Alternative models for stock price dynamics

Mikhail Chernov[a], A. Ronald Gallant[b], Eric Ghysels[b,*],
George Tauchen[c]

[a] *Columbia Business School, Division of Finance and Economics, 413 Uris Hall, 3022 Broadway,
New York, NY 10027, USA*
[b] *Department of Economics, University of North Carolina at Chapel Hill CB 3305, Chapel Hill,
NC 27599-3305, USA*
[c] *Department of Economics, 305 Social Science, Box 90097, Duke University, Durham,
NC 27708-0097, USA*

Abstract

This paper evaluates the role of various volatility specifications, such as multiple stochastic volatility (SV) factors and jump components, in appropriate modeling of equity return distributions. We use estimation technology that facilitates nonnested model comparisons and use a long data set which provides rich information about the conditional and unconditional distribution of returns. We consider two broad families of models: (1) the multifactor loglinear family, and (2) the affine-jump family. Both classes of models have attracted much attention in the derivatives and econometrics literatures. There are various tradeoffs in considering such diverse specifications. If pure diffusion SV models are chosen over jump diffusions, it has important implications for hedging strategies. If logarithmic models are chosen over affine ones, it may seriously complicate option pricing. Comparing many different specifications of pure diffusion multifactor models and jump diffusion models, we find that (1) log linear models have to be extended to two factors with feedback in the mean reverting factor, (2) affine models have to have a jump in returns, stochastic volatility or probably both. Models (1) and (2) are observationally equivalent on the data set in hand. In either (1) or (2) the key is that the volatility can move violently. As we obtain models with comparable empirical fit, one must make a choice based on arguments other than statistical goodness-of-fit criteria. The considerations include facility to price options, to hedge and parsimony. The affine specification with jumps in volatility might therefore be preferred because of the closed-form derivatives prices.
© 2003 Elsevier B.V. All rights reserved.

JEL classification: G13; C14; C52; C53

Keywords: Efficient method of moments; Poisson jump processes; Stochastic volatility models

---

* Corresponding author.
 E-mail address: eghysels@unc.edu (E. Ghysels).

0. Introduction

Stochastic volatility (SV) models are specifically designed to capture salient properties of volatility such as randomness and persistence. However, one of the most important recent findings is that these models are not able to characterize all aspects of asset returns distribution.[1] Indeed, given a reasonable fit to the conditional dynamics of volatility, SV models cannot match the high conditional kurtosis of returns (tail thickness) documented in the literature for many classes of financial assets, of which equities are the most prominent example.

This paper evaluates the role of various factors, such as additional volatility factors and jumps, in appropriate modeling of equity returns. To do so we estimate a variety of extensions of SV models using the same estimation technology with a common data set.

We are particularly interested in the role of SV factors and their functional form because this issue has not been considered in the prior work as much as the role of jump components has. Moreover, the evidence from option markets shows that adding a jump component to returns is not sufficient to fully capture the dynamics of financial series. Bakshi et al. (1997) and Bates (2000) find that the volatility of volatility coefficient, which is estimated from the underlying asset time series is much lower than the one estimated from the options cross-section. In addition, Pan (2002) using simultaneously equity returns and options prices finds evidence suggesting that the volatility of volatility is stochastic. This observation is confirmed by Jones (2003) who finds, based on the implied volatility series, that volatility of volatility is higher during the more volatile periods in the stock market.

This evidence suggests that an appropriate extension might involve two SV factors, thus breaking the link between tail thickness and volatility persistence. Depending on model specification—affine or logarithmic—a second SV factor may act as either a factor dedicated to exclusive modeling of tail behavior (the first factor is then often referred to as a long memory component), or as a stochastic volatility of volatility factor. In the latter specification, the volatility would be capable of making rapid moves, which is prohibited by a single SV specification. EJP propose to model the same feature by introducing a jump component to the SV factor.

There are various tradeoffs in considering these different specifications. If pure diffusion models are chosen over jump diffusions, it has important implications for hedging strategies. If logarithmic models are chosen over affine ones, it may seriously complicate option pricing. Finally, if we obtain models with comparable empirical fit, we would still have to make a choice based on arguments other than statistical goodness-of-fit criteria. Such arguments could be facility to price options, to hedge, or simply parsimony. Our approach allows us to address all these issues.

---

[1] Formal statistical diagnostics and rejections of SV models are reported in Andersen et al. (2002) (ABL hereafter), Chernov and Ghysels (2000), Eraker et al. (2003) (EJP hereafter), Jones (2003), and Pan (2002).

We estimate a total of ten different models, broadly classified as either affine or logarithmic. The benchmark for each class is a single SV model, i.e. the Heston (1993) model for affine class, and the Scott (1987) model for the logarithmic class. In the affine class the model is extended by considering two SV factors, a jump to returns (ABL), and a simultaneous jump to both returns and volatility (EJP). In the logarithmic class the extensions are achieved by adding a second SV factor, and by considering models with feedback (Gallant et al., 1999).

We consider a long data set, providing rich information about the conditional and unconditional distribution of returns under the objective probability measure. The common data set consists of returns on the Dow Jones industrial average (DJIA) index from January 1953 to July 1999, covering in addition to the market crash of October 1987, the more recent crashes of October 1997 and August 1998, as well as historical events such as the Cuban Missile Crisis in October 1962, or Arab Oil Embargo in October 1973. The long data set also offers more variety in the dynamics of volatility, which allows the determination of a more robust model. These are the longest series considered for such a study: corresponding to the combination of ABL, and EJP, who study 1953–1996 and 1980–1999, respectively. Although these authors use the S& P 500 index, our results should be qualitatively comparable because historically DJIA closely tracks the S& P 500 index.

Since the risk-neutral measure used in derivatives valuation has to coincide with the objective one up to sets of measure zero, the models of equities will retain the same factor structure under both probability measures. Therefore, despite the fact that we are not using options data in the present paper, we motivate our work by empirical results from both underlying and options literature. By the same token, our results will have implications for both.

The common estimation method is the efficient method of moments (EMM) of Gallant and Tauchen (1996). The advantages of using EMM, critical for comparison of several models, are that it offers: (1) formal statistical tests of a model fit, (2) formal diagnostics of model inadequacies and most importantly (3) nonnested specifications can be compared in a meaningful way since EMM forces all models to confront the same set of moment conditions.

The paper is organized as follows. In a first section, we describe the models that we consider in the study. The next section covers the estimation methods, briefly summarizing EMM procedure and the SNP model selection. Section three reports the empirical results. A final section concludes the paper.

1. Models specifications

In this section we describe the various classes of models we consider in our study, starting with single index volatility diffusion models in a first subsection. Special cases include affine models, constant elasticity of variance (CEV) and logarithmic models. In the second subsection, we discuss jump-diffusion models. In the third and final subsection we introduce a unified notation for the different models.

1.1. Single index volatility diffusion models

The starting point is the multifactor pure diffusion SV model. We consider models with at most four factors, namely,

$$dP_t/P_t = (\alpha_{10} + \alpha_{12} U_{2t}) \, dt + \sigma(\beta_{10} + \beta_{13} U_{3t} + \beta_{14} U_{4t})$$

$$\times (\psi_{11} \, dW_{1t} + \psi_{13} \, dW_{3t} + \psi_{14} \, dW_{4t}), \tag{1}$$

$$dU_{2t} = (\alpha_{20} + \alpha_{22} U_{2t}) \, dt + \beta_{20} \, dW_{2t}, \tag{2}$$

$$dU_{it} = (\alpha_{i0} + \alpha_{ii} U_{it}) \, dt + (\beta_{i0} + \beta_{ii} U_{it})^{\gamma_i} \, dW_{it}, \quad i = 3, 4. \tag{3}$$

In the above, P_t represents the financial price series evolving in continuous time (we reserve the notation U_{1t} for the logarithm of the price).

