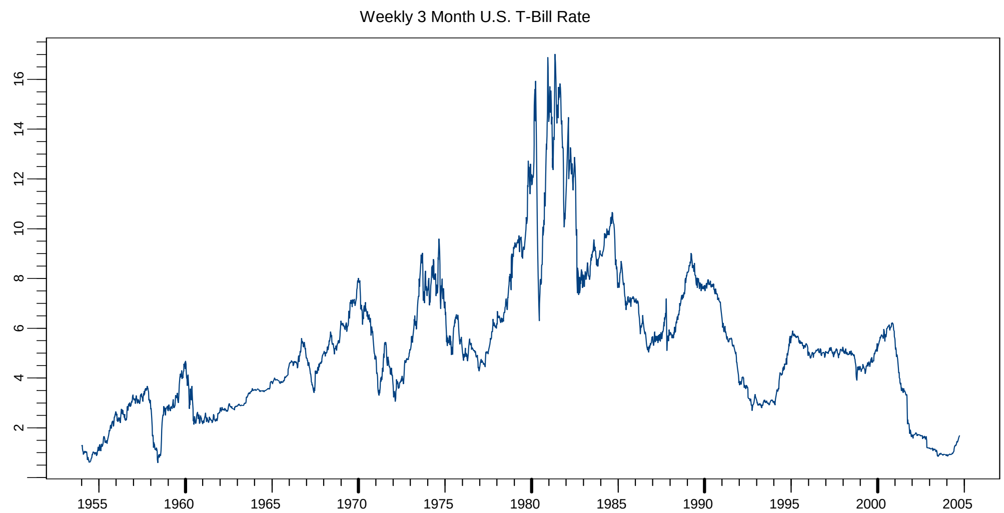


Figure 1.2: Time Series Plots

There are 2648 weekly observations of the 3-month T-Bill rates, ranging from January 4, 1954 to September 24, 2004. The raw data (percent) is plotted in panel (A); the first order difference of the raw data is presented in panel (B).

(A)



(B)

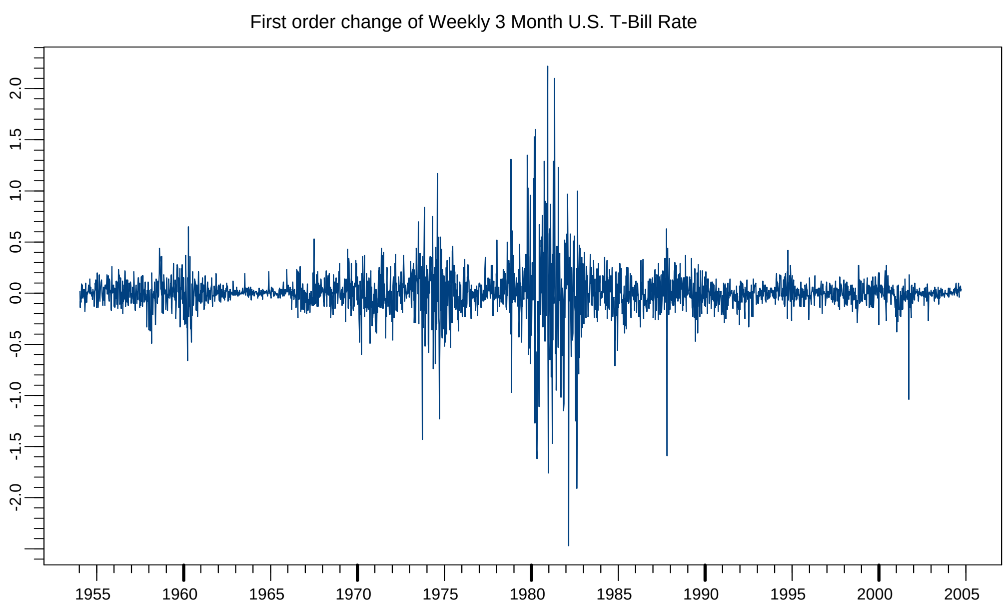
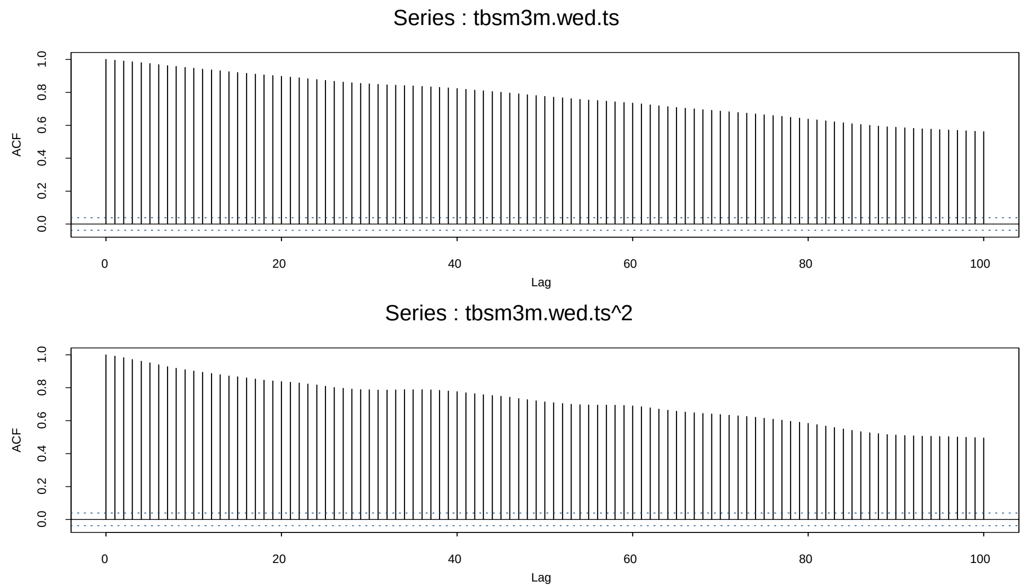


Figure 1.3: Autocorrelation Plots

The ACF plots for the raw data (percent) and the squared series are given in panel (A); the ACF plots of the first order difference of the raw data and its squared series are presented in panel (B).

