Estimating continuous-time stochastic volatility models of the short-term interest rate

Torben G. Andersen* , a, Jesper Lund b

aJ.L. Kellogg Graduate School of Management, Northwestern University, Evanston, IL 60208, USA
bDepartment of Finance, Aarhus School of Business, 8210 Aarhus V, Denmark

(Received November 1995; final version received January 1996)

Abstract

We obtain consistent parameter estimates of continuous-time stochastic volatility diffusions for the U.S. risk-free short-term interest rate, sampled weekly over 1954–1995, using the Efficient Method of Moments procedure of Gallant and Tauchen. The preferred model displays mean reversion and incorporates ‘level effects’ and stochastic volatility in the diffusion function. Extensive diagnostics indicate that the Cox–Ingersoll–Ross model with an added stochastic volatility factor provides a good characterization of the short rate process. Further, they suggest that recently proposed GARCH models fail to approximate the discrete-time short rate dynamics, while ‘Level-EGARCH’ models perform reasonably well.

Key words: Short-term interest rate; Stochastic volatility; Continuous-time estimation; Efficient method of moments; ARCH
JEL classification: C14; C15; C32; C51; E43; G13

1. Introduction

The default-free short-term interest rate is a key economic variable. 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, since

*Corresponding author.

We thank Tim Bollerslev, Bob Hodrick, Svend Jakobsen, seminar participants at University of Aarhus, and especially George Tauchen for helpful comments. The second author gratefully acknowledges financial support from the Danish Social Social Science Research Council.
expected equilibrium returns on risky assets are expressed in terms of excess returns relative to the risk-free rate, it is a general reference point for asset pricing. Finally, 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. Accordingly, the short rate is a natural state variable in both dynamic term structure models and full-fledged macroeconomic models.

Consequently, an enormous amount of work has been directed towards modeling and estimation of the short rate dynamics in recent years. In the process, the intrinsic difficulties of developing models that are attractive from a theoretical perspective and perform satisfactorily in empirical applications have become increasingly apparent. The leading theoretical models specify continuous-time processes for the interest rate, originating with the arithmetic Brownian motion representation of Merton (1973). While convenient, this specification has serious drawbacks, such as allowing negative interest rates, and represents, at best, a rough approximation to the actual process. Later modifications incorporate more appealing empirical features, e.g., the mean-reverting component in Vasicek (1977) allows the process to be stationary and dramatically reduces the likelihood of obtaining negative interest rates, while the square-root model of Cox, Ingersoll, and Ross (1985) (CIR) guarantees positivity and captures an aspect of conditional heteroskedasticity by having volatility increase with the level of the interest rate, the so-called 'level effect'. Both the Vasicek and CIR models may be extended to multi-factor models and provide popular closed form solutions for applied work. Hence, a large number of studies have estimated these models from discrete time data on short-term interest rates. However, not surprisingly, more flexible empirical specifications appear necessary in order to obtain an adequate characterization of the actual short rate process. For concreteness, consider the (generalized) CIR model

\[ dr_t = \kappa(\mu - r_t)dt + \sigma r_t^{\gamma} dW_t, \]

where \( r_t \) denotes the short-term interest rate, \( W_t \) is a standard Brownian motion, \( \kappa, \mu, \) and \( \sigma \) are nonnegative parameters, and CIR impose \( \gamma = \frac{1}{2}. \) In this model \( r_t \) mean-reverts towards the level \( \mu, \kappa \) measures the speed of the reversion, and the \( r_t^{1/2} \) term (the level effect) assures positive rates. Chan, Karolyi, Longstaff, and Sanders (1992) (CKLS) demonstrate that relaxing the power restriction results in point estimates of \( \gamma \) around \( \frac{1}{2} \) and significantly in excess of unity. At the same time, evidence from estimation of (generalized) autoregressive conditionally heteroskedastic, or (G)ARCH, models, developed in Engle (1982) and Bollerslev (1986), points towards extremely high degrees of volatility persistence in the interest rate process. These findings have inspired Brenner et al. (1994) and Koedijk et al. (1994) who nest the GARCH and approximate CKLS models under more general discrete-time specifications, and the earlier work of Longstaff and Schwartz (1992) who test a CIR term structure in a model.
including a stochastic volatility factor which they approximate by a discrete-time GARCH specification. These studies confirm the presence of rather extreme conditionally heteroskedastic volatility effects in the interest rate dynamics, while the levels effect weakens relative to the estimates from the CKLS model.

Although these findings are indicative of problems with the standard continuous-time models for the short-term rate, they are not themselves entirely satisfactory. First, due to both institutional features and data constraints, the estimates are typically obtained from monthly, or at best, weekly observations. The authors, accordingly, either directly specify the models in discrete time or employ a discrete-time approximation to the continuous-time process. In any case, the presence of a discretization bias makes it difficult, if not impossible, to map these discrete parameter estimates into corresponding continuous-time parameter estimates. This implies, secondly, that there is no transparent manner in which to gauge the seriousness of the misspecification of, say, the standard CIR model, and renders assessment of the pricing formulas derived from these models difficult. Third, we document that the discrete-time models proposed in the literature often are ill-behaved in the sense that their internal dynamics, at estimated parameter values, are excessively erratic. This severely limits their usefulness for numerical or simulation-based estimation procedures, and may pose a problem for practical modeling of the term structure or pricing of interest-rate-sensitive assets.

As an alternative, we propose direct, consistent estimation of a multi-factor diffusion process for the short-term interest rate process. The specification involves both a stochastic volatility factor and a level effect. Estimation is conducted by the Efficient Method of Moment (EMM) approach of Gallant and Tauchen (1994). The resulting parameter estimates are readily interpretable and shed new light on the requisite modifications of the standard CIR model. Specifically, our findings take the controversy over the size of the level effect 'full circle'. Our point estimates of $\gamma$ are invariably close to $\frac{1}{2}$, while the volatility persistence is high, but clearly within the strictly stationary region. Thus, while CIR initially was rejected relative to models with a stronger level effect, we accommodate the shortcomings of CIR in the diffusion setting solely through the addition of a stochastic volatility factor. This is consistent with the suggestions of Dybvig (1989), and broadly in line with the theoretical specification of Longstaff and Schwartz (1992). Moreover, while providing a good fit to the actual interest rate dynamics, the implied process is much less erratic than those implied by alternative discrete-time models that incorporate both level and stochastic volatility effects. Consequently, our estimated multi-factor diffusion constitutes a more reliable framework for analysis of term structure issues and pricing of interest rate derivative claims via simulation techniques. At the same time we hasten to add that this is not a consequence of continuous-time versus discrete-time modeling per se. Indeed, we demonstrate that the EMM approach
also provides an excellent setting for analysis of alternative discrete-time models, and we propose an alternative class of candidate discrete-time models that may be able to serve as reasonable vehicles for interest rate derivatives pricing and hedging.

While the present paper, to our knowledge, represents the first consistent estimation of a stochastic volatility diffusion for interest rates, there is already a sizeable literature on estimation of interest rate diffusion models and EMM estimation in general. One recent approach argues that a priori restrictions of the functional form for the mean or the volatility of the diffusion are incredible, and focuses on estimation with a minimal set of auxiliary conditions. This is exemplified by Conley et al. (1995) who exploit the moment generating techniques of Hansen and Scheinkman (1995) to obtain nonparametric estimates of the drift, Ait-Sahalia (1995a) who estimates the volatility function nonparametrically, and Stanton (1995) who provides nonparametric discrete-time approximations to the drift and volatility functions. These approaches are, however, fundamentally one-factor models. They do not allow for estimation of an additional stochastic volatility factor which appears necessary in order to accommodate the strong conditional heteroskedasticity in short-term interest rates. Second, the EMM approach has been applied to both discrete-time stochastic volatility models, e.g., Gallant, Hsieh, and Tauchen (1994) and Gallant and Tauchen (1994), and continuous-time stock return and interest rate models, e.g., Gallant and Tauchen (1995a), and Pagan, Hall, and Martin (1995). In addition, Ghysels and Jasiak (1995) use it in a combined stochastic volatility and time deformation framework. Finally, Gouriéroux and Monfort (1994) estimate the Merton (1973) model using the related method of Indirect Inference developed by Gouriéroux, Monfort, and Renault (1993), while Engle and Lee (1994) apply this methodology to a stochastic volatility diffusion for equity returns.

Another advantage of our empirical work is the use of a comparatively long sample. We rely on weekly observations for the annualized yield on three-month U.S. Treasury Bills over January 1954 to April 1995. Given the near instability of both the mean and volatility dynamics of the series, the longer time span may be especially important.

The remainder of the paper is organized as follows: Section 2 provides an overview of the various steps involved in our implementation of the EMM estimation procedure. Section 3 discusses recent modeling strategies for the short rate involving both ARCH and level effects. Next, our data source is described in Section 4, while Section 5 reports on our estimation results for the general class of ARCH models. In Section 6, we describe how the ARCH specification is expanded into full-fledged SNP conditional density estimation. The final step of the EMM procedure involves GMM estimation using simulations from the assumed diffusion. This procedure is outlined in Section 7, while the empirical results are presented in Section 8. Section 9 concludes.
2. Model specification and estimation technique

A standard representation of the short-term interest rate dynamics is to let the short rate solve a special case of the following stochastic differential equation:

\[ dr_t = \mu(r_t)dt + \sigma_t(r_t)dW_t, \]  

(2)

where \( W_t \) is a standard Brownian motion, and \( \mu_t \) and \( \sigma_t \), the drift and diffusion functions, depend on the short rate and possibly other state variables. Most empirical work allow for these functions to depend on the short rate only, i.e., \( \mu_t(r_t) \equiv \mu(r_t) \) and \( \sigma_t(r_t) \equiv \sigma(r_t) \). Conley et al. (1995), Aït-Sahalia (1995a), and Stanton (1995) apply nonparametric techniques to estimate this system without specifying the mean and/or the diffusion function. Hence, their approach generalize the standard models that impose specific restrictions on both the mean and volatility function. For example, letting \( \mu(r_t) = \kappa(\mu - r_t) \) and \( \sigma(r_t) = \sigma_t \) yields the generalized CIR, or CKLS, model in Eq. (1), while, instead, imposing \( \sigma(r_t) = \sigma \) results in the Vasicek (1977) model.

An important impetus for the generalizations of the standard parametric models is the numerous statistical rejections of these, e.g., Aït-Sahalia (1995b) find none of the usual parametric models to be adequate. This suggests a serious void of useful representations of the short rate dynamics. However, the work of Dybvig (1989), Longstaff and Schwartz (1992), Brenner et al. (1994), Koedijk et al. (1994), and Gallant and Tauchen (1995a) is strongly suggestive of the directions in which the standard models fail. In particular, they point towards the importance of incorporating a stochastic volatility factor in the models. Moreover, while the nonparametric procedures may avoid unduly restricting the functional forms, their associated Markov assumption does not allow for the estimation of unobserved volatility factor dynamics. Consequently, we propose as an alternative an extension of the standard diffusion models that accommodates both a level effect and a stochastic volatility factor. Furthermore, we retain the continuous time specification throughout, and we conduct elaborate specification tests that should detect any serious mis specification due to the choice of functional form.