We allow for a flexible drift specification via a stochastic factor U_{2t}, which evolves according to an Ornstein–Uhlenbeck process. This specification can accommodate the mild serial correlation appearing in the returns series, which may be explained by the nonsynchronous trading and unexpected stochastic dividend effects. An alternative strategy to incorporate these effects would be to prefilter the data as was done in ABL or Gallant and Tauchen (1993).

We model the diffusion coefficient $\sigma(.)$ as a function of the linear combination of the two stochastic volatility factors U_{3t} and U_{4t}, which are described by the usual mean reverting processes. The mean reversion parameters α_{ii} or respective volatility half-lives $\log 2/\alpha_{ii}$ measure persistence of these processes.

The use of a functional transformation of linear combination of factors is reminiscent of index models used in various areas of econometrics.[2] We can, therefore, refer to the volatility models as single index volatility (SIV) models. Different specifications of the index function will yield various classes of SV models, including affine and logarithmic two volatility factor models.

Finally, we parameterize $\psi_{11} = \sqrt{1 - \psi_{13}^2 - \psi_{14}^2}$ so that ψ_{13} and ψ_{14} are correlation coefficients. Note that, when we have multiple stochastic volatility factors, the correlation coefficients loose the interpretation of the leverage effect, i.e. the instantaneous correlation between returns and changes in variance. In this case, as can be easily shown, the leverage effect is equal to

$$\text{corr}(dU_{1t}, \beta_{13} \, dU_{3t} + \beta_{14} \, dU_{4t})$$

$$= \frac{\beta_{13} \psi_{13} (\beta_{30} + \beta_{33} U_{3t})^{\gamma_3} + \beta_{14} \psi_{14} (\beta_{40} + \beta_{44} U_{4t})^{\gamma_4}}{\sqrt{\beta_{13}^2 (\beta_{30} + \beta_{33} U_{3t})^{2\gamma_3} + \beta_{14}^2 (\beta_{40} + \beta_{44} U_{4t})^{2\gamma_4}}} \, dt. \tag{4}$$

As a result, the leverage effect is state dependent in these models.

---

[2] For cross-sectional applications see for instance Powell et al. (1989) and Stoker (1993), for time-series applications in term structure models see Ghysels and Ng (1998) and in portfolio allocation see Aït-Sahalia and Brandt (2001).

The classical models of Heston (1993) and Scott (1987) are obtained when U_{2t} and U_{4t} are switched-off. These models proved to be a substantial improvement over the Black and Scholes (1973) specification because of their formulation of volatility as a random persistent process. However, this persistence turned out to be the weakness of the model as well: extreme movements in returns occur more frequently in the observed data than would be implied by the Heston or Scott model calibrated to the data (see, for instance, ABL). These observations prompt us to explore generalizing the Heston and Scott models to better accommodate the data. In particular, we introduce a second stochastic volatility factor. The presence of two volatility factor breaks the aforementioned linkage between tail thickness and volatility persistence.

We have to note that the list of the models considered is by no means exhaustive. For example, Meddahi (2001) proposes to model the diffusion coefficient as a finite order expansion based on the eigenfunctions of the expectations of the state variables. In our notation his model can be represented as

$$\sigma^2(U_{3t}, U_{4t}) = \sum_{0 \leqslant i,j \leqslant p} a_{i,j} E_{3,i}(U_{3t}) E_{4,j}(U_{4t}), \tag{5}$$

where p is the order of expansion. The advantage of such a specification is that it nests all the models we consider here and offers more flexibility in modeling the diffusion term. However, as is often the case with nonparametric specifications, the intuition behind the specification is lost because of the higher-order expansions. The goal of this paper is to compare intuitive and commonly used specifications, therefore we leave nonparametric flexible form alternatives for future research. In the next subsections we consider (1) affine, (2) CEV and (3) logarithmic models, all particular cases of index volatility models.

1.1.1. Affine models

Affine diffusion models are characterized by linearity of the drift and variance functions. Dai and Singleton (2000) discuss the most general specification of such models including the identification and admissibility conditions. We consider a very simple representative of this class by specializing the SIV specification in (1)–(3) to

$$\sigma(u) = \sqrt{u}, \tag{6}$$

$$\gamma_i = 0.5, \quad i = 3, 4. \tag{7}$$

The volatility factors enter additively into the diffusion component specification, as in Engle and Lee (1999). Hence, they could be interpreted as short and long memory components. The long memory (persistent) component should be responsible for the main part of the returns distribution, while the short memory component will accommodate the extreme observations.

1.1.2. CEV models

The SIV model specializes to the CEV class when the volatility function of the affine models (6), is modified to allow γ_i in the range from 0.5 to 1. As extreme cases, the class contains the affine models, when $\gamma_i = 0.5$, and the GARCH diffusion models

(Nelson, 1990), when $\gamma_i = 1$. We do not provide a detailed discussion of the CEV models in this paper because our fit was very close to affine models, and, therefore, did not provide many new insights. Jones (2003) has success with these models, when allowing for volatility induced stationarity ($\gamma_i > 1$) and confronting it with the joint options and the underlying index data set.

1.1.3. Logarithmic models

In logarithmic models, the variance is an exponential function of the factors. We consider the following specializations of SIV (1)–(3):

$$\sigma(u) = \exp(u). \tag{8}$$

We study two different flavors of the logarithmic models, depending on the value of the coefficients γ_i.

When $\gamma_i = 0$, $i = 3, 4$ the volatility factors are described by Ornstein–Uhlenbeck processes. In this case, the drift and variance of the volatility factors are linear functions. Hence, this is a multifactor generalization of the Scott (1987), also known as log-linear, specification.[3] When $\gamma_i = 1$, $i = 3, 4$ we have volatility feedback, a feature which will turn out to be empirically relevant. The key property of interest is that it permits the volatilities of the volatility factors, via the terms $\beta_{33} U_{3t}$ and/or $\beta_{44} U_{4t}$, to be high when the volatility factors themselves are high. These terms are found to be important in Gallant et al. (1999).[4]

The logarithmic models with feedback violate the standard regularity conditions. Therefore, the stochastic integrals and solutions of the SDEs associated with these models are not defined. To remedy this problem one can splice the exponential volatility function in (8) with, essentially, the linear growth condition at the level of volatility so high that it is unlikely to be observed in the U.S. equity index returns. Fig. 6 compares the exponential function in (8) and the actual specification that we use. All the details are relegated to Appendix A.

The model with feedback also has a different volatility domain. As opposed to the affine and log-linear models, where $\sigma(u)$ ranges between zero and infinity, this model has a lower bound equal to $\exp(\beta_{10} - \beta_{13}\beta_{30}/\beta_{33} - \beta_{14}\beta_{40}/\beta_{44})$.[5] While there is no a priori consideration against this on pure modeling ground—after all volatility never reaches zero in practice—the fact that volatility boundary depends on the parameters may lead to nonstandard asymptotic behavior of estimators. Method of moment estimators are less prone to boundary problems than are maximum likelihood estimators.[6]

---

[3] We are not the first to suggest two-factor log-linear SV models, see for instance Alizadeh et al. (2002), Chacko and Viceira (1999), Gallant et al. (1999) and the two-factor GARCH model of Engle and Lee (1999).

[4] We use the terminology "feedback" to refer to a feature of the latent factors. One might argue that it is difficult to label and compare models through latent factor features. However, if we define feedback on the basis of the time variation of the variance of the instantaneous variance of returns, we would find that almost all SV models exhibit feedback. Our terminology is therefore based on distinct features of the latent volatility factors.

[5] When $\gamma_i = 1$, the volatility factors are equal to GARCH diffusion model shifted by β_{i0}/β_{ii}. Since the domain of GARCH diffusion is $[0, \infty)$, the domain of our volatility factors is $[-\beta_{i0}/\beta_{ii}, \infty)$.