(A)



(B)

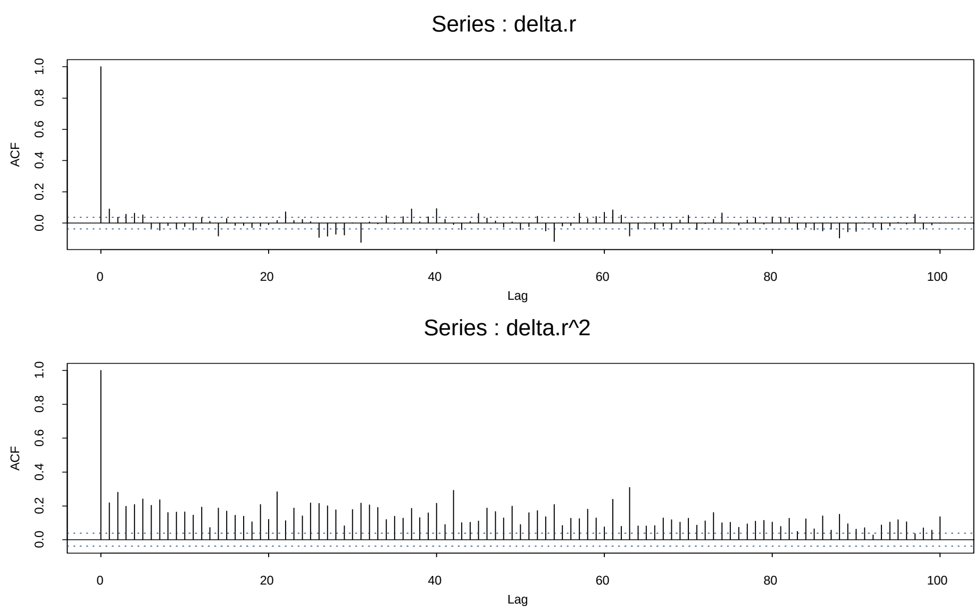
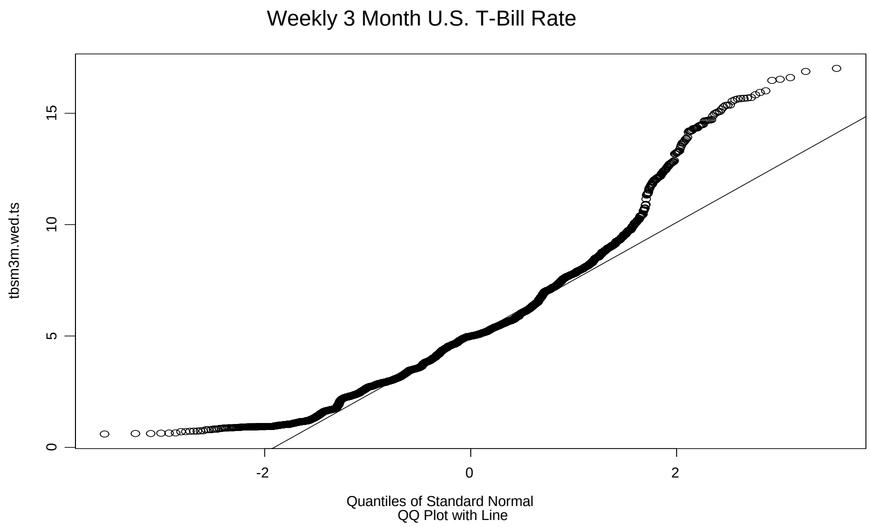


Figure 1.4: QQ Plots with 45° QQ Line

The qq-plot for the raw data (percent) is given in panel (A); the qq-plot of the first order difference of the raw data is presented in panel (B).

(A)



(B)