Our basic model extends the CKLS model by incorporating a stochastic volatility factor:

\[ dr_t = \kappa_1(\mu - r_t)dt + \sigma_t r_t^\gamma dW_{1,t}, \quad \gamma > 0, \]  

(3)

\[ d\log \sigma_t^2 = \kappa_2(\alpha - \log \sigma_t^2)dt + \xi dW_{2,t}, \]  

(4)

where \( W_{1,t} \) and \( W_{2,t} \) are independent standard Brownian Motion processes. The specification implies mean reversion of the interest rate level as well as the (log-)volatility. While this representation is ad hoc, it is inspired by the success of similar formulations for general financial time series, and as mentioned earlier, the adequacy of the resulting model will be tested thoroughly.
Maximum likelihood estimation is generally not feasible in the CKLS model, and the presence of an unobserved volatility exacerbates the problem.\(^1\) Instead, we utilize the Efficient Method of Moments (EMM) approach of Gallant and Tauchen (1994), which has certain advantages, even relative to maximum likelihood. The estimation strategy may be outlined as follows: in a first step, we obtain a semi-nonparametric estimate of the conditional density for the interest rate series, based on quasi-maximum likelihood (QML) estimation. When implemented with care, the resultant QML scores provide an operational representation of the salient features of the data. The second step then estimates the structural parameters by searching for the parameter vector that allows the assumed diffusion to mimic the essential characteristics of the data, as represented by the expected value of the QML scores, as closely as possible. Formally, the procedure follows the generalized method of moments (GMM) at this stage by minimizing a quadratic form in the expected scores. Since explicit calculation of these scores under the true model is infeasible, this step is implemented by simulation.\(^2\) Effectively, a Monte Carlo integration is performed to calculate the expectation of the scores for each candidate parameter vector. Because of the judicious choice of moments, the method can be shown, under regularity conditions, to be asymptotically equivalent to maximum likelihood which justifies the terminology 'Efficient Method of Moments'. Moreover, since the final step relies on GMM principles, the usual \(\chi^2\)-statistic for 'goodness-of-fit' of the overidentifying restrictions serves as an overall specification test, and, to the extent individual scores are associated with distinct characteristics of the data, the fit to each score indicates how well particular features of the data are accommodated.

The following sections document our implementation of the EMM procedure for the given continuous-time estimation problem. However, several of the individual steps are of interest in their own right. Specifically, our initial step involves fitting a semi-nonparametric model to the conditional density of the discrete observations. We follow Gallant and Tauchen (1995a) and rely on the SNP approach.\(^3\) The basic strategy is to utilize a squared Hermite polynomial expansion as an approximation to the conditional density. Usually, this involves employing a Gaussian density as the leading term and have the polynomials

\(^1\) However, it is feasible to apply maximum likelihood to discrete-time stochastic volatility models, if the state space for \(\sigma\) is severely restricted, e.g., \(\sigma\) may attain only a small set of finite values in the 'regime-switching' models of Hamilton (1990), Cai (1994), Gray (1995), Torous and Ball (1995), and Bekkaert, Hodrick, and Marshall (1995).

\(^2\) Consequently, the approach is in certain aspects similar to the Simulated Method of Moments (SMM) procedure of Duffie and Singleton (1993).

\(^3\) The theoretical justification for the approach is given by Gallant and Nychka (1987), building on work by Phillips (1983).
adapt to all non-Gaussian features of the process. However, as stressed by Gallant and Tauchen (1995a), the model’s ability to accommodate conditionally heteroskedastic data is much improved if the Hermite polynomial is relieved of this task. This suggests careful selection of the leading term, choosing the best possible candidate model for the interest rate process, while letting the Hermite expansion adapt to minor deviations from this leading term only. In this endeavor we may draw directly on recent evidence from parametric short rate modeling. Consequently, our implementation of EMM begins with a comprehensive analysis of alternative ARCH models that promise to provide a satisfactory fit to the conditionally heteroskedastic features of the process.

3. ARCH models of the short rate

It has long been acknowledged that discretely observed interest rate data display strong heteroskedasticity. The CKLS model captures an aspect of this feature by letting volatility be a function of the level of the interest rate. Alternatively, applications of GARCH models have documented very strong volatility persistence of the conditional heteroskedastic variety well known from most standard financial return series. While Kearns (1992) appears to be the first to model these features simultaneously, Brenner et al. (1994) and Koedijk et al. (1994) present the first maximum likelihood estimation results, and Longstaff and Schwartz (1992) rely on a related specification in their empirical work. Since these studies find both the level effect and the ARCH effects highly significant, they constitute a natural starting point for our search for a leading term for the SNP expansion.

Beyond requiring the candidate model to provide a good approximation to the conditionally heteroskedastic features of the data, however, we shall also be concerned with the stability properties of the models. A priori, we feel that the adopted specification should be strictly stationary. While it is entirely possible that there are structural shifts in the data-generating process over our sample – we explicitly test for this possibility – we are particularly insistent that the estimated process be nonexplosive. We simply do not know of any theoretical rationale for explosive interest rate series, and this type of instability will almost certainly render the implementation of the second, simulation-based, step of the EMM procedure infeasible.4 Perhaps, an even more compelling argument is that excessively erratic processes inherently are poor candidates for simulation-based analysis of the dynamics of the term structure or valuation of derivative securities. Since this class of models generally exclude closed form solutions, this

---

4 See Tauchen (1995) for a very readable introduction to the implementation of EMM.
is, indeed, a serious drawback. Our empirical results demonstrate that the issue of stability is an important practical concern for ARCH models fit to post-war U.S. interest rate data.

The initial ARCH model we fit is inspired by the Koedijk et al. (1994) specification, but resembles the models investigated by Brenner et al. (1994) as well. It takes the form:

\[ \Delta r_t = \phi_0 + \phi_1 r_{t-1} + \sum_{i=1}^{s-1} \phi_{i+1} \Delta r_{t-i} + r_{t-1}^y \sqrt{h_t} z_t, \]  
(5)

\[ h_t = \omega + \sum_{i=1}^{p} \beta_i h_{t-i} + \sum_{j=1}^{q} \alpha_j h_{t-j} z_{t-j}^2, \]  
(6)

where \( z_t \) is assumed i.i.d. \( \text{N}(0, 1) \). The AR terms in the conditional mean equation are required to accommodate the very high degree of persistence in the interest rate series, and our empirical results confirm that the series appear close to integrated. We use the 'Dickey–Fuller' representation, i.e., the differenced rates, in Eq. (5) purely for reasons of numerical stability, although the gain seems to be minor. Eq. (6) specifies the standard GARCH(\( p, q \)) model of Bollerslev (1986), except that the lagged squared innovations from Eq. (5), which ordinarily drive the GARCH dynamics, implicitly have been normalized by \( r_{t-1}^{2y} \) before entering Eq. (6). This constitutes the main difference from Brenner et al. (1994), and it is adopted to allow separate identification of the distinct heteroskedastic effects. In terms of the ability to approximate the actual volatility dynamics the two specifications appear indistinguishable. We also fit this model under the assumption that \( z_t \) follows a standardized Student-\( t \) distribution. However, the distribution governing \( z_t \) does not affect the conditionally heteroskedastic characteristics of the series, and the SNP expansion, which we explore later, is particularly well suited to handle the non-Gaussian features of the conditional density. In light of this and the quasi-maximum likelihood results of Bollerslev and Wooldridge (1992), we are more comfortable with the inference from the Gaussian version, although there is clear evidence of heavy tails in the conditional distributions.\(^5\)

Since we have no a priori reason to rule out asymmetric effects in the volatility dynamics, reminiscent of the well-documented leverage effect for equity return indices, we also explore a Level-AR(s)-EGARCH(\( p, q \)) which is obtained by substituting the EGARCH(\( p, q \)) specification of Nelson (1991) for Eq. (6). Our

\(^5\) Indeed, if the Student-\( t \) assumption is invalid, the maximum likelihood estimator is no longer consistent, while it retains consistency under the normality assumption, even in case of misspecification of the conditional density for \( z_t \). However, see Newey and Steigerwald (1994) for an approach that allows consistent estimation of a subset of the parameter by quasi-maximum likelihood under Student-\( t \) distributional assumptions.
specification of the EGARCH-dynamics is given by
\[
\log h_t = \omega + \sum_{i=1}^{p} \beta_i \log h_{t-i} + (1 + \alpha_1 L + \ldots + \alpha_q L^q) \\
\times [\theta_1 z_{t-1} + \theta_2 (b(z_{t-1}) - \sqrt{2/\pi})],
\]
where \( L \) denotes the lag-operator, \( z \) is standard Gaussian, and \( b(z) \) denotes a two times differentiable approximation to the absolute value function, \(|z|\).\(^6\)

Again, it is straightforward, but not necessarily beneficial, to allow for conditionally heavy tails by altering the distributional assumption on \( z \). In this case, one will generally have \( \mathbb{E}[g(z)] = \mathbb{E}[b(z)] - (2/\pi)^{1/2} \neq 0 \), but \( \theta_2 \) still measures the impact of an innovation in \( b(z) \), because \( \theta_2 \mathbb{E}[g(z)] \) is offset one-for-one in the estimated constant, \( \omega \), while the estimate of \( \theta_2 \) is unaffected.

The most obvious distinction between the EGARCH and GARCH models is the incorporation of the asymmetric volatility effect, captured by the coefficient, \( \theta_1 \). Nonetheless, a more subtle issue turns out to be important for the practical assessment of these models. The conditions for strict stationarity and covariance stationarity of the EGARCH volatility process are identical, while the GARCH model allows for strictly stationary processes that are not covariance-stationary.\(^7\) By itself, this feature is not critical, and GARCH estimates often place the model in the strictly, but covariance-nonstationary region of the parameter space, in particular when applied to interest rate series; see, e.g., the studies cited by Bollerslev, Chou, and Kroner (1992). However, in the Level-GARCH models there is an additional interaction via the interest rate level effect and the near integration in the mean dynamics which will induce further destabilizing elements into the interest rate dynamics. While theoretical parameter constraints that ensure strict stationarity of general Level-AR(1)GARCH models have not yet been derived, it is evident from the results of Broze, Scaillet, and Zakoïan (1995) that \( \gamma \leq 1 \) is a necessary condition. This condition is, in fact, routinely violated for Level-GARCH specifications,\(^8\) and, even when \( \gamma \) falls within the potentially stationary region, the GARCH models appear more unstable than the corresponding EGARCH variants. We conjecture that this is due to the fact that the point estimates governing the EGARCH dynamics invariably are

\(^6\) Specifically, \( b(z) = |z| \) for \( |z| \geq \pi/2K \) and \( b(z) = (\pi/2 - \cos(Kz))/K \) for \( |z| < \pi/2K \). We set \( K = 100 \), but any large value of \( K \) will ensure an extremely close and twice differentiable approximation to \( |z| \). This is similar to the approximation proposed by Gallant and Tauchen (1995b), and is necessary to circumvent convergence problems induced by the nondifferentiability of the EGARCH variance at the origin. We thank Jim Hamilton for the observation that a smooth approximation to the variance function may be helpful in this context.

\(^7\) See Nelson (1990), Nelson and Cao (1992), and Bollerslev and Engle (1993) for a discussion of these issues.

\(^8\) This point is also emphasized by Torous and Ball (1995).
consistent with strict and covariance stationarity, at least for the standard volatility models, i.e., in the absence of a level effect. Given the lack of theoretical criteria for stationarity we are nonetheless forced to explore the issue of stability of the estimated models by simulation methods. The empirical results are presented after the description of our data sources.

4. Data sources

Our empirical work uses weekly observations of the annualized yield on U.S. Treasury Bills with three months to maturity. The series was constructed from a daily series available from the Federal Reserve and corresponds to the three-month T-Bill rates included in the weekly H.15 release. Excerpts of the series are also included in Table 1.35 of the Federal Reserve Bulletin. The rates are calculated as unweighted averages of closing bid rates quoted by at least five dealers in the secondary market. The rates are posted on a bank discount basis, but are converted into continuously compounded yields prior to analysis.