[6] For further discussion of ML estimation with boundary parameters, see e.g. van der Vaart (2000).

Nonetheless, we strongly suspect that the densities are smooth enough at the boundary so that the parameter estimates follow the usual $\sqrt{N}$-asymptotics; however, we have been unable to prove this. An alternative strategy to address the boundary issue is to subtract the lower bound from the volatility specification in (8) as this would ensure that the lower bound is equal to zero. However, our estimation results showed that this specification is dominated by the one we consider here in terms of overall fit.

It is perhaps not so surprising that the feedback model has good empirical properties. Intuitively, the second factor not only takes care of the tail behavior, as the jump process does, it also features dynamics that seem appealing for modeling extreme market conditions. Indeed, the process can accommodate (mild) persistence in volatility during high volatility days, and when $\beta_{44} \neq 0$ (assuming the second factor determines tail behavior), the volatility of volatility increases as well. These properties cannot be accomplished by a simple Poisson jump process, which can accommodate tail behavior but not the dynamics of extreme events. It should also be noted that a nice feature of the logarithmic specification is the multiplicative effect of U_{3t} and U_{4t} on the volatility of returns. Neither affine models nor jump processes feature separate factors which scale multiplicatively the Brownian motion W_{1t}. Also, the ability of the exponential function to generate very high volatility values adds additional capability to model market stress. All these properties of logarithmic models facilitate mimicking the short-lived erratic behavior through the second volatility factor.

1.2. Affine jump diffusion models

As will be seen from empirical results, the log-linear models dominate the pure affine diffusion models and are not rejected by statistical tests. We would like to give affine models a fair chance and consider extensions of the Heston SV model via jump components.

As a benchmark, we will consider a constant intensity jump diffusion model. Namely the SV model is augmented by the jump to returns, $U_{1t} = \log P_t$, specified as

$$\mathrm{d}q_{1,t} = J_{1,t}\,\mathrm{d}N_t, \tag{9}$$

where

$$N_t \sim \mathrm{Poi}(\lambda_J), \tag{10}$$

$$J_{1,t} \sim \mathrm{N}(\mu_J, \sigma_J^2), \tag{11}$$

which is added to the affine version of (1) when $U_{4t} \equiv 0$. ABL constrain μ_J to be equal to zero. We estimate both constrained and unconstrained specifications.

EJP consider a jump to the volatility factor U_{3t} as well, namely,

$$\mathrm{d}q_{3,t} = J_{3,t}\,\mathrm{d}N_t, \tag{12}$$

where

$$J_{3,t} \sim \exp(\delta) \tag{13}$$

This specification means that jumps to returns and volatility are driven by the same Poisson process N_t, i.e. jumps occur at the same time. Such a specification allows the

introduction of correlation between jump sizes: the jump to return specification (11) should be replaced by

$$J_{1,t} \sim N(\mu_J + \psi_J J_{3,t}, \sigma_J^2) \tag{14}$$

EJP find this model to be the most successful affine specification in terms of the residual properties and the shape of implied volatilities smile.

1.3. Normalizations and model abbreviations

Some normalizations are needed to achieve identification of the various specifications described in the previous subsection. In the SIV specification (1)–(2) the long-run mean of the drift is simultaneously controlled by α_{10} and α_{20}, while the volatility of the drift volatility is controlled by α_{12} and β_{20}. Therefore, we impose

$$\alpha_{20} = 0, \beta_{20} = 1. \tag{15}$$

By analogy, for the general affine model in (3), (6) and (7) we impose the restrictions:

$$\beta_{10} = 0, \beta_{30} = 0, \beta_{33} = 1, \beta_{40} = 0, \beta_{44} = 1. \tag{16}$$

Finally, for the logarithmic specification (3) and (8) we set

$$\alpha_{30} = 0, \alpha_{40} = 0, \beta_{30} = 1, \beta_{40} = 1. \tag{17}$$

Note, that β_{10} is not equal to zero here, because it controls the long-run mean of the total volatility.

It proves convenient to have acronyms for the various models:

AFF1V means the simplest **AFF**ine **One** **V**olatility factor model appearing in (1)–(3), (6), and (7). This model with constant drift corresponds to the Heston (1993) model.

AFF2V stands for the **AFF**ine **Two** **V**olatility factor model, i.e. the most general model appearing in (1)–(3), (6), and (7). This model augments the previous one with an additional continuous path factor.

AFF1V-J0 represents the simplest **AFF**ine **One** **V**olatility factor model with **J**umps to returns appearing in (1)–(3), (6), and (7) in combination with the Poisson process as specified in (9)–(11) and mean jump size, μ_J, constrained to be equal to **0**. This is the ABL model.

AFF1V-J is the **AFF**ine **One** **V**olatility factor model with **J**umps to returns appearing in (1)–(3), (6), and (7) in combination with the Poisson process as specified in (9)–(11).

AFF1V-JJ is **AFF**ine **One** **V**olatility factor model appearing in (1)–(3), (6), and (7) with the Poisson process specified in (10) driving both **J**umps to volatility in (12) and (13) and **J**umps to returns in (9) and (14). This is the model introduced in Duffie et al. (2000) and estimated by EJP.

LL1V means the simplest **L**og **L**inear **One** **V**olatility factor model with no volatility feedback. This model with constant drift corresponds to the Scott (1987) model.

LL1VF means the **One** **V**olatility factor version of (1)–(3), (8), and $\gamma_i = 1$, $i = 3, 4$ with **F**eedback where $\beta_{14} = 0$ making the second volatility factor irrelevant.

Table 1
Model definitions and minimized chi-squared criterion

	α_{10}	α_{12}	α_{22}	α_{30}	α_{33}	α_{40}	α_{44}	β_{10}	β_{13}	β_{14}	β_{30}	β_{33}	β_{40}	β_{44}	ψ_3	ψ_4	λ_J	μ_J	σ_J	ψ_J	δ	$\chi^2(\hat\rho)$	df	p-value
AFF1V	*	*	*	*	*				*			1			*							20.196	9	0.0167
AFF2V	*	*	*	*	*	*	*		*	*				1		*						13.668	5	0.0179
AFF1V-J0	*	*	*	*	*				*			1			*		5.60		*			18.704	7	0.0092
AFF1V-J	*	*	*	*	*				*			1			*		1.70	*	*			11.213	6	0.0820
AFF1V-JJ	*	*	*	*	*				*			1			*		1.70	*	*	*	*	8.577	4	0.0726
LL1V	*	*	*		*			*	*		1				*							30.062	9	0.4e-04
LL1VF	*	*	*		*			*	*		1	*			*							22.752	8	0.0037
LL2V	*	*	*		*		*	*	*	*	1	*	1		*	*						15.037	6	0.0200
LL2VI	*	*	*		*		*	*	*	*	1	*	1	*	*	*						4.153	5	0.5276
LL2VF	*	*	*		*		*	*	*	*	1	*	1	*	*	*						3.171	4	0.5296

Notes: In this paper we consider the following model specifications:

$$\frac{dP_t}{P_t} = (\alpha_{10} + \alpha_{12} U_{2t})\,dt$$

$$+ \sigma(\beta_{10} + \beta_{13} U_{3t} + \beta_{14} U_{4t})\left(\sqrt{1 - \psi_{13}^2 - \psi_{14}^2}\,dW_{1t} + \psi_{13}\,dW_{3t} + \psi_{14}\,dW_{4t}\right) + (e^{J_{1,t}} - 1)\,dN_t,$$

$$dU_{2t} = \alpha_{22} U_{2t}\,dt + dW_{2t},$$

$$dU_{it} = (\alpha_{i0} + \alpha_{ii} U_{it})\,dt + (\beta_{i0} + \beta_{ii} U_{it})^{\gamma_i}\,dW_{it} + J_{3,t}\,dN_t, \quad i = 3, 4,$$

$$\sigma(u) = \sqrt{u}, \gamma_i = 0.5 \text{ for } \mathbf{AFF},$$

$$\sigma(u) = \exp(u), \gamma_i = 0 \text{ for } \mathbf{LL} \text{ or } \gamma_i = 1 \text{ for } \mathbf{LLF},$$

$$N_t \sim \mathrm{Poi}(\lambda_J),$$

$$J_{3,t} \sim \exp(\delta),$$

$$J_{1,t} \sim N(\mu_J + \psi_J J_{3,t}, \sigma_J^2).$$

*denotes a free parameter; 1 denotes a parameter pinned at unity; blank denotes a parameter set to zero. The reported results are based on a simulation of length $N = 75\,000$ simulated at $1/\Delta = 6048$ steps per year, or, equivalently 24 steps per day with 252 trading days per year.