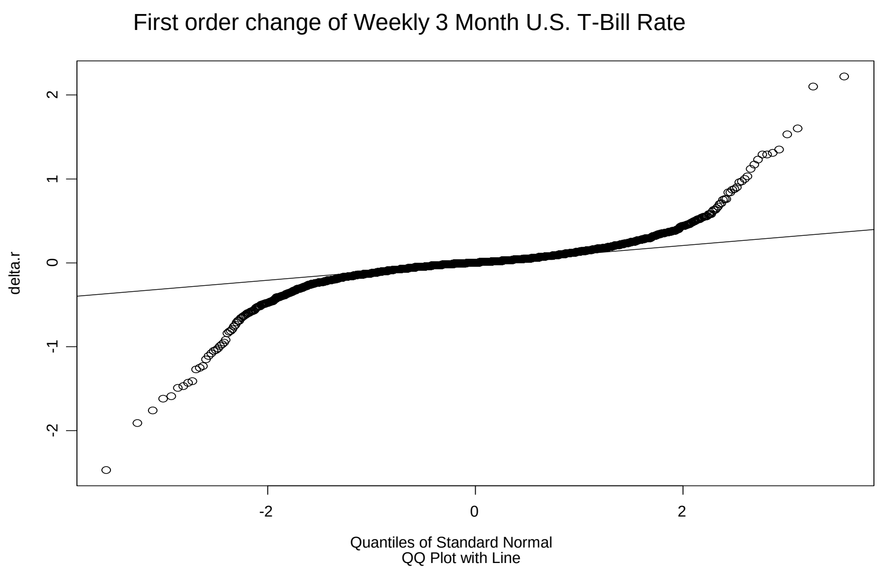


Figure 1.5: Simulated data from Fitted SNP Models

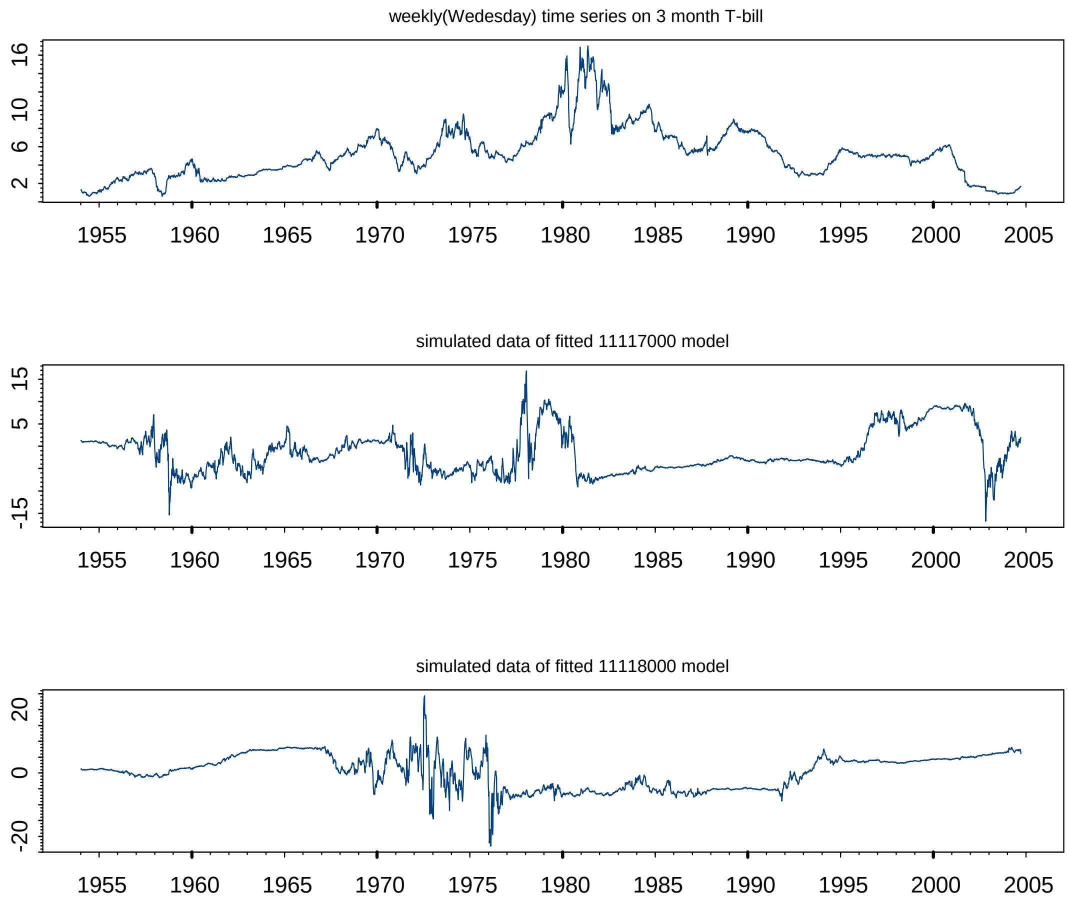
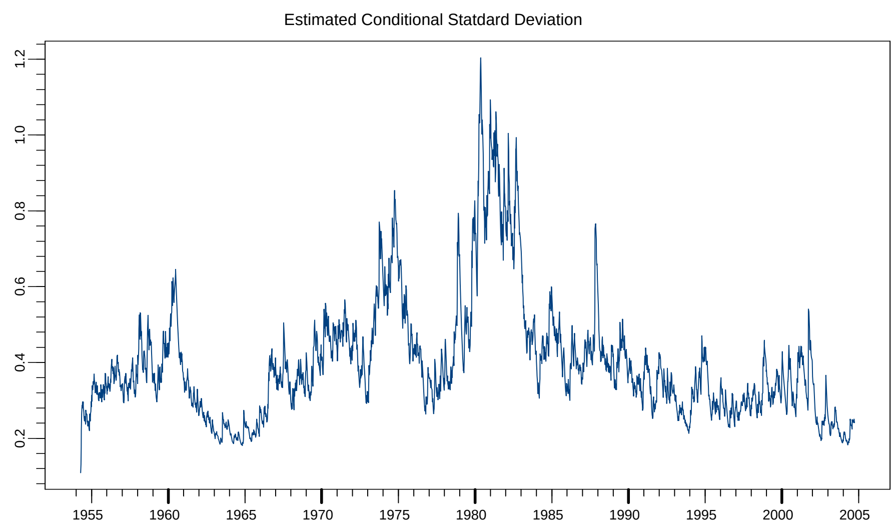


Figure 1.6: Diagnostic Tests for SNP Model 11117000

The panel (A) gives the estimated conditional volatilities of the data, which is persistent and volatile; the panel (B) shows the conditional density, which is peaked in the center with heavy tails.

(A)



(B)

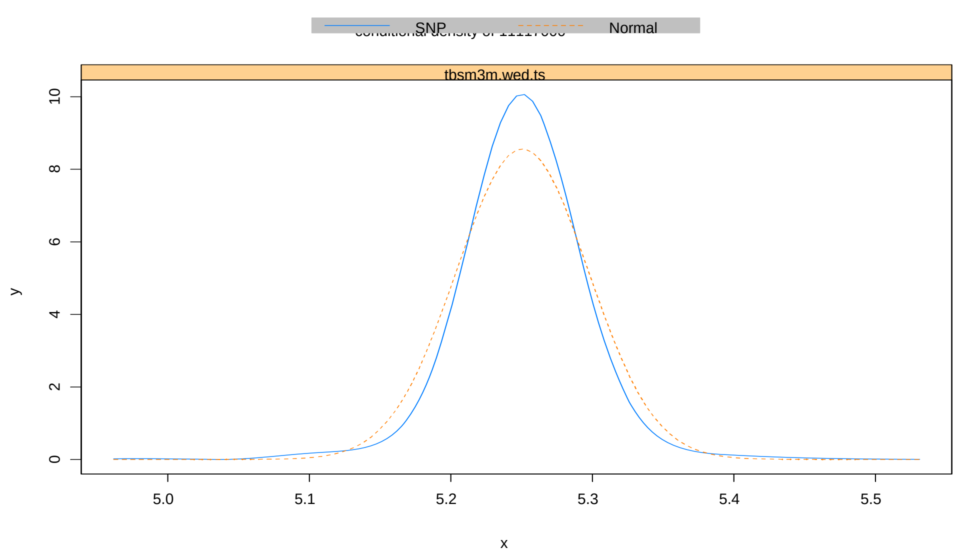
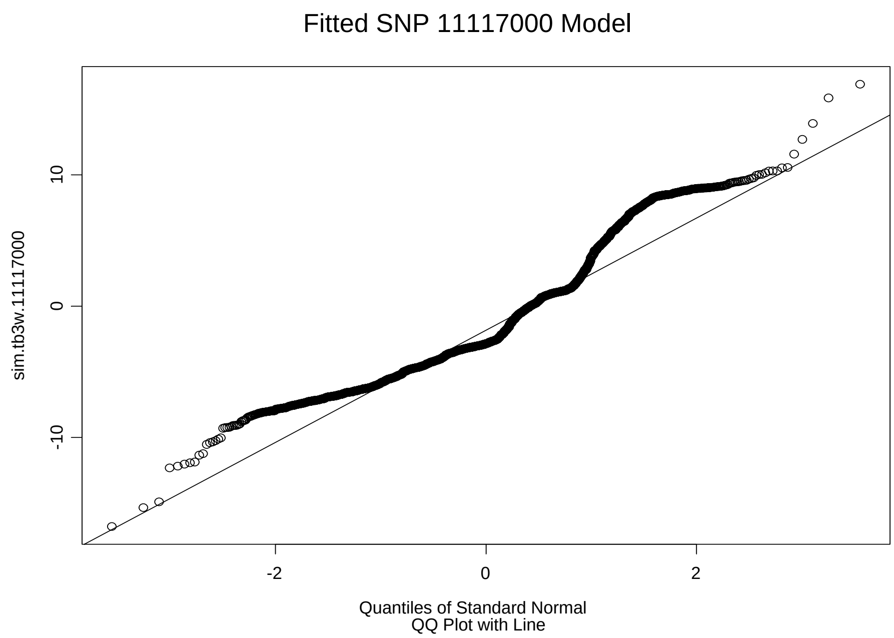