Due to some missing data points, possible holiday and week-day effects, and other potential problems associated with market microstructure effects, we limit our analysis to the weekly frequency. Wednesdays have the least number of missing observations, so we use the reported rate for this weekday. When a Wednesday rate is missing, we use the Tuesday rate. This ensures a valid observation for all weeks over the full January 1954 to April 1995 sample, and results in a total of 2155 data points. This series represents the longest weekly set of observations on the U.S. risk-free rate that we are aware of. The full series is displayed in Fig. 1. The use of a long sample may be particularly important in this context in order to facilitate the identification of the mean and volatility dynamics that are borderline stationary.

5. Empirical assessment of ARCH models for the short rate

This section presents evidence on the fit of ARCH models to the short-term interest rate process. ARCH models should be interpreted in its broadest sense, meaning all specifications that allow for potentially time-varying volatility driven by past observables. Thus, this class covers, as special cases, almost all discrete-time models that have been proposed for the short rate. The ARCH

---

9 Specifically, the daily data were obtained from an Internet-site supported by the Federal Reserve System. We thank Ken Kroner for providing us with a shorter data set that was used in an initial pilot-study.

10 The use of Treasury Bill yield as the economically relevant short-term interest rate is not entirely uncontroversial, however. See, e.g., Duffee (1995) for a discussion of the evidence regarding time-varying segmentation between the T-Bill market and the main markets for near riskless private debt.
Fig. 1. Weekly U.S. T-Bill rates, three-month maturity, January 1954 to April 1995.

class is of particular interest because of the recent suggestions that Level-GARCH models provide a good framework for pricing and hedging of interest rate derivative assets. The results reported below are contrary to this view, but suggest that modeling efforts based on the related Level-EGARCH model may provide a useful alternative. The findings are, furthermore, indicative of the appropriate choice for the leading term in the SNP modeling strategy presented in the following section.

Our two basic models of the Level-AR(s)-(E)GARCH\((p,q)\) variety, given by Eqs. (5) and (6), or (5) and (7), respectively, were estimated for various choices of the lag lengths \(s, p, \) and \(q\), and for \(z\), being distributed standard Gaussian and Student-\(t\). The quality of the fit across the various models was evaluated using standard information criteria, i.e. the Akaike (AIC), Bayesian (BIC), and Hannan–Quinn (HQC) criteria, as well as specification tests.

To establish a benchmark we present the fit of the Gaussian models that were selected on the basis of the HQC.\(^{11}\) The chosen Gaussian GARCH specification

\(^{11}\)In general, we expect the AIC to be the more liberal criterion and the BIC to be the more conservative. Thus, AIC will tend to select a richer and BIC a more parsimonious model parameterization than HQC. We found the HQC closer to BIC than AIC, with the latter showing signs of overparameterization. See, e.g., Lütkepohl (1991) for an account of model selection criteria.
is a Level-AR(1)-GARCH(1,2):
\[
    r_t = 0.0138 + 0.9979 r_{t-1} + (r_{t-1})^{0.5635} h_{t-1}^{1/2} z_t,
\]
\[
    (0.0050) \quad (0.0012) \quad (0.1471)
\]
\[
    h_t = 0.0002 + 0.8850 h_{t-1} + 0.2576 h_{t-1} z_{t-1}^2 - 0.1304 h_{t-2} z_{t-2}^2,
\]
\[
    (0.0001) \quad (0.0134) \quad (0.0602) \quad (0.0592)
\]

while the chosen Gaussian EGARCH is a Level-AR(2)-EGARCH(1,2):
\[
    r_t = 0.0111 + 0.9981 r_{t-1} + 0.0767 \Delta r_{t-1} + (r_{t-1})^{0.6395} h_{t-1}^{1/2} z_t,
\]
\[
    (0.0054) \quad (0.0013) \quad (0.0280) \quad (0.2072)
\]
\[
    \log h_t = -2.864 + 0.9818 \log h_{t-1} + (1 - 0.2335 L - 0.2287 L^2)
\]
\[
    (0.3586) \quad (0.0069) \quad (0.1674) \quad (0.1406)
\]
\[
    \times \left[ -0.0141 z_{t-1} + 0.4546(\log(z_{t-1}) - (2/\pi)^{1/2}) \right],
\]
\[
    (0.0388) \quad (0.0842)
\]

where robust standard errors are given below the parameter estimates. Note that the coefficient on \( r_{t-1} \) corresponds to \( 1 - \phi_1 \). The near unit root in the AR-polynomial is consistent with the highly persistent mean dynamics documented in prior studies. Nonetheless, our long sample helps accumulate some, albeit weak, evidence against the hypothesis of a unit root due to the small standard errors on the AR coefficient. Note also that \( E[r_t] = -\phi_0/\phi_1 \), which translates into an implied estimate of the unconditional mean of 6.66% in the Level-GARCH model and 5.97% in the Level-EGARCH. Further, we observe that the level coefficient, \( \gamma \), is significantly below unity and not significantly different from \( \frac{1}{2} \) in both models. The remaining features of the volatility process are governed by Eqs. (6) and (7). For the GARCH model, the relevant volatility persistence measure is \( \beta_1 + \alpha_1 + \alpha_2 = 1.0086 \), which indicates that the short rate process fails to be covariance-stationary and has an infinite (unconditional) variance. Hence, the results based on our long weekly sample are consistent with the findings of extreme volatility persistence in the earlier literature. In contrast, for standard EGARCH models the volatility persistence is governed by the \( \beta \) coefficients alone, and we find \( \beta_1 = 0.9819 \). This is consistent with strict as well as covariance stationarity in the standard models, and although the persistence is high, it is suggestive of a more smooth interest rate process than that implied by the Level-GARCH estimates. Moreover, \( 1 + \alpha_1 + \alpha_2 = 0.6395 \) measures the short-run sensitivity of the EGARCH process to innovations in the standardized interest rate residual function, \( \theta_1 z_t + \theta_2 \left[b(z_t) - (2/\pi)^{1/2}\right] \). The insignificant \( \theta_1 \) and positive \( \theta_2 \) verify the existence of strong conditional heteroskedasticity in the interest rate dynamics, but there is no evidence of asymmetric volatility responses beyond the level effect.

In order to document the robustness of these findings, we summarize our estimation results for the models selected by each of the three information
criteria within each category: the Level-GARCH and Level-EGARCH, and the Gaussian and Student-\(t\) distributional assumptions. Tables 1 and 2 report the results for the Level-GARCH and Level-EGARCH, respectively. It is apparent that all main qualitative features of the fitted models are similar. The only noteworthy points are, first, the higher level effect for EGARCH compared to GARCH which may reflect, or counteract, the associated decline in volatility persistence for the EGARCH models, and, second, the drop in the level coefficient, \(\gamma\), when moving from the Gaussian to the Student-\(t\) models. The latter may signify some complementarity between the parameters governing the conditional tails of \(z_t\) and the strength of the level effect. In other words, the extremely fat tails implied by the low values of the tail index, \(v\), in the Student-\(t\) models may induce a drop in the other parameters determining the sensitivity of the volatility process to interest rate innovations. Whether this is warranted or stems from unduly restrictive distributional assumptions is not clear, but it does suggest that some account should be taken of the conditionally heavy tails. The SNP strategy in the following section allows for this without imposing specific distributional assumptions on the standardized innovations.

Since these models do not represent our final specification and our findings generally are consistent with those reported elsewhere, e.g., Brenner et al. (1994), we only briefly summarize our specification tests. Ljung–Box tests for autocorrelation in the squared residuals confirm that all the models successfully remove

---

**Table 1**

Estimates of Level-GARCH models for weekly observations on U.S. three-month Treasury Bills

<table>
<thead>
<tr>
<th>Gaussian model</th>
<th>(-\phi_0/\phi_1)</th>
<th>(1 + \phi_1)</th>
<th>(\sum(\beta_i + \alpha_i))</th>
<th>(\gamma)</th>
<th>(\nu)</th>
</tr>
</thead>
<tbody>
<tr>
<td>HQC: (1, 1, 2)</td>
<td>6.56</td>
<td>0.9979</td>
<td>1.0086</td>
<td>0.5635</td>
<td></td>
</tr>
<tr>
<td>BIC: (1, 1, 2)</td>
<td>6.56</td>
<td>0.9979</td>
<td>1.0086</td>
<td>0.5635</td>
<td></td>
</tr>
<tr>
<td>AIC: (6, 1, 3)</td>
<td>5.88</td>
<td>0.9975</td>
<td>1.0057</td>
<td>0.5514</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Student-(t) model</th>
<th>(-\phi_0/\phi_1)</th>
<th>(1 + \phi_1)</th>
<th>(\sum(\beta_i + \alpha_i))</th>
<th>(\gamma)</th>
<th>(\nu)</th>
</tr>
</thead>
<tbody>
<tr>
<td>HQC: (1, 1, 3)</td>
<td>6.50</td>
<td>0.9984</td>
<td>1.0106</td>
<td>0.4138</td>
<td>5.00</td>
</tr>
<tr>
<td>BIC: (1, 1, 1)</td>
<td>6.20</td>
<td>0.9981</td>
<td>1.0166</td>
<td>0.4004</td>
<td>4.87</td>
</tr>
<tr>
<td>AIC: (6, 1, 3)</td>
<td>5.54</td>
<td>0.9979</td>
<td>1.0102</td>
<td>0.4287</td>
<td>5.04</td>
</tr>
</tbody>
</table>

The Level-GARCH(\(s, p, q\)) model is:

\[
\begin{align*}
    r_t &= \phi_0 + (1 + \phi_1)r_{t-1} + \sum_{i=1}^{s-1} \phi_{1+i} \Delta r_{t-i} + \epsilon_t^2 h_t^{1/2} z_{t-j} \\
    h_t &= \omega + \sum_{i=1}^{p} \beta_i h_{t-i} + \sum_{j=1}^{q} \alpha_j h_{t-j} z_{t-j}^2
\end{align*}
\]

The models were estimated by Maximum Likelihood, using the annualized yield on the weekly T-Bills with 90 days to maturity from January 1954 to April 1995. Within each class the indicated models were selected by the information criteria, Hannan–Quinn (HQC), Bayesian (BIC), and Akaike (AIC). The selected models are indexed by the triple \((s, p, q)\).
Table 2
Estimates of Level-EGARCH models for weekly observations on U.S. three-month Treasury Bills

<table>
<thead>
<tr>
<th>Gaussian model</th>
<th>$-\phi_0/\phi_1$</th>
<th>$1 + \phi_1$</th>
<th>$\sum \beta_i$</th>
<th>$\gamma$</th>
<th>$1 + \sum \alpha_j$</th>
<th>$\theta_1$</th>
<th>$\theta_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>HQC: (2,1,2)</td>
<td>5.97</td>
<td>0.9981</td>
<td>0.9818</td>
<td>0.6395</td>
<td>0.5378</td>
<td>-0.0141</td>
<td>0.4546</td>
</tr>
<tr>
<td>BIC: (2,1,1)</td>
<td>6.01</td>
<td>0.9980</td>
<td>0.9780</td>
<td>0.6938</td>
<td>0.6245</td>
<td>-0.0208</td>
<td>0.4326</td>
</tr>
<tr>
<td>AIC: (6,1,2)</td>
<td>5.47</td>
<td>0.9974</td>
<td>0.9830</td>
<td>0.6350</td>
<td>0.4990</td>
<td>-0.0098</td>
<td>0.4577</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Student-t$_\nu$ model</th>
<th>$-\phi_0/\phi_1$</th>
<th>$1 + \phi_1$</th>
<th>$\sum \beta_i$</th>
<th>$\gamma$</th>
<th>$1 + \sum \alpha_j$</th>
<th>$\theta_1$</th>
<th>$\theta_2$</th>
<th>$\nu$</th>
</tr>
</thead>
<tbody>
<tr>
<td>HQC: (1,1,2)</td>
<td>6.58</td>
<td>0.9985</td>
<td>0.9921</td>
<td>0.5275</td>
<td>0.5401</td>
<td>-0.0237</td>
<td>0.3863</td>
<td>5.24</td>
</tr>
<tr>
<td>BIC: (1,1,1)</td>
<td>6.52</td>
<td>0.9985</td>
<td>0.9908</td>
<td>0.5590</td>
<td>0.6162</td>
<td>-0.0286</td>
<td>0.3705</td>
<td>5.07</td>
</tr>
<tr>
<td>AIC: (6,1,2)</td>
<td>5.59</td>
<td>0.9981</td>
<td>0.9908</td>
<td>0.5380</td>
<td>0.5284</td>
<td>-0.0268</td>
<td>0.3861</td>
<td>5.22</td>
</tr>
</tbody>
</table>