LL2V is the model (1)–(3), (8), and $\gamma_i = 0$, $i = 3, 4$ meaning **L**og **L**inear, **Two Volatility** Factors without volatility feedback.

LL2VI is the model meaning **L**og **L**inear, **Two Volatility** Factors—the **Intermediate** case—one without volatility feedback and the other one with the feedback, i.e. $\gamma_3 = 0$ and $\gamma_4 = 1$.

LL2VF is the most general model, where the acronym means **L**og **L**inear, **Two Volatility** Factors, which feature **F**eedback via $\beta_{33} \neq 0$ and $\beta_{44} \neq 0$ from the Gallant, Hsu, and Tauchen (1999) exploredvolatility factors to their own volatilities. This is the Gallant et al. (1999) model.

The various models are summarized in Table 1. In what follows, ρ denotes the parameters of the underlying SDE that is to be estimated. For example, for the largest logarithmic specification **LL2VF** the parameter vector is

$$\rho = (\alpha_{10}, \alpha_{12}, \alpha_{22}, \alpha_{33}, \alpha_{44}, \beta_{10}, \beta_{13}, \beta_{14}, \beta_{33}, \beta_{44}, \psi_{13}, \psi_{14}). \tag{18}$$

2. Efficient method of moments

Let $\{y_t\}_{t=-\infty}^{\infty}$, $y_t \in \mathfrak{R}^M$, be a discrete stationary time series. In this paper, $\{y_t\}$ is $100 \times [\log(P_t) - \log(P_{t-1})]$, where P_t is the daily DJIA. When, as here, $\{y_t\}$ comes from a discretely sampled SDE system, then the SDE specification implicitly determines the density $p(y_{t-L}, \ldots, y_t | \rho)$ of a contiguous stretch of length $L + 1$ from $\{y_t\}$, where $\rho \in \mathfrak{R}^{p_\rho}$ is a vector of unknown parameters of the generic diffusion process (1). The fundamental problem that blocks straightforward application of standard statistical methods is that an analytic expression for $p(y_{t-L}, \ldots, y_0 | \rho)$ is not available. (see for instance, Aït-Sahalia, 2002; Elerian et al., 2001, Durham and Gallant, 2000 for further discussion). However, by using simulation, an expectation of the form

$$\mathscr{E}_\rho(g) = \int \cdots \int g(y_{-L}, \ldots, y_0) \, p(y_{-L}, \ldots, y_0 | \rho) \, \mathrm{d} y_{-L} \cdots \mathrm{d} y_0$$

can be computed for given ρ. That is, for given ρ, one can generate a simulation $\{\hat{y}_t\}_{t=1}^{N}$ from the system and put

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

with N large enough that Monte-Carlo error is negligible.

Gallant and Tauchen (1996) propose a minimum chi-squared estimator for ρ in this situation, which they termed the efficient method of moments (EMM) estimator. Being minimum chi-squared, the optimized chi-square criterion can be used to test system adequacy. The moment equations that enter the minimum chi-squared criterion of the EMM estimator are obtained from the score vector $(\partial/\partial\theta) \log f(y_t | x_{t-1}, \theta)$ of an auxiliary model $f(y_t | x_{t-1}, \theta)$ where x_{t-1} is a lagged state vector. The auxiliary model is termed the score generator. Gallant and Long (1997) show that if the score

generator is the SNP density $f_K(y|x, \theta_K)$ described below, then the efficiency of the EMM estimator can be made as close to that of maximum likelihood as desired by taking K large enough. The first step in computing the EMM estimator $\hat{\rho}_n$ is to use the score generator

$$f(y_t|x_{t-1}, \theta) \quad \theta \in \Theta \tag{19}$$

to summarize the data $\{\tilde{y}_t, \tilde{x}_{t-1}\}_{t=1}^n$ by computing the quasimaximum likelihood estimate

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

and the corresponding estimate of the information matrix

$$\tilde{\mathscr{I}}_n = \frac{1}{n} \sum_{t=1}^n \left[\frac{\partial}{\partial \theta} \log f(\tilde{y}_t|\tilde{x}_{t-1}, \tilde{\theta}_n) \right] \left[\frac{\partial}{\partial \theta} \log f(\tilde{y}_t|\tilde{x}_{t-1}, \tilde{\theta}_n) \right]'. \tag{20}$$

Estimator (20) presumes the score generator (19) provides an adequate statistical approximation to the transition density of the data, so that $\{(\partial/\partial \theta) \log f(\tilde{y}_t|\tilde{x}_{t-1}, \tilde{\theta}_n)\}$ is essentially serially uncorrelated. If (19) is not adequate, then one of the more complicated expressions for $\tilde{\mathscr{I}}_n$ set forth in Gallant and Tauchen (1996) must be used, although the EMM estimator is still consistent and asymptotically normal. Define

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

which is computed by averaging over a long simulation

$$m(\rho, \theta) \doteq \frac{1}{N} \sum_{t=1}^N \frac{\partial}{\partial \theta} \log[f(\hat{y}_t|\hat{x}_{t-1}, \theta)]. \tag{21}$$

The EMM estimator is

$$\hat{\rho}_n = \underset{\rho \in \mathfrak{R}^{p_\rho}}{\text{argmin}} \, m'(\rho, \tilde{\theta}_n)(\tilde{\mathscr{I}}_n)^{-1} m(\rho, \tilde{\theta}_n). \tag{22}$$

The estimator is consistent and asymptotically normally distributed with asymptotic distribution given in Gallant and Tauchen (1996). Under the null hypothesis that $p(y_{-L}, \ldots, y_0|\rho)$ is the correct model, n times the minimized value of the objective function is asymptotically chi-squared on $p_\theta - p_\rho$ degrees of freedom where p_θ and p_ρ are, respectively, the lengths of parameter vectors θ and ρ.

The EMM estimation involves simulating continuous path diffusions which has been covered extensively in the literature. We rely on a standard Euler discretization scheme. The simulations involve a sampling frequency with twenty four steps per trading day. The trading day was set equal to $\frac{1}{252}$, therefore the models parameters have annual scaling.

We use a nonstandard approach to simulate the affine diffusions from (1)–(3), (6), and (7).[7] Instead of a naive discretization of U_{3t} and U_{4t}, we first derive the dynamics of $\log U_{3t}$ and $\log U_{4t}$ using the Itô's lemma. Then we apply the Euler scheme to these processes. As is well known, square-root processes require constraints on the coefficients for the processes to stay positive (e.g. Feller, 1971). Given our normalizations in (16), these constraints translate into $\alpha_{i0} > 0.5$ for $i = 3, 4$. If we directly simulate the affine processes these constraints impose numerical burdens, as it becomes hard to take numerical derivatives and even simulate for the borderline cases. When we simulate the log-versions of U_{3t} and U_{4t}, we are not concerned with the positivity of the processes, so we can let the parameters α_{i0} change freely. This manipulation improves the stability of the procedure tremendously. Therefore, although affine diffusions satisfy the standard regularity conditions, we might expect, on this basis, that simulating the log provides some increase in numerical accuracy. This approach is related to the Doss transformation which improves the speed of convergence of simulation-based estimates (Detemple et al., 2002).