Figure 1.6: Diagnostic Tests for SNP Model 11117000 (Conj)

The panel (C) gives the qq-plot of the simulated series from the preferred SNP model 1117000; the panel (D) shows the qq-plot of the change of the simulated series.

(C)



(D)

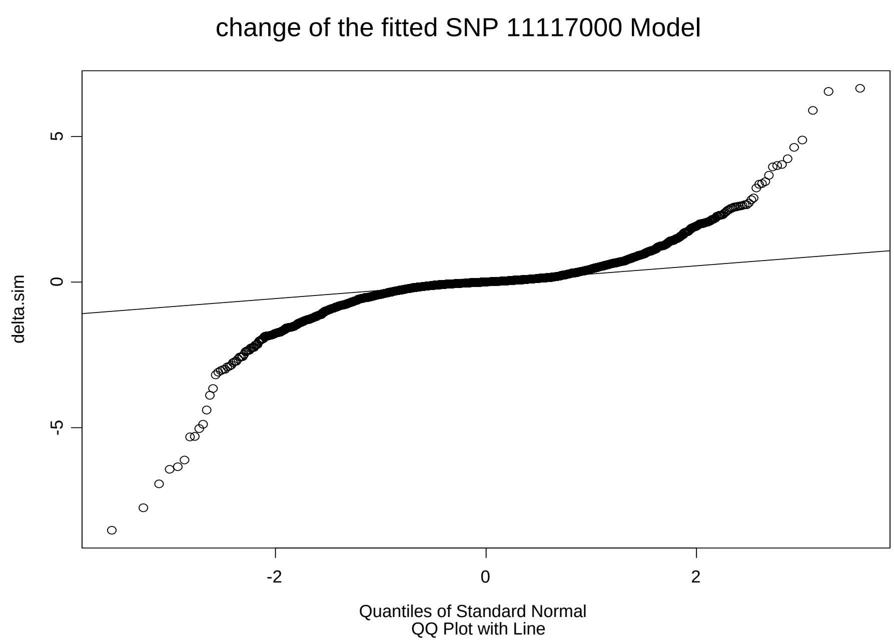
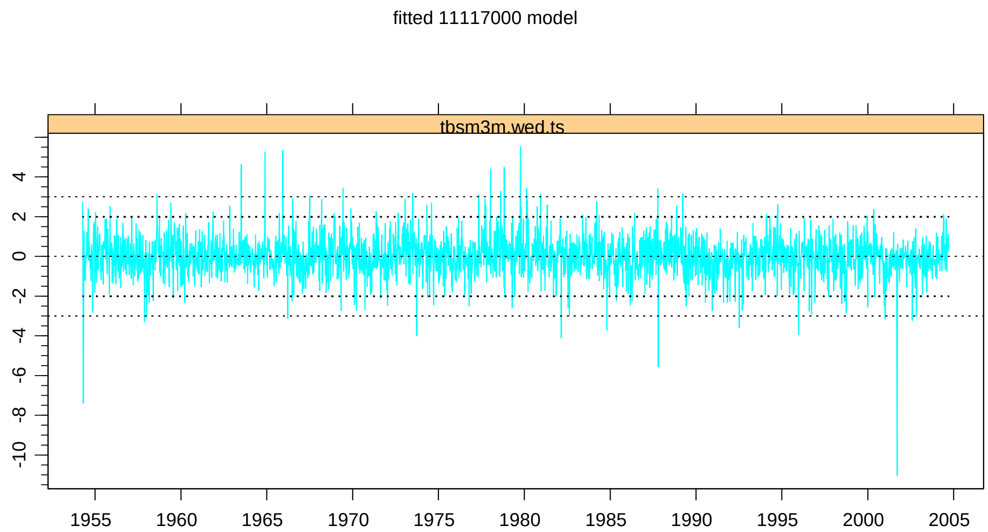


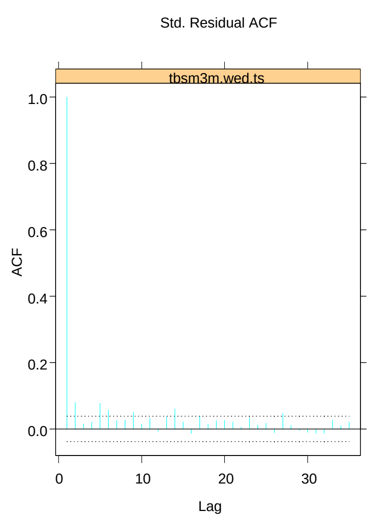
Figure 1.6: Diagnostic Tests for SNP Model 11117000 (Conj)

The panel (E) represents the standardized residuals, which seems to resemble a Gaussian white noise process. Lastly, the panel (F) provides the autocorrelation plot for the residuals and the squared residuals, implying no significant autocorrelation for both of them.