The Level-EGARCH($s,p,q$) model is:

$$r_t = \phi_0 + (1 + \phi_1)r_{t-1} + \sum_{i=1}^{s-1} \beta_i \Delta r_{t-i} + \sum_{j=1}^{p} \alpha_j h_{t-j}^{1/2} z_t$$

$$\log h_t = \omega + \sum_{i=1}^{s} \beta_i \log h_{t-i} + \left(1 + \sum_{j=1}^{q} \alpha_j L^j \right)$$

$$\times [\theta_1 z_{t-1} + \theta_2 (b(z_{t-1}) - (2/\pi)^{1/2})]$$

The models were estimated by Maximum Likelihood, using the annualized yield on the weekly T-Bills with 90 days to maturity from January 1954 to April 1995. Within each class the indicated models were selected by the information criteria, Hannan–Quinn (HQC), Bayesian (BIC), and Akaike (AIC). The selected models are indexed by the triple ($s,p,q$).

The systematic second-order dependency in the series. However, the corresponding tests for the residuals themselves reveal rather strong evidence of autocorrelation, especially at higher-order lags. This is symptomatic of the strong drift in the interest rates over the sample period, and we are not aware of any models that have provided a satisfactory fit to this aspect of the short rate series. Further, there is, of course, overwhelming evidence of excess kurtosis in the standardized residuals from the Gaussian models, but, as mentioned, we address this feature in later sections. Finally, we conducted a string of conditional moment tests for model misspecification based on the robust procedures of Wooldridge (1990).12 First, we found that the tests reinforce the above points. Second, we investigated the period covering the so-called ‘Monetary Experiment’ of the Federal Reserve System between October 1979 and October 1982. Our results are consistent with such an erratic episode which again confirms results from Brenner et al. (1994) and also Torous and Ball (1995). Indeed, we show below that the problem with some of these models is their tendency to induce episodes far more volatile than this event, not the opposite. Finally, we

---

12 See Brenner et al. for an in-depth exposition of these tests. Most significantly, they are robust to misspecified distributional assumptions.
detect weak evidence against the symmetric volatility response functions implied by the GARCH models.

At this point one may be tempted to label the ARCH evidence presented above a relative success. The models appear to provide an adequate fit to the highly erratic short rate series, and they establish convincingly that both standard ARCH effects and a level effect are necessary to account for the volatility dynamics. Furthermore, the qualitative conclusions appear robust to minor variations in specification, and the switch from a GARCH to an EGARCH representation appears to produce only minor changes in results so the choice between the models may seem arbitrary. This is, however, rather deceiving. The following simulation evidence reveals a remarkable difference in the stability properties of the two model classes.

We fixed parameter values at the point estimates and constructed a random 100,000 long sample of realizations from each candidate model. This sample was then broken down into consecutive subsamples of 1000 observations. This corresponds roughly to 100 twenty-year samples constructed randomly according to the particular data-generating process for each model. Clearly, there is some dependency between two consecutive samples, but if the null hypothesis that the models are valid representations of the short rate, then this correlation should be minor, and samples that are not adjacent ought to be close to independent, so the 100 consecutive twenty-year samples should be sufficient to reveal the basic features of the models. Invariably, the simulated Level-GARCH models display extremely erratic behavior. Fig. 2 plots the maximum realizations attained by the interest rate process over each twenty-year simulated sample for two Level-AR-GARCH models selected by the HQC. The Gaussian model is the Level-AR(2)-GARCH(1, 2) for which parameter estimates were provided above. To obtain an acceptable scaling, the log-interest rates are reported in the figure. The overall maximum for this series is in excess of 66,000%, followed by 51,117% and 18,603%, and more than 25 samples have rates in excess of 100%. In general, the evidence for the Level-AR-GARCH-t models is even more disturbing. Specifically, for the model selected by HQC, Level-AR(1)-GARCH(1, 3)-t, we obtain an overall maximum of over 22.5 million percent, followed by other sample maxima in excess of 11, 3.1, and 2.1 million percent. In total, 17 samples attain rates of over 1,000%. Fig. 2 illustrates the erratic behavior of the implied process. For both these models the maximum rate exceeds 25% in about half the samples.

The results for the Level-EGARCH models stand in sharp contrast to the above. Fig. 3 presents the same evidence regarding the maximum short rates obtained over 100 consecutive simulated twenty-year samples, but generated by the models selected by HQC within the EGARCH class, as reported in Table 1, so the Gaussian Level-EGARCH model corresponds to the one whose parameter estimates were provided earlier. For the simulated EGARCH models the rates never exceed 75%, and only a single sample has rates in excess of 37%. The
Fig. 2. Maximum interest rates over consecutive twenty-year periods from simulated Level-AR-GARCH models.

Fig. 3. Maximum interest rates over consecutive twenty-year periods from simulated Level-AR-EGARCH models.
number of samples with maximum rates in excess of 25% is 7 and 4, respectively, for the Gaussian and Student-t models. In summary, the estimated EGARCH models produce sample realizations that generally conform to a priori perceptions of a short rate process. By inspection of the range of values attained over the subsamples (not reported), it is evident that these models still generate highly volatile series, and quite readily accommodate episodes such as the 'Monetary Experiment' – indeed, they predict that these may occur with nonnegligible frequency. However, the series also produce long periods of relative tranquility, and they do not tend to generate the near-explosive behavior that occurs with alarming frequency in the corresponding GARCH models.

Our simulation results are robust to alterations in specification. We have documented similar, or even worse, behavior for all relevant Level-GARCH models reported in Brenner et al. (1994) and Koedijk et al. (1994). Dynamic stability is crucial for implementation of the second step of the EMM procedure since it avoids having the parameter estimates of the structural (diffusion) model stray outside the stationary and ergodic region. These issues are discussed extensively in Tauchen (1995). In conclusion, we do not recommend Level-GARCH models for analysis of short rate dynamics. Fortunately, the related Level-EGARCH class appears to perform satisfactorily, and we use the Gaussian Level-EGARCH as the leading term for our SNP expansion of the conditional density for the short rate process investigated in the following section.

6. Expanding into SNP models

As discussed in Section 2, the EMM estimation principle relies on a two-step procedure. The first step involves specifying and estimating an auxiliary time series model, called the score generator, and is roughly comparable to the choice of moment conditions in an ordinary GMM estimation. However, while the moment conditions for the typical GMM (or SMM) estimation are chosen more or less arbitrarily, the EMM procedure offers considerable guidance on the choice of moment conditions.

Ideally, we would, of course, prefer to estimate the parameters of the structural model by maximum likelihood. By the prediction error decomposition we can write the log-likelihood function for a time series of the short-term interest of length \( T \), \( \{r_t\} \), as

\[
\log \mathcal{L}(r_1, \ldots, r_T; \rho) = \sum_{t=1}^{T} \log p(r_t | F_{t-1}; \rho),
\]

where \( p(r_t | F_{t-1}; \rho) \) is the conditional density function of \( r_t \) given the information set, \( F_{t-1} \), generated by \( \{r_{t-j}, j > 0\} \). In theory, the functional form of this
conditional density can be derived from the drift and diffusion coefficients in (3) and (4), but computational problems render the approach infeasible.\textsuperscript{13}

Fortunately, Gallant and Long (1995) show that careful implementation of the EMM scheme can achieve the same asymptotic efficiency as maximum likelihood while retaining computational tractability. The key condition is that the score function of the auxiliary model asymptotically spans the score function of the true (but unknown) conditional density function for \( r_t \). They further show that the score function of the SNP density of Gallant and Nychka (1987) spans the score of most relevant distributions, provided the number of terms in the SNP expansion, \( K \), grows with the sample size such that \( K \to \infty \) when \( n \to \infty \) a.s. As long as this requirement is fulfilled, any data-dependent scheme can be used to choose \( K \), including information criteria such as the AIC and BIC. Thus, the SNP density can model any aspect of the conditional heterogeneity of \( \{r_t\} \), but the finite-sample properties are improved if the Gaussian leading term provides a good approximation to the conditional mean and variance, typically represented by an AR-ARCH model. The theoretical results developed in the earlier EMM/SNP literature are based on Markovian score generators, which means that the leading term has to be pure ARCH. However, Gallant and Long (1995) show that non-Markovian score generators are applicable as well. This is important because the conditional heteroskedasticity of most financial time series is well captured by low-order GARCH\((p,q)\) or EGARCH\((p,q)\) models, with \( p \) equal to 1 or 2, while a corresponding fit for the pure ARCH\((q)\) usually requires a very high-order model. Although the small-sample properties of the SNP/EMM estimator have not yet been investigated, we conjecture, based on the evidence for the ordinary GMM estimator in the stochastic volatility context, e.g., Andersen and Sørensen (1996), that it is important to conserve on the number of elements in the score generator (the moments conditions of EMM).

In view of this as well as the earlier empirical results we use an SNP density with a Gaussian Level-AR\((s)\)-EGARCH\((p,q)\) leading term. To facilitate discussion, let \( x_t \) denote a vector containing \( M \) lags of \( r_t \), and let \( K = \max(K_x, K_z) \). Our score generator (auxiliary model) is then given by

\[
    f_K(r_t | x_t; \eta) = \frac{[P_K(z_t, x_t)]^2 \phi(z_t)}{\int_{-\infty}^{\infty} [P_K(u, x_t)]^2 \phi(u) \, du} \frac{1}{r_t^2 \sqrt{h_t}},
\]

\[
    (9)
\]

\textsuperscript{13} First, while the joint conditional density function for \( \{r_t, \sigma_t\} \) is known to solve the Kolmogorov forward and backward equations, cf. Lo (1988), closed from expressions for these partial differential equations (PDE's) are only known for special cases such as the CIR model. Second, even if a solution could be obtained (e.g., by solving the PDE numerically) we still need to integrate the joint density function over the unobserved \( \sigma_t \) in order to obtain the marginal likelihood for \( \{r_t\} \) in Eq. (8). See, e.g., Melino (1994) for an extensive discussion of these issues.
where $\phi(\cdot)$ is the standard normal density, and

$$z_t = (r_t - \mu_t)/(r_{t-1}^{\eta} \sqrt{h_t}),$$

$$\mu_t = \phi_0 + (1 + \phi_1) r_{t-1} + \sum_{i=1}^{s-1} \phi_{i+1} \Delta r_{t-i},$$

$$P_k(z,x_t) = \sum_{i=0}^{K_z} a_i(x_t) z^i = \sum_{i=0}^{K_z} \left( \sum_{j=0}^{K_x} a_{ij} x_t^j \right) z^i,$$  \quad a_{00} = 1,$$  \quad a_{00} = 1,$$

with $h_t$ given by Eq. (7), and $\eta$ is a vector containing the parameters of the Level-AR(s)-EGARCH($p,q$) term as well as the SNP polynomial (12). The condition $a_{00} = 1$ is imposed for identification purposes. If $M = 1$, (12) is a $K_x$-th order polynomial in $z$, whose coefficients $a_i(x_t)$ are represented by a polynomial of degree $K_x$ in the scalar $x_t$.$^{14,15}$

Full asymptotic efficiency relative to maximum likelihood is not the only advantage of using the SNP model as score generator. Compared to, say, using only the Gaussian leading term as score generator, the structural model (diffusion) is required to fit the entire conditional distribution of the time series, not just the first and second conditional moments. This more stringent requirement increases the power of the EMM specification tests against misspecification of the structural model.