We took the following approach with respect to jump component simulation. We opted a profiling approach, where the EMM objective function is optimized with respect to the parameters ρ appearing in (18) and the jump size parameters. Since we focus on a standard Merton type jump process the size distribution is Gaussian and involves two parameters. The jump frequency is drawn from a Poisson process, with its intensity parameter fixed and moved over a grid to appraise the overall fit of the model. The jump process was implemented by drawing durations between jumps from a exponential distribution. When the durations fell inside the discretization interval, the size of the jump was attributed time proportionally to the hourly observations bracketing the jump event. In practice, this scheme is equivalent to the one in Platen and Rebolledo (1985) and hence achieves the same convergence.

The best choice of a moment function to implement simulated method of moments is the score of a auxiliary model that closely approximates the system dynamics where the parameter vector of the auxiliary model is evaluated at its quasi maximum likelihood estimate. The SNP density of Gallant and Tauchen (1989,1992,2003), which is derived as a location-scale transform of an innovation density represented as a Hermite expansion leads to a useful, general purpose auxiliary model. We give a brief description. Here, y_t represents the observed process and, for now, $x_{t-1} = (y_{t-L}, \ldots, y_{t-1})$. We frequently drop the time subscripts and write y and x generically.

If one expands $\sqrt{p(x, y \mid \rho^0)}$ in a Hermite series, that is, expands the square root of the stationary density of system (1) in a Hermite series, and derives the approximation to the transition density $p(y \mid x, \rho^0)$ of the system that corresponds to the truncated expansion, then one obtains an approximating transition density $f_K(y_t \mid x_{t-1})$ that has the form of a location-scale transform

$$y = R_x z + \mu_x \tag{23}$$

---

[7] We are greatful to Michael Johannes for suggesting this.

of an innovation z_t, where R_x is an upper triangular matrix (see Gallant et al., 1997).[8] The density function of the innovation z_t, is

$$h_K(z|x) = \frac{[\mathscr{P}(z,x)]^2 \phi(z)}{\int [\mathscr{P}(u,x)]^2 \phi(u)\,du}, \tag{24}$$

where $\mathscr{P}(z,x)$ is a polynomial in (z,x) of degree K and $\phi(z)$ denotes the multivariate normal density function with dimension M, mean vector zero, and variance–covariance matrix the identity.

It proves convenient to express the polynomial $\mathscr{P}(z,x)$ in a rectangular expansion

$$\mathscr{P}(z,x) = \sum_{|j|=0}^{K_z} \left(\sum_{|i|=0}^{K_x} a_{ij} x^i \right) z^j, \tag{25}$$

where $K = (K_z, K_x)$, i and j are multiindexes, and $|\cdot|$ denotes the degree of an index. Because $[\mathscr{P}(z,x)]^2 / \int [\mathscr{P}(u,x)]^2 \phi(u)\,du$ is a homogeneous function of the coefficients of the polynomial $\mathscr{P}(z,x)$, $\mathscr{P}(z,x)$ can only be determined to within a scalar multiple. To achieve a unique representation, the constant term a_{00} of the polynomial $\mathscr{P}(z,x)$ is put to one. With this normalization, $h_K(z|x)$ has the interpretation of a series expansion whose leading term is the normal density $\phi(z)$ and whose higher-order terms induce departures from normality.

The advantage of a rectangular expansion is that it gives the polynomial $\mathscr{P}(z,x)$ the interpretation of a polynomial in z of degree K_z whose coefficients are polynomials of degree K_x in x. This is useful in applications because putting $K_x = 0$ implies that the innovation density $h_K(z_t|x_{t-1})$ does not depend on x_{t-1} and is therefore homogeneous. That is, if $K_x=0$ none of the moments of the innovation density $h_K(z|x_{t-1})$ will depend on the past. Conversely, if $K_x > 0$, then the shape of the innovation distribution does depend on the history $x_{t-1}=(y_{t-L},\ldots,y_{t-1})$ of the process $\{y_t\}_{t=-\infty}^{\infty}$. In the empirical application we will compare parameter estimates obtained from both homogeneous and heterogeneous scores for certain models.

The location function takes the form of an autoregression $\mu_x = b_0 + \sum_{k=1}^{L_u} B_k y_{t-k}$. Consequently, the density determined by the location-scale transform $y = Rz + \mu_x$ together with the innovation density $h_K(z|x)$ is a Gaussian vector autoregression if $K_z = K_x = 0$. It is a semi-parametric autoregression along the lines of Engle and Gonzales-Rivera (1991) if $K_z > 0$ and $K_x = 0$, and is a fully nonparametric nonlinear process if $K_z > 0$ and $K_x > 0$. The two choices of R_x that have given good results in applications are an ARCH-like moving average specification and a GARCH-like ARMA specification which are discussed in Gallant and Tauchen (1997). In summary, L_u, L_g, and L_r determine the location-scale transformation $y=R_x z_t + \mu_x$ and hence determine the nature of the leading term of the expansion. The number of lags in the location function μ_x is L_u and the number of lags in the scale function R_x is $L_u + L_r$. The number of lags

---

[8] Although R does not depend on x in this derivation, it proves advantageous in applications to allow the scale matrix R_x to depend on x because it reduces the degree K_x required to achieve an adequate approximation to the transition density $p(y|x,\rho^0)$.

that go into the x part of the polynomial $\mathscr{P}(z,x)$ is L_p. The parameters K_z, K_x determine the degree of $\mathscr{P}(z,x)$ and hence the nature of the innovation process $\{z_t\}$.

3. Empirical findings

In a first subsection we cover the estimation of the auxiliary model. The second subsection reports and discusses the estimates. A final subsection discusses reprojection of the factors and their properties.

3.1. Data and auxiliary model

The raw data for analysis consist of 11 717 daily observations January 2, 1953, to July 16, 1999, on the (geometric) percent movement

$$y_t = 100 * [\log(P_t) - \log(P_{t-1})] \tag{26}$$

of the DJIA, P_t. As noted earlier, we use the raw series and do not perform any transformation on the raw data which are plotted in Fig. 1. The first step is to project the data $\{y_t\}$ onto an auxiliary model, which is the SNP model described above. We reserve the first 47 data points for forming lags leaving 11 670 observations, net. The tuning parameters L_u, L_g, L_r, L_p, K_z, and K_x are selected by moving upward along an expansion path using the BIC criterion:

$$\mathrm{BIC} = s_n(\tilde{\theta}) + (1/2)(p_K/n)\log(n),$$

where the objective function $s_n(\theta)$ is given by

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

to guide the search. Models with small values of BIC are preferred.