(E)



(F)



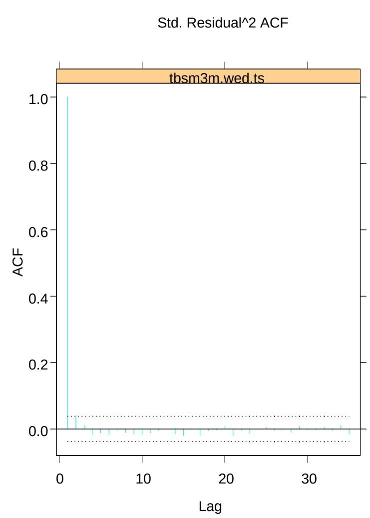
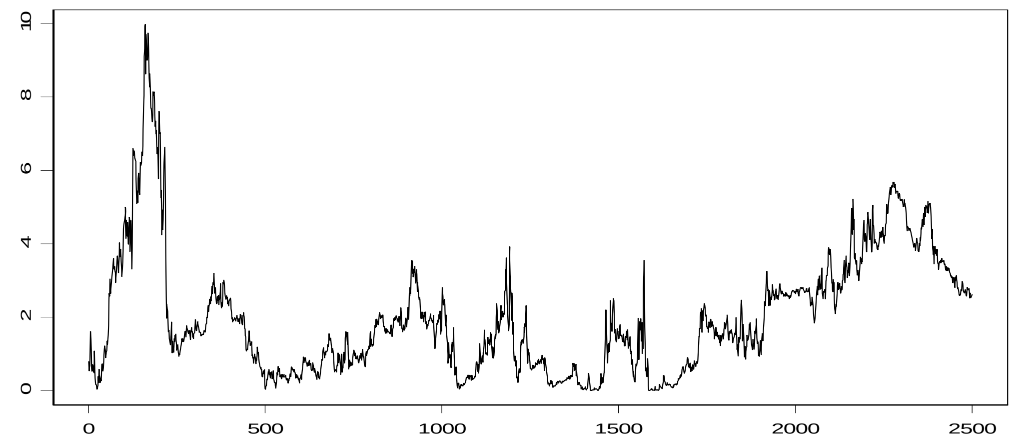
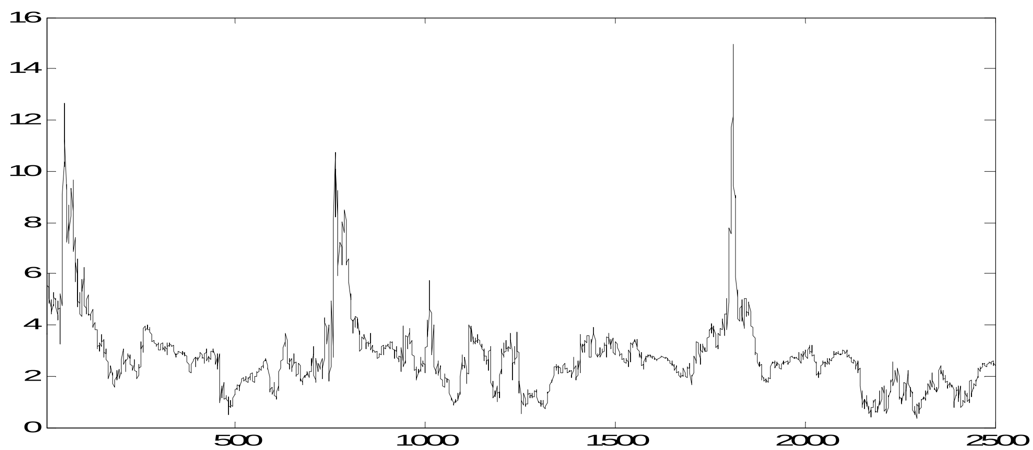


Figure 1.7: Plots of Simulations from Preferred Models

(A) Simulation from the fitted two-factor SV model



(B) Simulation from the fitted three-factor SV model



(C) Simulation from the fitted RS-in-$\sigma + Level$ model

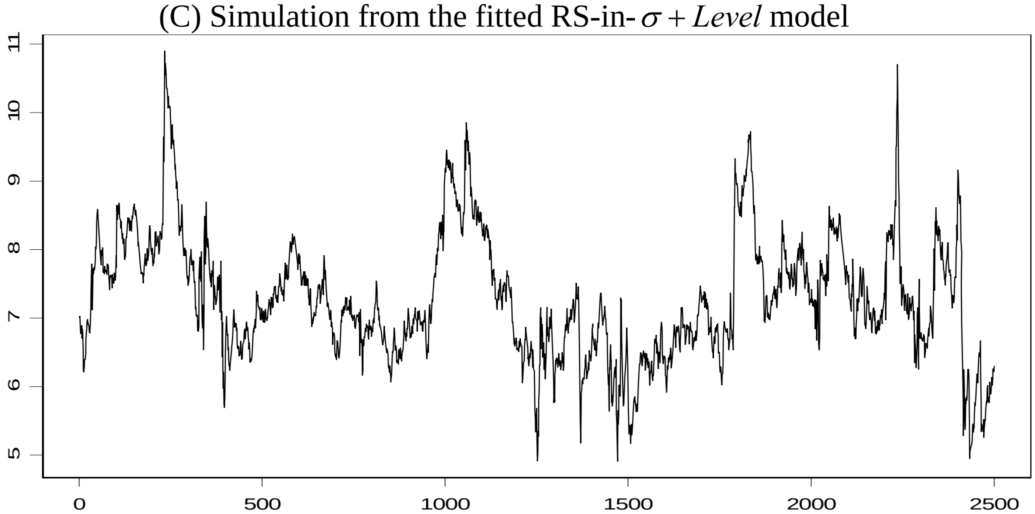


Table 1.1: SNP Tuning Parameters

A useful taxonomy of SNP models is defined by putting certain restrictions on the tuning parameters, according to Gallant and Tauchen (1997).