Tables 3 and 4 contain estimation results for a number of different SNP models. In the Hermite polynomial we consider various combinations of $K_z = 4, 6$, or 8, $K_x = 0$ or 1, and $M = 1$. We vary the order of the AR polynomial from 1 to 5, while the EGARCH($p,q$) part is either (1,1) or (1,2). This selection includes most, if not all, relevant models for the three-month T-Bill series. To make the results comparable, the first five observations were set aside as lags in the AR(s) part and the Hermite polynomial, leaving an effective sample size of 2150. For each model, we report the (maximized) value of the log-likelihood function, as well as the AIC, BIC, and HQC. For linear ARMA models BIC is known to be a consistent information criterion, i.e., it selects the correct model asymptotically, while the AIC always overfits the model. The corresponding theory for semi-nonparametric models is still being developed, and results have yet to be extended to SNP density estimation for dependent

$^{14}$ For $M > 1$, the expression $x^i$ becomes more complex. Since we only consider $M = 1$ in the sequel, we defer to Gallant and Tauchen (1989) for details on the general case.

$^{15}$ During implementation, we rewrite the polynomial in $z$ using orthogonal Hermite polynomials, see Fenton and Gallant (1995) or Abramowitch and Stegun (1972, Ch. 22) for details. This changes neither the order of the polynomial nor the functional form of the SNP model, but enhances overall numerical stability.
Table 3
Hermite ($K_z, K_x$) SNP models with a Gaussian Level-AR($s$)-EGARCH($p, q$) leading term

<table>
<thead>
<tr>
<th>s</th>
<th>p</th>
<th>q</th>
<th>$K_z$</th>
<th>$K_x$</th>
<th>Log lik.</th>
<th>AIC</th>
<th>BIC</th>
<th>HQC</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>4</td>
<td>0</td>
<td>1053.41</td>
<td>-0.4844</td>
<td>-0.4685</td>
<td>-0.4786</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>1</td>
<td>4</td>
<td>0</td>
<td>1055.41</td>
<td>-0.4848</td>
<td>-0.4677</td>
<td>-0.4786</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>1</td>
<td>4</td>
<td>0</td>
<td>1055.68</td>
<td>-0.4845</td>
<td>-0.4660</td>
<td>-0.4777</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>1</td>
<td>4</td>
<td>0</td>
<td>1056.27</td>
<td>-0.4843</td>
<td>-0.4645</td>
<td>-0.4771</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>1</td>
<td>4</td>
<td>0</td>
<td>1059.30</td>
<td>-0.4853</td>
<td>-0.4641</td>
<td>-0.4775</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>6</td>
<td>0</td>
<td>1060.39</td>
<td>-0.4867</td>
<td>-0.4682</td>
<td>-0.4799</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>1</td>
<td>6</td>
<td>0</td>
<td>1064.46</td>
<td>-0.4881</td>
<td>-0.4683</td>
<td>-0.4809</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>1</td>
<td>6</td>
<td>0</td>
<td>1064.65</td>
<td>-0.4877</td>
<td>-0.4666</td>
<td>-0.4800</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>1</td>
<td>6</td>
<td>0</td>
<td>1065.70</td>
<td>-0.4878</td>
<td>-0.4653</td>
<td>-0.4796</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>1</td>
<td>6</td>
<td>0</td>
<td>1069.97</td>
<td>-0.4893</td>
<td>-0.4655</td>
<td>-0.4806</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>8</td>
<td>0</td>
<td>1062.91</td>
<td>-0.4869</td>
<td>-0.4658</td>
<td>-0.4792</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>1</td>
<td>8</td>
<td>0</td>
<td>1064.75</td>
<td>-0.4873</td>
<td>-0.4649</td>
<td>-0.4791</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>1</td>
<td>8</td>
<td>0</td>
<td>1065.09</td>
<td>-0.4870</td>
<td>-0.4633</td>
<td>-0.4783</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>1</td>
<td>8</td>
<td>0</td>
<td>1066.57</td>
<td>-0.4872</td>
<td>-0.4622</td>
<td>-0.4781</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>1</td>
<td>8</td>
<td>0</td>
<td>1071.23</td>
<td>-0.4889</td>
<td>-0.4626</td>
<td>-0.4793</td>
</tr>
</tbody>
</table>

The model is given by:

$$f_k(r_t | x_t; \eta) = \frac{[P_K(z_t, x_t)]^2 \phi(z_t)}{\left[\sum_{u} [P_K(u, x_t)]^2 \phi(u) du \right] r_{t-1} \sqrt{h_t}},$$

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

$$z_t = \frac{r_t - \mu_t}{r_{t-1} \sqrt{h_t}},$$

$$\mu_t = \phi_0 + (1 + \phi_1)r_{t-1} + \sum_{i=1}^{s-1} \phi_{i+1} \Delta r_{t-i},$$

$$P_K(z, x) = \sum_{i=0}^{K_z} a_1(z) z^i = \sum_{i=0}^{K_x} \left( \sum_{\beta=0}^{K_x} a_\beta x_\beta^i \right) z^i, \quad a_{00} = 1,$$

$$\log h_t = \omega + \sum_{i=1}^{p} \beta_i \log h_{t-i} + (1 + a_1 L + \cdots + a_q L^q)[\theta_1 z_{t-1} + \theta_2 (b(z_{t-1}) - \sqrt{2/\pi})],$$

with $b(z)$ representing a close, but twice differentiable approximation to the absolute value function defined in Footnote 6, $x_t$ denotes a set of lagged values of $r_t$, and $\eta$ is a vector containing all unknown parameters.

variables (time series).\(^{16}\) Nonetheless, the results of Eastwood (1991) suggest that the AIC is optimal in the SNP setting. However, our experience with EMM shows that overfitting in the score generator may lead to numerical problems, probably because we then attempt to fit the model to noise in the data. Our

\(^{16}\) Andrews (1991), Eastwood (1991), and Gallant and Souza (1991) analyze the asymptotic properties of semi-nonparametric estimators and data-dependent schemes for selection of the number of terms to include.
Table 4
Hermite $(K_x, K_x)$ SNP models with a Gaussian Level-AR(s)-EGARCH$(p, q)$ leading term

<table>
<thead>
<tr>
<th>$s$</th>
<th>$p$</th>
<th>$q$</th>
<th>$K_x$</th>
<th>$K_x$</th>
<th>Log Lik.</th>
<th>AIC</th>
<th>BIC</th>
<th>HQC</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>4</td>
<td>1</td>
<td>1058.65</td>
<td>-0.4845</td>
<td>-0.4621</td>
<td>-0.4763</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>1</td>
<td>4</td>
<td>1</td>
<td>1060.83</td>
<td>-0.4850</td>
<td>-0.4613</td>
<td>-0.4763</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>1</td>
<td>4</td>
<td>1</td>
<td>1061.05</td>
<td>-0.4847</td>
<td>-0.4596</td>
<td>-0.4755</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>1</td>
<td>4</td>
<td>1</td>
<td>1061.48</td>
<td>-0.4844</td>
<td>-0.4580</td>
<td>-0.4748</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>1</td>
<td>4</td>
<td>1</td>
<td>1064.63</td>
<td>-0.4854</td>
<td>-0.4577</td>
<td>-0.4753</td>
</tr>
</tbody>
</table>

| 1   | 1   | 1   | 6     | 1     | 1069.52  | -0.4877 | -0.4600 | -0.4775 |
| 2   | 1   | 1   | 6     | 1     | 1072.84  | -0.4888 | -0.4597 | -0.4781 |
| 3   | 1   | 1   | 6     | 1     | 1072.88  | -0.4883 | -0.4580 | -0.4772 |
| 4   | 1   | 1   | 6     | 1     | 1073.62  | -0.4882 | -0.4565 | -0.4766 |
| 5   | 1   | 1   | 6     | 1     | 1077.65  | -0.4896 | -0.4566 | -0.4775 |

| 1   | 1   | 2   | 4     | 0     | 1056.60  | -0.4854 | -0.4682 | -0.4791 |
| 2   | 1   | 2   | 4     | 0     | 1058.47  | -0.4858 | -0.4673 | -0.4790 |
| 3   | 1   | 2   | 4     | 0     | 1058.89  | -0.4855 | -0.4657 | -0.4783 |
| 4   | 1   | 2   | 4     | 0     | 1059.52  | -0.4854 | -0.4642 | -0.4776 |
| 5   | 1   | 2   | 4     | 0     | 1062.80  | -0.4864 | -0.4640 | -0.4782 |

| 1   | 1   | 2   | 6     | 0     | 1062.14  | -0.4870 | -0.4673 | -0.4798 |
| 2   | 1   | 2   | 6     | 0     | 1065.72  | -0.4882 | -0.4671 | -0.4805 |
| 3   | 1   | 2   | 6     | 0     | 1065.86  | -0.4878 | -0.4654 | -0.4796 |
| 4   | 1   | 2   | 6     | 0     | 1065.09  | -0.4879 | -0.4642 | -0.4793 |
| 5   | 1   | 2   | 6     | 0     | 1071.57  | -0.4896 | -0.4645 | 0.4804 |

The model is given by:
\[
f_k(r_t | x_t; \eta) = \frac{[P_k(z_t, x_t)]^2 \phi(z_t)}{\int_{-\infty}^{+\infty} [P_k(u, x_t)]^2 \phi(u) du} r_{t-1}^{-1/2} h_t^{-1},
\]
where \( \phi(\cdot) \) is the standard normal density, and
\[
z_t = \frac{r_t - \mu_t}{r_{t-1}^{1/2} h_t},
\]
\[
\mu_t = \phi_0 + (1 + \phi_1) r_{t-1} + \sum_{i=1}^{s-1} \phi_{i+1} \Delta r_{t-i},
\]
\[
P_k(z_t, x_t) = \sum_{i=0}^{K_x} a_i(x_t) z^i = \sum_{i=0}^{K_x} \left( \sum_{\delta = 0}^{K_x} a_{i\delta} x_t^\delta \right) z^i, \quad a_{00} = 1,
\]
\[
\log h_t = \omega + \sum_{i=1}^{s} \beta_i \log h_{t-i} + (1 + \alpha_1 L + \cdots + \alpha_s L^s) [\theta_1 z_{t-1} + \theta_2 (b(z_{t-1}) - \sqrt{2/\pi})],
\]
with \( b(z) \) representing a close, but twice differentiable approximation to the absolute value function defined in Footnote 6, \( x_t \) denotes a set of lagged values of \( r_t \), and \( \eta \) is a vector containing all unknown parameters.

The choice of score generator(s) is therefore also guided by the more conservative HQC and BIC criteria.