The expansion path has a tree structure. Rather than examining the full tree, the strategy is to expand first in L_u with $L_g = L_r = L_p = K_z = K_x = 0$ until BIC turns upward. For ARCH-type specifications, we expand L_r with $L_g = L_p = K_z = K_x = 0$, then expand K_z with $K_x = 0$, and lastly L_p and K_x. It is useful to expand in K_z, L_p and K_x at a few intermediate values of L_r because it sometimes happens that the smallest value of BIC lies elsewhere within the tree. For GARCH-type specifications, the strategy is similar: we put $L_g = L_r = 1$, then expand K_z, L_p and K_x as above. We then check $L_g = L_r = 2$. These two are the only GARCH-type specifications considered, which is consistent with standard practice among GARCH practitioners. There is the difficulty that increases in K_x add a plethora of parameters. We control this by restricting the coefficients a_{ij} of the Hermite expansion (25) to be zero when $|j| > 2$ and $|i| \geqslant 1$, which was motivated by inspecting t-statistics on Hermite coefficients of larger models without such restrictions. The net effect of the restrictions is that the Hermite coefficients of (25) are state dependent, i.e. dependent upon x, only up through quadratic terms; the Hermite coefficients of z_j are constant for cubics and higher.
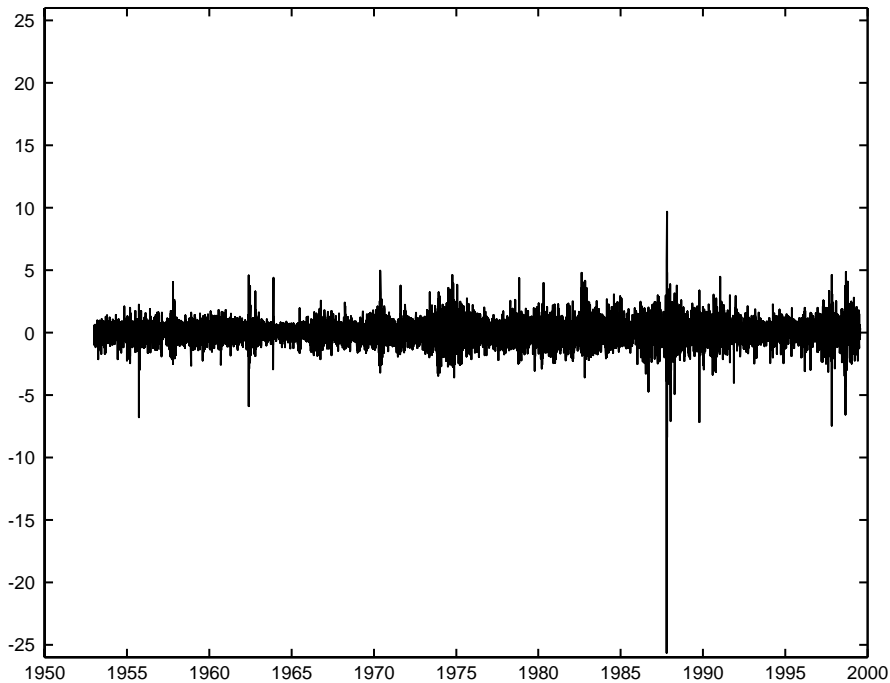
Fig. 1. DJIA 1953–1999.

The final SNP model selected via this procedure has

$$L_u = 1, L_r = 1, L_g = 1, L_p = 1, K_z = 8, K_x = 1. \tag{27}$$

This SNP model, preferred under BIC, can be characterized as a GARCH(1,1) with a nonparametric error density represented as an eighth-degree Hermite expansion where the Hermite coefficients up through quadratic terms are state dependent. The model is akin to the semiparametric GARCH of Engle and Gonzales-Rivera (1991), except their nonparametric error density is represented as a state-independent kernel density. Unlike SNP, the kernel representation of the semiparametric GARCH precludes state dependence of the error density, which is found to be empirically important for this data set.

We generate starting values for the optimization by first estimating the models based on a homogeneous score where the state dependence of the Hermite polynomial is not incorporated, namely the tuning parameters are set to

$$L_u = 1, L_r = 1, L_g = 1, L_p = 1, K_z = 8, K_x = 0. \tag{28}$$

We use the EMM package capability to process a sequence of input parameter files with many randomly perturbed starting values from each input file and, therefore, bad starting values leading to local optima are not a concern. This strategy

yielded satisfactory fits sometime substantially improving results of previous work.

3.2. Estimation results

Table 1 shows the various model specifications along with the minimized value of the EMM objective function appearing in (22), scaled to follow an asymptotic chi-squared on $p_\theta - p_\rho$ degrees of freedom. Tables 3 and 4 report parameter estimates of the affine and logarithmic models, respectively. The parameters correspond to returns expressed in decimal form on a yearly basis. The models diagnostics via t-ratios of individual SNP score elements are provided in Table 5. We start the discussion with the benchmark case of single factor SV models, and then analyze various extensions.

3.2.1. Benchmark case: single factor SV models

It is not surprising that all three single factor SV specifications, **AFF1V**, **LL1V**, and **LL1VF**, are rejected. The t-ratios indicate that because of the misspecification the models can match only some aspects of the returns distribution exemplified by components of the SNP score. All of these models seem to capture the tails of the distribution foregoing matching more intuitively appealing GARCH components. The **AFF1V** model is the most dramatic example: its speed of mean reversion, α_{33} is about 28 times larger than that of logarithmic models, which indicates highly erratic behavior capable of generating extreme tails, but missing the main bulk of the returns distribution.

In fact, the estimated α_{33} is about 40 times larger than the speed coefficient in affine models estimated by ABL and EJP. In order to understand this puzzling result better, we evaluate the **AFF1V** estimation results based on the homogeneous score 11 118 000 appearing in (28). This score mitigates the influence of the tails, and, therefore, a misspecified model should be able to match the GARCH components of the score better.

Table 2 reports the parameter estimates of **AFF1V** based on the homogeneous score 11 118 000. The most interesting feature of Table 2 is that it offers two sets of parameters for one model. The parameter estimates differ particularly with respect to α_{33} measuring the speed of mean reversion in the volatility process. The *intuitive* fit yields estimates with slow mean reversion, i.e. α_{33} equals -3.39, or half-life of 2.5 months, which conforms to the usual empirical findings. However, the *intuitive* fit turns out to be a local minimum of the EMM objective function equal to 31.815, as there is a better fit, which we refer to as the *best* with a lower χ^2 of 17.886 and, unlike previous findings, with very fast mean reversion, which corresponds to volatility half-life of 1.3 days. [9] Panel B of Table 2 shows the EMM t-ratio diagnostics. We learn that the *intuitive* fit violates the moment conditions associated with Hermite polynomial coefficients

---

[9] The better fit was discovered with the help of the heterogenous score, but theoretically it can be found via meticulous grid search of the starting values, so the heterogeneous score is not required for this.

Table 2
Parameter estimates, standard errors and t-ratio diagnostics for the AFF1V model, homogeneous score case

	Intuitive		Best	
	Est	SE	Est	SE
Panel A. Parameter estimates and standard errors				
α_{10}	0.1096	0.0652	0.0989	0.0203
α_{12}	0.8992	1.5165	5.7842	2.3474
α_{22}	−1.1644	3.5986	−44.5889	30.8810
α_{30}	0.8445	0.1460	1.1864	0.0136
α_{33}	−3.3857	0.7865	−132.2989	36.0200
β_{13}	0.0604	0.0115	1.1245	0.0783
ψ_{13}	−0.2786	0.1459	−0.1574	0.0948
	$\chi_6^2 = 31.815$		$\chi_6^2 = 17.886$	

			Intuitive	Best
Panel B. t-ratio diagnostics				
AR	b_0		−1.806	2.725
AR	b_1		1.771	1.159
GARCH	τ_0		2.102	2.335
GARCH	τ_{1a}		0.849	3.352
GARCH	τ_{1g}		1.671	3.111
Hermite	a_{01}		−1.753	2.531
Hermite	a_{02}		2.748	2.767
Hermite	a_{03}		−2.305	2.733
Hermite	a_{04}		2.527	2.786
Hermite	a_{05}		−2.560	0.973
Hermite	a_{06}		2.028	2.903
Hermite	a_{07}		−2.662	0.023
Hermite	a_{08}		1.762	2.670

Notes: Entries to the table show the parameter estimates along with conventional Wald-type standard errors determined by numerical differentiation for one-factor affine models:

$$\frac{dP_t}{P_t} = (\alpha_{10} + \alpha_{12} U_{2t})\,dt + \sqrt{\beta_{13} U_{3t}}\left(\sqrt{1 - \psi_{13}^2}\,dW_{1t} + \psi_{13}\,dW_{3t}\right),$$

$$dU_{2t} = \alpha_{22} U_{2t}\,dt + dW_{2t},$$

$$dU_{3t} = (\alpha_{30} + \alpha_{33} U_{3t})\,dt + \sqrt{U_{3t}}\,dW_{3t}.$$

fitting the tail behavior, whereas the *best* fit fails at mimicking the GARCH volatility persistence moment conditions.