Parameter Setting	Characterization of
$L_u = 0, L_g = 0, L_r = 0, L_p \geq 0, K_z = 0, K_x = 0$	iid Gaussian
$L_u > 0, L_g = 0, L_r = 0, L_p \geq 0, K_z = 0, K_x = 0$	Gaussian VAR
$L_u > 0, L_g = 0, L_r = 0, L_p \geq 0, K_z > 0, K_x = 0$	Semiparametric VAR
$L_u \geq 0, L_g = 0, L_r > 0, L_p \geq 0, K_z = 0, K_x = 0$	Gaussian ARCH
$L_u \geq 0, L_g = 0, L_r > 0, L_p \geq 0, K_z > 0, K_x = 0$	Semiparametric ARCH
$L_u \geq 0, L_g > 0, L_r > 0, L_p \geq 0, K_z = 0, K_x = 0$	Gaussian GARCH
$L_u \geq 0, L_g > 0, L_r > 0, L_p \geq 0, K_z > 0, K_x = 0$	Semiparametric GARCH
$L_u \geq 0, L_g \geq 0, L_r \geq 0, L_p > 0, K_z > 0, K_x > 0$	Nonlinear nonparametric

Table 1.2: Descriptive Statistics

The statistics summary is given in Panel (A) for the raw data, which are 2648 weekly observations of the 3-month T-Bill rates, ranging from January 4, 1954 to September 24, 2004. The panel (B) shows the statistics summary for the change of the raw data.

(A)

Sample Quantiles	Min: 0.6	1Q: 3.18	Median: 4.99	3Q: 6.67	Max:17.01
Sample Moments	Mean: 5.246	Std. Dev. : 2.849	Skewness: 1.065	Kurtosis: 4. 712	

(B)

Sample Quantiles	Min: -2.47	1Q: -0.07	Median: 0	3Q: 0.077	Max:2.22
Sample Moments	Mean:1.5e-4	Std. Dev. : 0.236	Skewness: -0.522	Kurtosis: 24.81	

Table 1.3: SNP Fitting Strategy

The SNP score generator has a leading (G)ARCH term with L_u lags in conditional mean. The standardized innovation has a normal density stretched by a squared Hermite polynomial with degree of K_z. Similarly, the coefficient of the z-polynomial may depend on the lagged observations through a K_x degree polynomial. For univariae SNP density, the interaction polynomial terms, I_z and I_x, are ignored. The fitting strategy is shown by the following diagram.

<u>VAR-ARCH leading:</u>

```
   10010000 →→→→ 10110000
→  20010000      10210000
   30010000      …
   …             10810000→ 10814000
   70010000                10815000
   80010000                …
                           10818000 → 10818010; 10818020
```

<u>VAR-GARCH leading:</u>

```
10010000 → 11110000
           11114000
           …
           11117000  → 11117010; 11117020 (check conditional heterogeneity)
           11118000  → 11118010; 11118020
```

Table 1.4: SNP Estimation and Selection

The continued table reports the choice of SNP density and the BIC value, based on which we choose our preferred specification. We find that a VAR-GARCH model 11117000 is the BIC preferred model for T-Bill rates using the searching strategy specified in Table 3.

ARCH-Leading	L_u	L_g	L_r	L_p	K_z	I_z	K_x	I_x	BIC
10110000	1	0	1	1	0	0	0	0	-1.377
10210000	1	0	2	1	0	0	0	0	-1.4851
10310000	1	0	3	1	0	0	0	0	-1.5307
10410000	1	0	4	1	0	0	0	0	-1.5618
10510000	1	0	5	1	0	0	0	0	-1.5729
10610000	1	0	6	1	0	0	0	0	-1.5836
10710000	1	0	7	1	0	0	0	0	-1.5866
10810000	**1**	**0**	**8**	**1**	**0**	**0**	**0**	**0**	**-1.587**
10814000	1	0	8	1	4	0	0	0	-1.6238
10815000	1	0	8	1	5	0	0	0	-1.6221
10816000	**1**	**0**	**8**	**1**	**6**	**0**	**0**	**0**	**-1.6301**
10817000	1	0	8	1	7	0	0	0	-1.6287
10818000	1	0	8	1	8	0	0	0	-1.6285
10816010	1	0	8	1	6	0	1	0	-1.6247
10816020	1	0	8	1	6	0	2	0	-1.6152
GARCH-Leading	L_u	L_g	L_r	L_p	K_z	I_z	K_x	I_x	BIC
11110000	1	1	1	1	0	0	0	0	-1.5957
11114000	1	1	1	1	4	0	0	0	-1.6374
11115000	1	1	1	1	5	0	0	0	-1.6359
11116000	1	1	1	1	6	0	0	0	-1.6344
11117000	**1**	**1**	**1**	**1**	**7**	**0**	**0**	**0**	**-1.6407**
11118000	1	1	1	1	8	0	0	0	-1.6395
11117010	1	1	1	1	7	0	1	0	-1.6325
11117020	1	1	1	1	7	0	2	0	-1.622
11118010	1	1	1	1	8	0	1	0	-1.6298
11118020	1	1	1	1	8	0	2	0	-1.618

Table 1.5: Parameter Estimates of Projected SNP Density

This preferred SNP model of 11117000 is a GARCH (1,1) with a nonparametric error density represented as a seven-degree Hermite expansion where the Hermite coefficients are state independent. The auxiliary model and the conditional density 1117000 are given by

$$y_t = \mu_0 + b_1 x_{t-1} + \sigma_t z_t$$

$$\sigma_t = \rho_0 + p_1 \mid y_{t-1} - \mu_{x_{t-2}} \mid + g_1 \sigma_{t-1}$$

$$f(y_t \mid x_{t-1}, \theta) = \left[P(z_t, x_{t-1})\right]^2 N(y_t; \mu_0 + b_1 y_{t-1}, \sigma_t^{\ 2})$$

$$\text{with } P(z_t, x_{t-1}) = \sum_{i=0}^{K_z=7} a_i z_t^i$$

Parameter	Estimate	Standard Error	T-statistics
Hermite a_0	0.0000	0. 0000	0.000
Hermite a_1	-0.0431	0. 0330	-1.3072
Hermite a_2	-0.3073	0. 0215	-14.3144
Hermite a_3	0.0299	0. 0189	1.5790
Hermite a_4	0.0523	0. 0059	8.8977
Hermite a_5	-0.0071	0. 0033	-21358
Hermite a_6	-0.0021	0. 0004	-5.3421
Hermite a_7	0.0004	0. 0002	2.8532
Mean μ_0	0.0021	0. 0012	1.7891
Mean b_1	0.9996	0. 0009	1106.6166
ARCH ρ_0	0.0017	0. 0001	12.3629
GARCH p_1	0.2463	0. 0150	16.4166
GARCH g_1	0.8517	0. 0072	117.8965
BIC: -1.6418	HQ: -1.6503	AIC: -1.6551	Log L: 4371.665