Initially, we consider, in Table 3, models with an EGARCH(1, 1) leading term and homogeneous innovation density, i.e., \( K_x = 0 \). The AIC points towards the
AR(5)-Hermite(6,0) model. The more general Hermite(8,0) has a higher AIC, and the additional two Hermite parameters are not significant in a likelihood ratio test. The BIC prefers the simpler Hermite(4,0) with an AR(1) leading term, but specification tests (not reported) reveal that the AR(1) model leaves a fair amount of autocorrelation in the residuals, with higher-order AR models doing much better in this regard. Furthermore, the four additional parameters in the AR(5) specification are significant. With s > 1, the BIC also selects the Hermite(6,0), although with an AR(2) leading term. This model is also selected by the HQC, but it is only slightly better than the AR(5)-Hermite(6,0). Since the AR(5) leading term does better on the specification tests, and the additional three parameters are significant at the 1% level in a likelihood ratio test, the AR(5)-Hermite(6,0) model remains our preferred model in Table 3.

Table 4 consider two extensions. First, the Hermite polynomial is expanded to $K_x = 1$, i.e., a nonhomogeneous innovation density, and, second, we let $q = 2$ in the EGARCH term. The case $K_x = 8$ is not examined, partly due to convergence problems. None of the extensions appear particularly appealing, except perhaps when judged on the liberal AIC benchmark. Therefore, we decided to use the AR(5)-EGARCH(1,1)-Hermite(6,0) model for the EMM estimations in the following sections.

Table 5 summarizes the parameter estimates for the relevant ARCH coefficients in the chosen SNP representation. There are only minor changes from the pure ARCH results reported earlier. The main difference is the weak evidence of an asymmetric volatility response to interest rate innovations, i.e. $\theta_1$ has turned decidedly more negative. All other inference is qualitatively the same as before.

Our SNP results differ from those of Tauchen (1995) in two respects. First, he ends up with an AR(1)-mean specification, whereas we choose the longer AR(5). The studies use different data, but we conjecture that the discrepancy is due to our longer sample that allows for better identification of the complex mean dynamics. It remains, however, an open question whether this mean dependency is economically significant. Second, he selects $K_x = 1$, which allows conditional heterogeneity into the innovation density, while we find no evidence that such an extension is required. This is in all likelihood due to the different selection of the leading term for the SNP expansion. We rely on recent results of Gallant and Long (1995) to justify the use of a non-Markovian term, Level-EGARCH,

---

It should be pointed out, though, that the Ljung–Box tests for residual autocorrelation are generally significant, even for the AR(5) model. However, inspection of the correlogram reveals no systematic pattern, so the results are likely due to the combination of the strong drift in the interest rate process over the sample period and a few large outliers.

Since we end up with a homogeneous innovation density, our SNP approach ends up mirroring the spirit of the semi-parametric ARCH estimator proposed by Engle and González-Rivera (1991).
Table 5
Estimation of the Hermite(6,0) SNP model with an Level-AR(5)-EGARCH(1,1) leading term for weekly observations on U.S. three-month Treasury Bill rate

<table>
<thead>
<tr>
<th>Parameter</th>
<th>( -\phi_0/\phi_1 )</th>
<th>1 + ( \phi_1 )</th>
<th>( \sum \beta_i )</th>
<th>( \gamma )</th>
<th>1 + ( \sum \alpha_j )</th>
<th>( \theta_1 )</th>
<th>( \theta_2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Estimate</td>
<td>4.09</td>
<td>0.9977</td>
<td>0.9876</td>
<td>0.6190</td>
<td>0.6220</td>
<td>-0.0724</td>
<td>0.4440</td>
</tr>
<tr>
<td>(s.e.)</td>
<td>(0.0013)</td>
<td>(0.0037)</td>
<td>(0.1226)</td>
<td>(0.0787)</td>
<td>(0.0310)</td>
<td>(0.0427)</td>
<td></td>
</tr>
</tbody>
</table>

The Level-EGARCH\((s,p,q)\) model is:

\[
\begin{align*}
    r_t &= \phi_0 + (1 + \phi_1)r_{t-1} + \sum_{i=1}^{s-1} \phi_{i-1} \Delta r_{t-i} + r_{t-1} Z_t^{1/2} Z_t \\
    \log h_t &= \omega + \sum_{i=1}^{p} \beta_i \log h_{t-i} + \left(1 + \sum_{j=1}^{q} \alpha_j L^j\right) \\
    &\times [\theta_1 Z_{t-1}^{1/2} + \theta_2 (b(z_{t-1}^2) - (2/\gamma)^{1/2})]
\end{align*}
\]

The models were estimated by Maximum Likelihood, using the annualized yield on the weekly T-Bills with 90 days to maturity from January 1954 to April 1995. The standard errors are computed from the robust formula (Bollerslev and Wooldridge, 1992).

which provides a parsimonious representation of the conditional distribution, while Tauchen uses the Markovian pure ARCH. Given the evidence high volatility persistence and significant level effects, it is clear that a relatively low-order ARCH cannot provide an adequate fit to the conditional density. Thus, the Hermite polynomial is forced to accommodate these features in Tauchen's analysis. In our case, the insignificance of the \( K_x \) expansion likely reflects the extensive analysis of candidate discrete-time models. Our SNP-results are entirely consistent with the conclusion that the Gaussian Level-EGARCH model captures the bulk of the conditional heteroskedasticity in the interest rates. Hence, the role of our Hermite expansion is merely to accommodate the non-Gaussian shape of the innovation density, e.g., fat tails and possible asymmetries, and not to compensate for missing level or ARCH effects. This is potentially an advantage of our approach and may explain why our final EMM estimation step is remarkably successfully, relative to prior empirical work, in identifying the level effect and stochastic volatility dynamics separately.

7. Estimation by efficient method of moments

The EMM estimator is closely related to the GMM estimator of Hansen (1982). The moment condition utilized by EMM is the expectation of the score function of the auxiliary model (our SNP model), i.e.,

\[
m(\rho, \hat{\eta}) = \int \frac{\partial \log f_{X}(r | X; \hat{\eta})}{\partial \eta} \, dP(r, X; \rho), \tag{13}
\]
where \( P(r, X; \rho) \) is the (true) joint probability measure for the short rate, \( r \), and the conditioning arguments in \( X \),\(^{19}\) and \( \hat{\eta} \) is the QML estimate of the parameters in the score generator. Note that the moment conditions depend only on the vector of structural parameters, \( \rho \), through the integration (probability) measure \( P \). An analytical computation of the expectation in Eq. (13) is rarely feasible, and Monte Carlo integration is used in its place. Specifically, for a candidate value of the parameter vector, \( \rho \), we simulate a time series of length \( N \) from the structural model, denoted by \( \{ r_t(\rho), X_t(\rho) \} \), and estimate (13) by

\[
m_N(\rho, \hat{\eta}) = \frac{1}{N} \sum_{t=1}^{N} \frac{\partial \log f_k(r_t(\rho)|X_t(\rho); \hat{\eta})}{\partial \eta}.
\]

Since the structural model is assumed stationary and ergodic, \( m_N(\rho, \hat{\eta}) \to m(\rho, \hat{\eta}) \) as \( N \to \infty \) almost surely. Preferably, \( N \) should be large enough that Monte Carlo error can be ignored, i.e., \( m_N(\rho, \hat{\eta}) \) can be substituted for \( m(\rho, \hat{\eta}) \) when deriving the asymptotic properties of the estimator. There are a number of implementation issues that arise at this step of the EMM procedure, such as choice of \( N \), treatment of initial conditions, and the simulation/discretization scheme for the diffusion model. We discuss these issues in relation to our stochastic volatility diffusion for the short-term interest rate in the Appendix.

The EMM estimator of \( \rho \) is obtained by minimizing a quadratic form in the vector of moment conditions, \( m_N(\rho, \hat{\eta}) \), using a weighting matrix \( W_T \):

\[
\hat{\rho} = \arg \min_{\rho} m_N(\rho, \hat{\eta})^T W_T m_N(\rho, \hat{\eta}).
\]

Since all of the standard GMM theory applies, the optimal EMM estimator is obtained if the weighting matrix \( W_T \) is a consistent estimator of the inverse of the asymptotic covariance matrix for the moment conditions.\(^{20}\) Note that the elements of Eq. (13) are stochastic only because of the dependence on the estimated \( \eta \). If the score generator provides a good description of the dynamic properties of \( \{ r_t \} \), which is implicitly assumed when using SNP, the scores \( \{ \partial \log f_k(r_t|X_t; \hat{\eta})/\partial \eta \} \) display (near) martingale difference behavior, and the covariance matrix of \( T^{1/2} m(\rho, \hat{\eta}) \) may be estimated from the ‘outer product of the gradient’ (OPG) formula (Gallant and Tauchen, 1994),

\[
I_T(\hat{\eta}) = \frac{1}{T} \sum_{t=1}^{T} \frac{\partial \log f_k(r_t|X_t; \hat{\eta})}{\partial \eta} \frac{\partial \log f_k(r_t|X_t; \hat{\eta})}{\partial \eta'}.
\]

\(^{19}\) In the non-Markovian setting \( X \) contains an infinite number of lagged observations (or at least functions thereof), so an exact characterization of the moment condition relies on limit operations. We refer to Gallant and Long (1995) for a thorough discussion.

\(^{20}\) Using the optimal weighting matrix is, of course, a precondition for obtaining full asymptotic efficiency with a SNP score generator. However, the statistical inference presented below also relies on \( W_T \) being optimal, cf. Hansen (1982).
Note that, contrary to what is usually the case, the estimator of the optimal weighting matrix does not depend on the structural parameters $\rho$, thus avoiding the need of using an iterated GMM estimator.

The (estimated) asymptotic covariance matrix of the EMM estimator $\hat{\rho}$ is given by
\[
\text{cov}(\hat{\rho}) = \frac{1}{T} \left( \frac{\partial m_N(\hat{\rho}, \hat{\eta})'}{\partial \rho} I_T(\hat{\eta})^{-1} \frac{\partial m_N(\hat{\rho}, \hat{\eta})}{\partial \rho'} \right)^{-1}
\equiv \frac{1}{T} (M'_\rho I_T(\hat{\eta})^{-1} M_\rho)^{-1},
\]
and if the structural model is correctly specified, the minimized value of $T$ times the EMM criterion function is distributed as $\chi^2$ with $\text{dim}(\eta) - \text{dim}(\rho)$ degrees of freedom, where $\text{dim}(w)$ denotes the number of elements in the vector $w$. This is an omnibus test of the overidentifying restrictions, and if the test rejects the model, it is useful to examine the individual elements of the score vector $m_N(\hat{\rho}, \hat{\eta})$.

Inference should, of course, be based on the $t$-statistics, i.e., the elements of the score vector divided by their standard errors. The requisite standard errors for $m_N(\hat{\rho}, \hat{\eta})$ are given by the square root of the diagonal elements of the matrix
\[
\frac{1}{T} (I_T(\hat{\eta}) - M_\rho(M'_\rho I_T(\hat{\eta})^{-1} M_\rho)^{-1} M'_\rho).
\]

Note that Eq. (18) involves no unknown matrices beyond those used in the calculation of the covariance matrix for $\hat{\rho}$. The Jacobian $M_\rho$ must, in general, be computed by numerical differentiation.\(^{21}\)

8. Estimation results for the continuous-time models

The second step of the EMM estimation procedure requires the simulation of a long sample from the continuous-time model. Our simulation uses the simple Euler scheme with 25 subdivisions per time unit, i.e., we make 25 draws in order to record one simulated (weekly) data point. Moreover, we make use of antithetic variables in order to reduce the Monte Carlo variability associated with the evaluation of the integral in Eq. (13), which is calculated on the basis of a final simulated sample of 75,000 'weekly observations'. Further details regarding the implementation are provided in the Appendix.