This evidence indicates that there is a dilemma in accommodating at the same time volatility persistence and tail behavior via a single SV factor.[10] The **AFF1V** model combined with the homogenous score can put emphasis either on the persistence in the

---

[10] Meddahi (2001) gives an excellent theoretical discussion of this issue in the framework of the discrete-time **LL1V** model.

Table 3
Parameter estimates and standard errors for affine models

	AFF1V		AFF2V		AFF1V-J0		AFF1V-J		AFF1V-JJ	
	Est	SE	Est	SE	Est	SE	Est	SE	Est	SE
α_{10}	0.1008	0.0184	0.0985	0.0202	0.1070	0.0198	0.08625	0.0236	0.0843	0.0226
α_{12}	4.4309	0.1233	4.2116	1.5694	1.5575	0.4724	3.2190	1.8839	3.1880	1.0081
α_{22}	−27.0252	3.2074	−2.1428	1.4014	−3.0035	1.7006	−10.9593	13.2870	−11.6233	7.1798
α_{30}	1.2241	0.0009	1.1598	0.2894	0.9756	0.1703	1.0197	0.1152	0.7494	0.0826
α_{33}	−137.8742	0.1663	−33.0876	17.4321	−5.7521	1.0712	−2.7898	0.5440	−3.6407	0.8597
α_{40}			1.0655	0.0602						
α_{44}			−92.8250	13.6036						
β_{13}	1.0477	0.0605	0.2975	0.1258	0.0729	0.0125	0.0427	0.0101	0.0742	0.0124
β_{14}			0.0050	0.0344						
ψ_{13}	−0.1991	0.0039	−0.4058	0.0534	−0.4830	0.0144	−0.4762	0.0987	−0.5199	0.0904
ψ_{14}			0.9031	0.0221						
λ_J					5.60		1.70		1.70	
μ_J							−0.0301	0.0023	−0.0089	0.0026
σ_J					0.0213	0.0002	0.0078	0.0010	0.0173	0.0010
ψ_J									−0.8767	2.0354
δ									0.0181	0.0420

Notes: Entries to the table show the parameter estimates along with conventional Wald-type standard errors determined by numerical differentiation for the affine models:

$$\frac{dP_t}{P_t} = (\alpha_{10} + \alpha_{12} U_{2t})\, dt,$$

$$+ \sqrt{\beta_{13} U_{3t} + \beta_{14} U_{4t}}\left(\sqrt{1 - \psi_{13}^2 - \psi_{14}^2}\, dW_{1t} + \psi_{13}\, dW_{3t} + \psi_{14}\, dW_{4t}\right) + (e^{J_{1,t}} - 1)\, dN_t,$$

$$dU_{2t} = \alpha_{22} U_{2t}\, dt + dW_{2t},$$

$$dU_{it} = (\alpha_{i0} + \alpha_{ii} U_{it})\, dt + \sqrt{U_{it}}\, dW_{it} + J_{3,t}\, dN_t, \quad i = 3, 4,$$

$$N_t \sim \text{Poi}(\lambda_J),$$

$$J_{3,t} \sim \exp(\delta), \quad J_{1,t} \sim \text{N}(\mu_J + \psi_J J_{3,t}, \sigma_J^2).$$

Table 4
Parameter estimates and standard errors for logarithmic models

	LL1V		LL1VF		LL2V		LL2VI		LL2VF	
	Est	SE	Est	SE	Est	SE	Est	SE	Est	SE
α_{10}	0.0831	0.0200	0.0841	0.0195	0.0780	0.0337	0.0589	0.0225	0.0674	0.0279
α_{12}	0.6787	0.0513	0.8870	0.1397	0.9833	0.4820	1.1670	0.7872	2.1927	0.7281
α_{22}	−0.6087	0.8540	−1.0531	0.3425	−1.1219	1.0043	−1.1701	2.5137	−7.0195	5.9997
α_{33}	−6.3778	0.9315	−4.3016	0.3601	−0.0041	0.0291	−0.0512	0.0410	−0.1203	0.1227
α_{44}					−74.7610	10.7994	−52.6673	3.1107	−51.3082	8.2119
β_{10}	−2.2882	0.0320	−2.2585	0.0253	−2.0659	1.0294	−2.1969	0.0414	−2.2143	0.0486
β_{13}	1.3708	0.1059	1.2051	0.0410	0.0367	0.0261	0.0863	0.0400	0.1348	0.0695
β_{14}					3.5477	0.5162	2.7688	0.2597	2.7442	0.3130
β_{33}			0.5342	0.1168					0.0408	0.3672
β_{44}							1.9228	0.2260	2.2169	0.3655
ψ_{13}	−0.6482	0.0216	−0.6365	0.0107	−0.3382	0.3285	−0.2966	0.0240	−0.3403	0.1077
ψ_{14}					−0.3538	0.0793	−0.2915	0.0408	−0.2804	0.0564

Notes: Entries to the table show the parameter estimates along with conventional Wald-type standard errors determined by numerical differentiation for logarithmic models:

$$\frac{dP_t}{P_t} = (\alpha_{10} + \alpha_{12}U_{2t})\,dt + \exp(\beta_{10} + \beta_{13}U_{3t} + \beta_{14}U_{4t})(\sqrt{1 - \psi_{13}^2 - \psi_{14}^2}\,dW_{1t} + \psi_{13}\,dW_{3t} + \psi_{14}\,dW_{4t}),$$

$$dU_{2t} = \alpha_{22}U_{2t}\,dt + dW_{2t},$$

$$dU_{it} = \alpha_{ii}U_{it}\,dt + (1 + \beta_{ii}U_{it})^{\gamma_i}\,dW_{it}, \quad i = 3, 4,$$

$$\gamma_i = 0 \text{ for } \textbf{LL} \text{ or } \gamma_i = 1 \text{ for } \textbf{LLF}.$$

Table 5
t-ratio diagnostics

		LL1V	LL1VF	LL2V	LL2VI	LL2VF	AFF1V	AFF2V	AFF1V-J0	AFF1V-J	AFF1V-JJ
AR	b_0	−1.509	−0.081	1.706	0.410	−0.110	1.428	0.912	−3.120	0.566	1.553
AR	b_1	1.593	0.521	0.941	0.011	−0.409	0.397	1.122	0.266	−0.137	0.456
GARCH	τ_0	3.828	3.519	2.364	1.842	1.675	3.580	0.858	2.996	−0.699	−1.884
GARCH	τ_{1a}	2.804	2.641	3.167	1.521	1.761	3.992	1.844	2.079	−0.448	−2.335
GARCH	τ_{1g}	3.664	3.164	2.765	1.708	1.688	3.973	1.483	2.905	−0.553	−2.133
Hermite	a_{10}	0.033	0.665	−0.748	−0.273	−0.177	−0.497	−1.853	−0.919	−1.379	−0.614
Hermite	a_{01}	0.694	0.094	2.117	0.407	0.051	1.085	2.240	−0.837	2.509	0.992
Hermite	a_{11}	1.938	0.531	0.740	−0.473	−0.825	−0.604	0.955	1.623	−0.504	−0.047
Hermite	a_{02}	3.899	3.660	3.719	0.789	−0.034	2.325	2.471	2.106	0.556	1.827
Hermite	a_{12}	−1.189	−0.743	−1.477	−0.841	−0.469	−0.920	−2.653	−1.738	−1.213	−0.591
Hermite	a_{03}	−0.080	−0.233	2.134	0.136	0.040	0.827	1.064	−0.981	2.432	0.562
Hermite	a_{04}	4.313	4.192	4.012	1.106	0.115	1.633	2.758	1.942	0.430	1.173
Hermite	a_{05}	−1.243	−0.790	0.252	−0.451	−0.076	0.582	−0.826	−1.119	1.580	−0.115
Hermite	a_{06}	3.502	3.557	1.587	1.344	0.392	1.597	2.464	1.539	0.075	0.253
Hermite	a_{07}	−1.841	−1.075	−1.095	−0.912	−0.276	0.253	−1.706	−1.103	0.806	−0.343
Hermite	a_{08}	2.528	2.710	1.271	1.079	0.358	1.833	2.168	1.290	−0.420	−0.754

volatility or the tail behavior whereas the heterogeneous score restricts the one factor model to emphasizing the tail behavior only.