Table 1.6: SNP Models used by Selected Papers for Financial Data

This table lists a number of SNP models that have been utilized for EMM estimations in selected papers with applications of the stochastic volatility modeling, the term structural dynamics and long memory study for the interest rates and equity returns.

Selected Papers	SNP model	Application
Andersen and Lund (1997)	51116000[*]	Stochastic volatility models of the short-term interest rate.
Bansal and Zhou (2000)	10514300	Term structure models using the bivariate dynamics of the yields on the six-month bill and the five-year note.
Ahn et al. (2003)	11114300	Term structure models using the bivariate dynamics of the yields on the six-month bill and the three-year note.
Dai and Singleton (2000)	10214000	Affine term structure models using the swap rates of maturities from six months to ten years.
Andersen et al. (2002)	01118000*	Stochastic volatility models of the S&P 500 Index return.
Chernov et al. (2000)	11118000	Stochastic volatility models and jump diffusion models of the Dow Jones Industrial Average Index return.
Liu (2000)	0025018000[**]	Long memory of equity returns.

[*] Andersen and Lund (1997) used an SNP model with EGARCH (1,1), instead of Level-GARCH (1,1), as the leading term.

[**] This SNP model is a VAR(0) with ARCH(25) conditional variance and the nonparametric error is represented by a stage-independent Hermite Polynomial of degree 18.

Table 1.7a: EMM Model Estimations I

The EMM estimations are given for the one-factor Gaussian diffusion model based on the CKLS model (CKLS-N) and corresponding two-factor and three-factor SV models (SV2 and SV3), which are laid out in section 3.1. The CKLS-N model refers to the model (1.9); the SV2 model refers to the one (1.11); and the SV3 model refers to the one (1.12) with three stochastic factors.

Parameter	CKLS*	CKLS-N	SV2	SV3
ϕ_0	0.0408 (0.022)	0.4818 (0.02593)	0.8428 (0.2669)	-
$-\phi_1$	-0.5921 (0.382)	-0.1927 (0.01830)	-0.2956 (0.1816)	-1.0947 (0.8463)
ϕ_0 / ϕ_1 **	0.0690 (-)	2.5003 (-)	2.8512 (-)	- (-)
γ	1.4999 (0.252)	0.3076 (0.05062)	0.6659 (0.1163)	0.5167 (0.0755)
σ	1.6704 (2.169)	1.3624 (0.19325)	-	-
ω_0	-	-	-0.5912 (0.5122)	-1.4491 (0.3003)
ω_1	-	-	-0.5629 (0.1983)	-0.9250 (0.1990)
ξ	-	-	1.7765 (0.0474)	2.5433 (0.0959)
υ_0	-	-	-	2.5494 (0.7520)
υ_1	-	-	-	-0.9801 (16.5030)
ζ	-	-	-	0.4715 (2.1140)
χ^2	-	50.47	10.35	5.55
$p-value$	-	3.32e-08	0.1107	0.2357
$d.o.f$ ***	-	8	6	4

* The model of Chan et al (1992) with monthly short-term interest rates over period of 6/1964 to 12/1989.
** The fitted long-run reverting mean.
*** The "d.o.f" stands for the "degree of freedom".

Table 1.7b: EMM Model Estimations II

The EMM estimations are given for the non-Gaussian diffusion models based on the CKLS model, which is specified by

$$dr_t = (\phi_0 - \phi_1 r_t)dt + \sigma r_t^{\gamma} dL_t = k_r(\mu_r - r_t)dt + \sigma r_t^{\gamma} dL_t$$

where the L_t is stable Lèvy process with shape variable α, and skewness variable β. The CKLS-N model where L_t is Wiener process is listed for the comparison with the CKLS-S(α, β) models.

Parameter	CKLS*	CKLS-N	CKLS-S (1.8,0)	CKLS-S (1.9,0)	CKLS-S (1.95,0)	CKLS-S (1.9,0.1)
ϕ_0	0.0408	0.4818	0.9043	0.7546	0.7202	1.2154
	(0.022)	(0.02593)	(0.05396)	(0.1272)	(0.06407)	(0.08063)
$-\phi_1$	-0.5921	-0.1927	-1.2370	-0.5325	-0.4502	-4.7425
	(0.382)	(0.01830)	(1.237)	(0.373)	(0.12936)	(0.31014)
ϕ_0 / ϕ_1	0.0690	2.5003	0.7310	1.4171	1.5997	0.2563
	(-)	(-)	(-)	(-)	(-)	(-)
γ	1.4999	0.3076	0.3727	0.3664	1.0675	0.4905
	(0.252)	(0.05062)	(0.01508)	(0.0177)	(0.06533)	(0.03837)
σ	1.6704	1.3624	-0.5207	0.9694	1.0675	0.8965
	(2.169)	(0.19325)	(0.19325)	(0.0171)	(0.02534)	(0.05622)
χ^2	-	50.47	82.19	33.39	23.89	15.8
$p-value$	-	3.32e-08	1.765e-14	5.238e-05	1.195e-04	0.0454**
$d.o.f$	-	8	8	8	8	8

[*] The model of Chan et al (1992) with monthly short-term interest rates over period of 6/1964 to 12/1989.
[**] The corresponding p-value with degree of freedom of six is 0.0149.