\(^{21}\)The square root of the diagonal elements of $I_T(\hat{\eta})$ can be used as an approximation to these standard errors, but this approach ignores the fact that the score vector $m_N(\hat{\rho}, \hat{\eta})$ is calculated at an estimate of $\rho$, not at the true value $\rho_0$. Because of this 'pre-whitening' effect, the approximate $t$-statistics are biased downwards, leading to a loss of power if they are compared to percentiles of the normal distribution. Gallant and Tauchen (1995a) denote these statistics 'quasi-$t$-ratios'.
Table 6
Estimation of the continuous-time two-factor stochastic volatility model by EMM, U.S. three-month T-Bills, weekly, January 6, 1954 to April 19, 1995

<table>
<thead>
<tr>
<th>Variable</th>
<th>Estimate</th>
<th>Std. error</th>
<th>$t$-stat.</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\kappa_1$</td>
<td>0.163</td>
<td>0.057</td>
<td>2.86</td>
</tr>
<tr>
<td>$\alpha$</td>
<td>-0.282</td>
<td>0.180</td>
<td>-1.57</td>
</tr>
<tr>
<td>$\kappa_2$</td>
<td>1.04</td>
<td>0.127</td>
<td>8.17</td>
</tr>
<tr>
<td>$\zeta$</td>
<td>1.27</td>
<td>0.059</td>
<td>21.4</td>
</tr>
<tr>
<td>$\mu$</td>
<td>5.95</td>
<td>0.464</td>
<td>12.8</td>
</tr>
<tr>
<td>$\gamma$</td>
<td>0.544</td>
<td>0.084</td>
<td>6.45</td>
</tr>
</tbody>
</table>

Function value: 24.25  Degrees of freedom: 12  $p$-value: 0.019

Score generator diagnostics

<table>
<thead>
<tr>
<th>Moment</th>
<th>Estimate</th>
<th>Std. error</th>
<th>$t$-stat.</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\phi_0$</td>
<td>0.111</td>
<td>0.220</td>
<td>0.506</td>
</tr>
<tr>
<td>$\phi_1$</td>
<td>-0.030</td>
<td>0.734</td>
<td>-0.041</td>
</tr>
<tr>
<td>$\phi_2$</td>
<td>-0.045</td>
<td>0.023</td>
<td>-1.964</td>
</tr>
<tr>
<td>$\phi_3$</td>
<td>0.007</td>
<td>0.021</td>
<td>0.310</td>
</tr>
<tr>
<td>$\phi_4$</td>
<td>-0.023</td>
<td>0.022</td>
<td>-1.042</td>
</tr>
<tr>
<td>$\phi_5$</td>
<td>-0.053</td>
<td>0.023</td>
<td>-2.348</td>
</tr>
<tr>
<td>$\gamma$</td>
<td>-0.001</td>
<td>0.001</td>
<td>-0.438</td>
</tr>
<tr>
<td>$\sigma_1$</td>
<td>0.005</td>
<td>0.009</td>
<td>-0.517</td>
</tr>
<tr>
<td>$\beta_1$</td>
<td>0.220</td>
<td>0.409</td>
<td>0.538</td>
</tr>
<tr>
<td>$\theta_1$</td>
<td>-0.003</td>
<td>0.016</td>
<td>-0.218</td>
</tr>
<tr>
<td>$\theta_2$</td>
<td>-0.007</td>
<td>0.013</td>
<td>-0.551</td>
</tr>
<tr>
<td>$\gamma$</td>
<td>0.000</td>
<td>0.004</td>
<td>0.103</td>
</tr>
<tr>
<td>$\alpha_{10}$</td>
<td>-0.009</td>
<td>0.039</td>
<td>-0.218</td>
</tr>
<tr>
<td>$\alpha_{20}$</td>
<td>-0.048</td>
<td>0.035</td>
<td>-1.354</td>
</tr>
<tr>
<td>$\alpha_{30}$</td>
<td>0.028</td>
<td>0.039</td>
<td>0.731</td>
</tr>
<tr>
<td>$\alpha_{40}$</td>
<td>-0.030</td>
<td>0.035</td>
<td>-0.849</td>
</tr>
<tr>
<td>$\alpha_{50}$</td>
<td>0.042</td>
<td>0.036</td>
<td>1.178</td>
</tr>
<tr>
<td>$\alpha_{60}$</td>
<td>0.145</td>
<td>0.041</td>
<td>3.524</td>
</tr>
</tbody>
</table>

The two-factor model is:

$$dr_t = \kappa_1(\mu - r_t)dt + \sigma_t r_t \, dW_{1,t}$$

$$d \log \sigma_t^2 = \kappa_2(\alpha - \log \sigma_t^2)dt + \zeta \, dW_{2,t}$$

$W_{1,t}$ and $W_{2,t}$ are two independent Brownian Motion processes, and $\gamma > 0$.

Estimation is performed by EMM with a Hermite(6,0) SNP model with Level-AR(5)-EGARCH(1,1) as a leading term. This 'score-generator' model is described in the notes to Table 5. The simulated series for computation of the expected scores were 75,000 long and antithetic variables were used to reduce the Monte Carlo variability. The simulations were generated by the Euler scheme using 25 subintervals per week.
The estimation results for the two-factor model are given in Table 6. All important parameters are highly significant and estimated with a reasonable degree of precision. In particular, it is reassuring that both \( \kappa_1 \) and \( \kappa_2 \) are strictly positive since this rules out the possibility of a unit root in the mean and volatility dynamics, respectively. The parameter estimates are furthermore readily interpretable. First, note that \( \mu \) provides an estimate of the unconditional mean of the short-term interest rate of 5.95\%, while the level coefficient, \( \gamma \), is estimated at 0.54 and again not significantly different from \( \frac{1}{2} \). Second, for interpretation of the strength of the implied mean reversion in the mean and volatility dynamics, note that the solution of Eq. (4) takes the form

\[
\log \sigma_t^2 = \exp(-\kappa_2(s-t)) \log \sigma_r^2 + \alpha(1 - \exp(-\kappa_2(s-t))) \\
+ \xi \int_t^s \exp(-\kappa_2(s-v)) dW_v. \tag{19}
\]

This discrete-time specification identifies the first component on the right-hand side of (19) as the relevant discrete-time autoregressive coefficient, while the second and third terms represent a constant drift and an i.i.d. mean zero error term, respectively. Because the time span from \( t = 1 \) to \( t = 2 \) corresponds to one year, the discrete-time autoregressive coefficient for the weekly frequency is obtained by letting \( s = t + 1/52 \), or \( \exp(-\kappa_2(1/52)) = \exp(-1.04/52) = 0.9802 \). Thus, the implied estimated measure of (log-)volatility persistence is about 0.98 at the weekly level which is in line with the discrete time evidence as well as the separate evidence of Torus and Ball (1995). The identical calculation for the discrete-time autoregressive coefficient in the mean dynamics produces \( \exp(-\kappa_1(1/52)) = \exp(-0.163/52) = 0.9969 \). Thus, the mean dynamics is extremely close to being nonstationary, although we statistically can reject the unit root hypothesis with a fairly large degree of confidence.

While the overall \( \chi^2 \)-test with an associated \( p \)-value of 2\% provides mild evidence against our two-factor diffusion model, the result is actually quite encouraging for two reasons. It should, first of all, be judged in the light of the prior findings of Tauchen (1995) for a one-factor specification similar to the CKLS model. He is unable to identify the level coefficient, \( \gamma \), with an acceptable degree of precision, and his overall specification test indicates an overwhelming rejection of the CKLS-type diffusion model. We present corroborating evidence in Table 7, which provides EMM estimates of the one-factor CKLS-model in Eq. (1). This model attains a \( p \)-value of \( 2 \times 10^{-5} \). Clearly, the introduction of a stochastic volatility factor has an enormous effect in terms of enabling the diffusion to match the characteristics of the data. Further, we also find that the level coefficient, \( \gamma \), increases in the absence of the stochastic volatility factor in Table 7, and the estimates become markedly less precise, again confirming the estimation results from discrete-time models and the qualitative features of Tauchen’s results. Second, we have in some sense stacked the deck against our
Table 7
Estimation of the continuous-time one-factor model by EMM, U.S. three-month T-Bills, weekly, January 6, 1954 to April 19, 1995

<table>
<thead>
<tr>
<th>Variable</th>
<th>Estimate</th>
<th>Std. error</th>
<th>t-stat.</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\kappa_1$</td>
<td>0.082</td>
<td>0.075</td>
<td>1.089</td>
</tr>
<tr>
<td>$\mu$</td>
<td>6.279</td>
<td>0.677</td>
<td>9.254</td>
</tr>
<tr>
<td>$\sigma$</td>
<td>0.670</td>
<td>0.051</td>
<td>13.13</td>
</tr>
<tr>
<td>$\gamma$</td>
<td>0.676</td>
<td>0.149</td>
<td>4.534</td>
</tr>
</tbody>
</table>

Function value: 46.98  Degrees of freedom: 14  $p$-value: 1.93 $\cdot 10^{-5}$

Score generator diagnostics

<table>
<thead>
<tr>
<th>Moment</th>
<th>Estimate</th>
<th>Std. error</th>
<th>t-stat.</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\phi_0$</td>
<td>-0.121</td>
<td>0.101</td>
<td>-1.193</td>
</tr>
<tr>
<td>$\phi_1$</td>
<td>-0.742</td>
<td>0.503</td>
<td>-1.475</td>
</tr>
<tr>
<td>$\phi_2$</td>
<td>-0.048</td>
<td>0.023</td>
<td>-2.109</td>
</tr>
<tr>
<td>$\phi_3$</td>
<td>-0.006</td>
<td>0.025</td>
<td>-0.240</td>
</tr>
<tr>
<td>$\phi_4$</td>
<td>-0.035</td>
<td>0.023</td>
<td>-1.471</td>
</tr>
<tr>
<td>$\phi_5$</td>
<td>-0.052</td>
<td>0.024</td>
<td>-2.215</td>
</tr>
<tr>
<td>$\omega$</td>
<td>-0.003</td>
<td>0.002</td>
<td>-1.869</td>
</tr>
<tr>
<td>$\alpha_1$</td>
<td>0.007</td>
<td>0.011</td>
<td>0.707</td>
</tr>
<tr>
<td>$\beta_1$</td>
<td>1.135</td>
<td>0.503</td>
<td>2.258</td>
</tr>
<tr>
<td>$\theta_1$</td>
<td>0.009</td>
<td>0.019</td>
<td>0.459</td>
</tr>
<tr>
<td>$\theta_2$</td>
<td>-0.011</td>
<td>0.015</td>
<td>-0.702</td>
</tr>
<tr>
<td>$\gamma$</td>
<td>0.001</td>
<td>0.005</td>
<td>0.191</td>
</tr>
<tr>
<td>$a_{10}$</td>
<td>0.004</td>
<td>0.031</td>
<td>0.123</td>
</tr>
<tr>
<td>$a_{20}$</td>
<td>-0.154</td>
<td>0.040</td>
<td>-3.878</td>
</tr>
<tr>
<td>$a_{30}$</td>
<td>-0.015</td>
<td>0.043</td>
<td>-0.359</td>
</tr>
<tr>
<td>$a_{40}$</td>
<td>-0.043</td>
<td>0.043</td>
<td>-1.014</td>
</tr>
<tr>
<td>$a_{50}$</td>
<td>0.006</td>
<td>0.042</td>
<td>0.135</td>
</tr>
<tr>
<td>$a_{60}$</td>
<td>0.186</td>
<td>0.043</td>
<td>4.342</td>
</tr>
</tbody>
</table>

The one-factor model is: \[ dr = \kappa_1(\mu - r_t)dt + \sigma tr_t^\gamma dW_{1,t} \]

$W_{1,t}$ is a Brownian Motion process, and $\gamma > 0$.