Table 1.7c: EMM Model Estimation III

The following estimations are given for the different types of Markov Regime Switching (RS) models, which are laid out in section 3.3. The first two models are OU-based RS models, given by equations of (1.14) and (1.16) respectively. The last two models are CKLS-based, given by equations of (1.17) and (1.18) respectively.

Parameter	OU Based		CKLS Based	
	RS-in-σ	RS-in-σ + SV	RS-in-σ + Level	RS-in-σ + Level + SV
ϕ_0	0.2580 (0.0336)	0.2240 (0.6623)	0.1769 (0.5818)	0.4027 (0.3227)
$-\phi_1$	-0.0284 (0.1640)	-0.0332 (0.5006)	-0.0285 (0.2976)	-0.0408 (1.0771)
ϕ_0 / ϕ_1	9.0845 (-)	6.7470 (-)	6.2070 (-)	9.8701 (-)
γ	-	-	0.5076 (0.1266)	0.0063 (4.5077)
σ_1	0.1472 (0.0628)	-	0.0389 (0.0914)	-
σ_2	0.4613 (0.0556)	-	0.1400 (0.0874)	-
ω_{01}	-	-2.1187 (1.6037)	-	-2.8481 (0.3137)
ω_{02}	-	-1.0078 (0.6564)	-	-1.3966 (0.1915)
ω_1	-	-0.5326 (1.5275)	-	-0.1570 (0.4406)
ξ	-	0.2902 (0.3272)	-	0.2210 (0.2212)
P_1	0.98	0.98	0.98	0.98
P_2	0.91	0.94	0.89	0.94
χ^2	11.01	8.94	6.53	5.55
$p-value$[*]	0.0881	0.0626	0.2916	0.0253
$d.o.f$	6	4	5	3

[*] The p-values are calculated based on the degree of freedom equal to the $d.o.f$ less two.

Table 1.8a: Models Diagnostic T-Ratios I

The adjusted t-rations* are reported for different model specifications based on the same score generator (11117000), for which the parameters refers to the following equations. The adjusted t-ratios are testing whether the fitted sample moments are equal to zero, as predicted by population moments of the SNP density.

$$y_t = \mu_0 + b_1 x_{t-1} + \sigma_t z_t,\ \text{where}\ \sigma_t = \rho_0 + p_1 | y_{t-1} - \mu_{x_{t-2}} | + g_1 \sigma_{t-1},$$

$$f(y_t | x_{t-1}, \theta) = \left[P(z_t, x_{t-1})\right]^2 N(y_t; \mu_0 + b_1 y_{t-1}, \sigma_t^{\,2}) \quad \text{with} \quad P(z_t, x_{t-1}) = \sum_{i=0}^{K_z=7} a_i z_t^i$$

Parameter	CKLS-N	CKLS-S(1.95, 0)	CKLS-S(1.9, 0.1)	SV2	SV3
Hermite a_1	-1.5845	-1.2618	0.03594	-0.3861	0.5721
Hermite a_2	**-2.4841**	-0.7672	-0.5429	0.08518	-0.8662
Hermite a_3	**-2.7178**	**-3.4528**	-1.884	-1.5801	-0.8818
Hermite a_4	**2.0221**	0.5788	-0.3106	1.05861	-0.0750
Hermite a_5	**-2.0172**	**-3.0735**	-1.9019	-1.2937	-0.7872
Hermite a_6	**2.7662**	1.1112	0.09164	1.30066	0.2805
Hermite a_7	-0.5218	-1.4579	-1.324	-0.2149	-0.1253
Mean μ_0	1.1283	0.8312	0.27349	1.71367	1.7819
Mean b_1	**-2.4661**	-1.6667	0.13483	-1.1658	-0.5640
GARCH ρ_0	-1.6397	0.3467	-1.8993	0.6306	0.5843
GARCH p_1	**-2.1221**	0.4447	-1.6631	0.2647	0.3493
GARCH g_1	-1.9152	0.2685	-1.6	-0.3861	0.4326
$p-value$	3.32e-08	1.20e-04	0.0454	**0.1107**	**0.2357**

* According to Gallant and Tauchen (2000), the unadjusted t-ratios are biased downward.

Table 1.8b: Models Diagnostic T-Ratios II

The t-rations are reported for different model specifications based on the same score generator (11117000), for which the parameters refers to the following equations. The t-ratios are testing whether the fitted sample moments are equal to zero, as predicted by population moments of the SNP density.

$$y_t = \mu_0 + b_1 x_{t-1} + \sigma_t z_t, \text{ where } \sigma_t = \rho_0 + p_1 | y_{t-1} - \mu_{x_{t-2}} | + g_1 \sigma_{t-1},$$

$$f(y_t | x_{t-1}, \theta) = \left[P(z_t, x_{t-1})\right]^2 N(y_t; \mu_0 + b_1 y_{t-1}, \sigma_t^{\,2}) \quad \text{with} \quad P(z_t, x_{t-1}) = \sum_{i=0}^{K_z=7} a_i z_t^i$$

Parameter	OU Based		CKLS Based	
	RS-in-σ	RS-in-σ + SV	RS-in-σ + Level	RS-in-σ + Level + SV
Hermite a_1	-0.7338	0.6262	0.7327	**-2.1586**
Hermite a_2	-1.7376	**-2.4572**	-1.5314	**-5.6508**
Hermite a_3	-0.7855	-0.2299	0.3636	-0.4873
Hermite a_4	-1.0537	-1.6070	-1.1830	**-2.5206**
Hermite a_5	-0.6400	-0.3290	-0.0961	-0.0609
Hermite a_6	-0.8775	-0.2241	-0.9017	-0.7921
Hermite a_7	-0.8047	0.3008	-0.9423	0.1670
Mean μ_0	-0.3481	1.7798	0.2709	-0.8511
Mean b_1	0.4950	-0.9013	0.9051	1.4379
GARCH ρ_0	-1.6486	-0.4648	-1.5102	-1.7896
GARCH p_1	-1.5090	-0.7646	-0.5848	**-2.7740**
GARCH g_1	-1.5210	-0.4723	-0.5863	**-2.5500**
$p-value$	0.0881	0.0626	0.2916	0.0253