Estimation is performed by EMM with a Hermite(6,0) SNP model with Level-AR(5)-EGARCH(1,1) as a leading term. The 'score-generator' model is described in the notes to Table 5. The simulated series for computation of the expected scores were 75,000 long and antithetic variables were used to reduce the Monte Carlo variability. The simulations were generated by the Euler scheme using 25 subintervals per week.

diffusion by including a string of, arguably, marginally significant higher-order autoregressive terms in our score generator. It is hard to see how the assumed diffusion should be able to generate any such complex mean dependency, and from the reported individual score generator diagnostics in Table 6, we do
indeed find that the scores corresponding to the second- and fifth-order autocorrelation coefficients contribute significantly to the overall high value of the GMM criterion.

This being said, there are also some indications of model misspecification among the scores associated with the Hermite polynomial in Table 6. This is likely due to a failure to generate innovations with sufficiently heavy tails. This does suggest that there is room for improvement vis-a-vis the assumed two-factor model. A potential extension of the diffusion model that may accommodate both the apparent strong, but time-varying, mean drift over the sample period and the lack of sufficient outlier activity involves the introduction of a third factor associated with the mean level, \( \mu \), towards which the short rate reverts. Time variation in this coefficient, \( \mu_t \), may be interpreted as variation in an underlying inflation rate, and may plausibly induce the type of prolonged mean drift that characterizes our sample period. In fact, in a simpler modeling context Balduzzi et al. (1995) find general support for extensions in this direction. However, investigation of this conjecture within our stochastic volatility framework is left for future research.

Finally, Fig. 4 displays a representative simulated sample from our estimated two-factor model. It is included to convey an impression of the type of sample paths that may be generated by the model. Although the particular sample does

![Figure 4](image_url) Fig. 4. Simulated series from the two-factor stochastic volatility diffusion.
not generate an episode as volatile as the Monetary Experiment, the figure shares a number of qualitative features with the actual interest rate series in Fig. 1.

8. Concluding remarks

This paper provides consistent estimates for the parameters of a two-factor continuous-time model for the short-term risk-free interest rate. The resultant specification effectively extends the standard CIR model to a stochastic volatility setting. The incorporation of the unobservable volatility factor is shown to greatly enhance the model's ability to fit the data. Put simply, the stochastic volatility factor seems to be an integral part of the short rate dynamics. By implication, we suspect that estimation techniques which focus on estimation of one-factor interest rate models may prove of limited applicability. Our estimates are obtained via implementation of the Efficient Method of Moments (EMM) procedure of Gallant and Tauchen (1994). As part of the implementation, we thoroughly investigate current popular discrete-time models for the short rate. We find the recently proposed Level-GARCH models to exhibit excessively volatile behavior, and thus constitute poor candidate models for the interest rate process, while the closely related Level-EGARCH models appear to provide an overall adequate fit. Using the latter representation as the leading term in an SNP expansion, we obtain an operational characterization of the salient features of the short rate process. Our use of a good candidate model for the leading term appears to enhance the identification and estimation efficiency in the final simulation based step of the implementation of EMM considerably. Finally, the powerful EMM diagnostics suggest interesting avenues for further extensions of the model.

Appendix

Numerical implementation of EMM

The appendix describes computational aspects of EMM estimation of continuous-time stochastic volatility models with emphasis on the Monte Carlo integration that computes the moment conditions used in the GMM procedure. Specifically, as discussed in Section 7, we need to approximate the expectation

\[ m(\rho, \hat{\eta}) = \int \frac{\partial \log f_K(r | X; \hat{\eta})}{\partial \eta} dP(r; X; \rho) \]  (A.1)

by the Monte Carlo estimate

\[ m_N(\rho, \hat{\eta}) = \frac{1}{N} \sum_{i=1}^{N} \frac{\partial \log f_K(r(t_i, \rho) | X(t_i, \rho); \hat{\eta})}{\partial \eta}, \]  (A.2)
where the simulated time series\footnote{The notation in the appendix differs slightly from the remainder of the paper due to the need to distinguish between the continuous-time process and the discretely observed (simulated) data. Therefore, in the following $r(t_i)$ denotes the $i$th observation on the interest rate, observed at time $t_i$.} $\{r(t_i, \rho, X(t_i, \rho))\}, \; i = 1, \ldots, N$, is generated from the probability distribution induced by the continuous-time process

$$
dr_t = \kappa(\mu - r_t) \, dt + \sigma_t r_t^\gamma \, dW_{1,t}, \quad \gamma > 0,$$

$$
d \log \sigma_t^2 = \kappa_2(\alpha - \log \sigma_t^2) \, dt + \xi \, dW_{2,t},$$

and $W_{1,t}$, $W_{2,t}$ are standard independent Brownian motion processes. This is done by dividing each week into $S$ subintervals, and simulating a total of $N \cdot S$ observations from an approximation to (A.3) and (A.4). We use the Euler scheme which replaces (A.3) and (A.4) by the (first-order) discrete-time approximation

$$
r_{t+\Delta} = r_t + \kappa_1(\mu - r_t) \Delta + \sigma_t r_t^\gamma \sqrt{\Delta z_{1t}},$$

$$
\log \sigma_{t+\Delta}^2 = \log \sigma_t^2 + \kappa_2(\alpha - \log \sigma_t^2) \Delta + \xi \sqrt{\Delta z_{2t}},$$

where $(z_{1t}, z_{2t})$ are two independent normal $N(0,1)$ variates and $\Delta = 1/(52 \cdot S)$. In our computer program we generate normal variates using the ratio of uniforms method (Kinderman and Monahan, 1977). The uniform $(0,1)$ random numbers are generated by combining two linear congruential generators, as discussed by L'Ecuyer (1988).

The Euler approximation, with $S = 1$, is often used when estimating parameters of stochastic differential equations from discretely observed data. Because we are using a simulation-based estimation technique, we can use any value of $S > 1$ which reduces the discretization bias. The empirical analysis in Section 7 is based on $S = 25$ subintervals. For a given $S$ the discretization bias could theoretically be reduced further by employing a higher-order simulation scheme, cf. Kloeden and Platen (1992). These schemes are similar (but not identical) to higher-order Taylor series expansions for deterministic functions. Therefore, they are more expensive in terms of computing time, and we suspect that the variance reduction technique discussed below (antithetic variables) would be less effective with a higher-order scheme due to the additional squared terms in $z_{1t}$ and $z_{2t}$. Furthermore, we observed only minor changes in the results for other values of $S$.

In the Monte Carlo integration (A.2) we implicitly assume that the simulated time series of the interest rate, and hence the score function, is generated from the stationary distribution. Strictly speaking, this is impossible since we are unable to draw the initial values $(r_0, \log \sigma_0^2)$ from their unconditional distribution. Instead, the state variables $r_t$ and $\log \sigma_t^2$ are initialized at their unconditional
means, $\mu$ and $\alpha$, respectively, and the first $N_1$ simulated values of $r_t$ are discarded. For large enough $N_1$, the effect of initial conditions wears off, and we are then drawing $r_t$ from the stationary distribution. The next step is to draw a simulated time series $\{r_t\}$ of length $N_2 + N$ and calculate the score function for the SNP auxiliary model. We discard the first $N_2$ draws of the score vector since the EARGARCH leading term itself depends on initial conditions ($\log h_0$). This leaves an effective sample of $N$ draws of the score vector, which is then used to calculate (A.2). In our empirical application we use $N_1 = 2000$ and $N_2 = 1000$. Other values were tried, but with no discernible effect on the results.

Finally, we need a choice of $N$, the effective simulated sample size. Ideally, $N$ should be so large that the Monte Carlo error is negligible. An easy way of assessing this requirement is to compute the EMM criterion function for two different simulated series of $\{r_t\}$, using two different initial seeds in the random number generator. However, this exercise revealed a significant degree of dependence on the choice of the random sequence, and experimentation with different values of $N$ suggested that we would have to simulate millions of observations in order to make the Monte Carlo error negligible. With a sample size that large one evaluation of the EMM criterion function would take more than an hour on a state-of-the-art PC. This problem stems from the extremely high persistence in the conditional mean and variance that is found in the actual short rate series and, of course, replicated in the simulated series. Furthermore, for realistic values of $N$ (e.g., 100,000) some of the parameter estimates showed a great deal of dependence on the choice of random sequence (initial seed, especially for $\alpha$, the mean of $\log \sigma_t^2$).

We consequently experimented with variance reduction techniques and found the antithetic variables technique to be quite effective, especially in reducing the Monte Carlo error in the estimates of the structural parameters. Briefly, the idea behind the antithetic variables techniques is to average two estimates of the integral (A.1) which are presumed to be negatively correlated. We first calculate (A.2) using the random sequence $\{z_{1t}, z_{2t}\}$, according to the procedure given above. The second estimate is then computed using the same random numbers, but with the opposite sign, i.e., the sequence $\{-z_{1t}, -z_{2t}\}$. Note that this sequence is also independently and normally distributed. In practice, the two simulated time series of $r_t$ are generated simultaneously, which avoids the need to regenerate the random numbers for the second estimate. All results reported in Section 7 are based on this technique with $N = 75,000$ observations in each of the two simulated samples of $\{r_t\}$.

---

23 The second round of discarding observations is necessary in order to draw from the stationary distribution of the score generator.

24 See Geweke (1995) for thorough discussion of the antithetic variables technique.
References

Bekaert, G., R.J. Hodrick, and D.A. Marshall, 1995, ‘Peso problem’ explanations for term structure anomalies, Manuscript (J.L. Kellogg Graduate School of Management, Northwestern University, Evanston, IL).
Conley, T., L.P. Hansen, E.G.J. Luttmer, and J. Scheinkman, 1995, Short-term interest rates as subordinated diffusions, Manuscript (University of Chicago, Chicago, IL).
Dybvig, P.H., 1989, Bond and bond option pricing based on the current term structure, Manuscript (Washington University, St. Louis, MO).
Geweke, J., 1995, Monte Carlo simulation and numerical integration, Staff report no. 192 (Federal Reserve Bank of Minneapolis, MN).
Ghysels, E. and J. Jasiak, 1995, Stochastic volatility and time deformation: an application to trading volume and leverage effects, Manuscript (University de Montréal, Montréal).
Gray, S.F., 1995, Modelling the conditional distribution of interest rates as a regime-switching process, Manuscript (Duke University, Durham, NC).
Kearns, P., 1992, Pricing interest rate derivative securities when volatility is stochastic, Manuscript (University of Rochester, Rochester, NY).
Newey, W. and D. Steigerwald, 1994, Consistency of quasi-maximum likelihood estimators for models with conditional heteroskedasticity, Manuscript (University of California, Santa Barbara, CA).
Pagan, A.R., A.D. Hall, and V. Martin, 1995, Exploring the relations between the finance and econometrics literature on the term structure, Manuscript (Australian National University, Canberra).
Stanton, R., 1995, A nonparametric model of term structure dynamics and the market price of interest rate risk, Manuscript (Haas School of Business, University of California, Berkeley, CA).
Toroos, W.N. and C.A. Ball, 1995, Regime shifts in short term riskless interest rates, Manuscript (Anderson Graduate School of Management, University of California, Los Angeles, CA).