3.2.2. Diffusion extensions: multiple SV models

The second SV factor in **AFF2V** leads to a tremendous improvement in capturing the returns dynamics. The *t*-ratios indicate that the model does a good job with the tails and significantly improves the fit for the GARCH components of the score. It is clear that for this specification one SV factor, U_3, is working on the main part of the distribution (notice that its persistence is much higher than that of the **AFF1V** model) and another factor, U_4, is matching the tails. The relative success of this model is evident in the dramatic decrease of the objective function value from 20.196 for **AFF1V** to 13.668. However, the speed of mean reversion of the persistent factor U_3 is still very high, which indicates potential misspecification of the model. Moreover, the loss of degrees of freedom associated with the increase in the number of parameters is not compensated by the decrease in the objective function value. The *p*-values for the two affine models are roughly the same and therefore **AFF2V** is rejected as well.

Note that while single factor models are known to have negative correlation between innovations in volatility and returns, this may not be the case for the two-factor model. The models have been parameterized so that the coefficients ψ_{13} and ψ_{14} come out as correlations. We find in Table 3 a correlation in **AFF1V** equal to −0.19, the correlations of returns with U_3 and U_4 are equal to −0.41 and 0.90, respectively. The leverage effect formula (4) will allow us to understand how these correlations affect overall relationship between returns and volatility.

Specializing the general formula (4) to the case of **AFF2V** we obtain

$$\text{corr}(dU_{1t}, \beta_{13}\,dU_{3t} + \beta_{14}\,dU_{4t}) = \frac{\beta_{13}\psi_{13}\sqrt{U_{3t}} + \beta_{14}\psi_{14}\sqrt{U_{4t}}}{\sqrt{\beta_{13}^2 U_{3t} + \beta_{14}^2 U_{4t}}}\,dt. \tag{29}$$

Since correlation coefficients take opposite signs, the leverage could become positive in certain states. To asses the likelihood of this happening, we first compute the "average" case. Ideally, we would have to compute the unconditional leverage effect. However, we cannot do this analytically. Therefore, we compute something that is very close to the unconditional leverage effect. Namely, we evaluate the conditional leverage effect at the unconditional, or long-run, means of the states, which are equal to α_{i0}/α_{ii}. After substituting values of the parameters and the long-run means into (29) we find the value of leverage effect equal to -0.3971.

In order to investigate a possibility of the positive value of the leverage effect, we consider a very unlikely scenario: factor U_3 is equal to 1/10th of its long-run mean and factor U_4 is equal to 10 times of its long-run mean. This scenario puts a lot of weight on the contribution of the positive correlation coefficient ψ_{14} to the overall leverage effect. In this case, the leverage is equal to -0.3175. So, as the factor U_4 increases, the leverage effect moves into the positive direction, but very slowly. As a result, positive leverage is theoretically feasible, but is highly unlikely.

The second factor in logarithmic models leads to improvements as well. However, since it enters the model multiplicatively, U_4 will work as stochastic volatility of volatility rather than the factor dedicated to the tails of distribution. The **LL2V** is the least successful specification. Despite its p-value of 2% being higher than those of all pure diffusive affine models and all single factor logarithmic models, it is still quite low for the model to be retained as an adequate model. In particular, when one examines the t-ratio diagnostics, they show only marginal improvements over single factor models. Moreover, some of the key model parameters related to the more persistent factor U_3 (α_{33}, β_{13}, and ψ_{13}) are insignificant.

The **LL2VF** specification, which adds feedback to the **LL2V** specification, dominates all diffusion models based on t-ratios, p-values, and objective function values. The estimated parameters clearly indicate extreme persistence, i.e. near unit root discretely sampled, of U_3 and extreme mean reversion, i.e. near white noise discretely sampled, of U_4.

Evaluating **LL2VF** more carefully, one observes that β_{33}, the coefficient controlling the feedback component of U_3 is not significantly different from zero. The factor's persistence parameter, α_{33}, is not significant either. Our last logarithmic specification **LL2VI** explores the possibility of modeling the persistent factor U_3 without feedback. We find that, despite an increase in the objective function, additional degree of freedom leads to p-value very close to that of **LL2VF**. Moreover, all nice t-ratio diagnostics remain intact and α_{33} becomes significant. These results suggest that the introduction of feedback to both SV factors is unnecessary, and **LL2VI** becomes our preferred logarithmic model.

Interestingly, the point estimates of the leverage effect coefficients ψ_{13} and ψ_{14} are -0.30 and -0.29, respectively. Hence, they are both negative and significant. The value

of ψ_{14} is quite a dramatic reversal as compared to **AFF2V**. This is additional evidence that affine and logarithmic specifications work in fundamentally different ways.

Indeed, specializing the general formula (4) to the case of **LL2VI** we obtain

$$\text{corr}(\mathrm{d}U_{1t}, \beta_{13}\,\mathrm{d}U_{3t} + \beta_{14}\,\mathrm{d}U_{4t}) = \frac{\beta_{13}\psi_{13} + \beta_{14}\psi_{14}(1 + \beta_{44}U_{4t})}{\sqrt{\beta_{13}^2 + \beta_{14}^2(1 + \beta_{44}U_{4t})^2}}\,\mathrm{d}t. \tag{30}$$

As we noted earlier in Section 1.1.3, the factor U_4 is a GARCH-diffusion shifted by $1/\beta_{44}$. Therefore, the expression $1 + \beta_{44}U_4$ will always be nonnegative. Since both correlation coefficients are below zero, the leverage effect will be negative in any state. For comparison with the affine model, we compute the conditional leverage effect evaluated at the long-run mean of U_4. It is equal to -0.3006. So, on average, logarithmic models produce a slightly smaller leverage effect.

3.2.3. Jump extensions

Our findings indicate that logarithmic diffusion models overwhelmingly dominate the affine diffusion models mainly because of the multiplicative specification of the volatility. However, we would like to investigate the role of jumps in affine models since they were shown to be important by several authors. This analysis will lead to a fair comparison between affine and logarithmic models.

The first specification, **AFF1V-J0**, offers a somewhat modest improvement in the EMM objective function value as compared to **AFF1V**. Moreover, because of the loss of one degree of freedom the p-value is slightly worse. Turning to the t-ratios we see a mild improvement in the GARCH components of the score. However, looking at the parameter estimates, it becomes clear that the diffusion part of the model becomes much more reasonable. In particular, the persistence of U_3 increases 24-fold. This is not surprising: incorporating jumps provides additional flexibility in fitting the tails of the returns distribution, relieving the volatility factor from this burden. Therefore, the volatility process coefficients are much more closely aligned with the *intuitive* fit in Table 2.

ABL, who study the specification of type (1) find that jump occur about six times per year and on average jump up and down by the same magnitude. Clearly, jump component is fitting the tail behavior. A priori, it seems that a second SV factor can perform the same task. Our results confirm this intuition: **AFF2V** does even a better job than **AFF1V-J0**.

The jumps in **AFF1V-J0** are symmetric (because μ_J was set to zero) and occur 5.6 times per year.[11] Casual observation of the return series in Fig. 1 indicates that this is not the case: there seem to be more negative than positive jumps, and the negative moves tend to more severe. Moreover, the extreme events seem to occur less frequently than six times a year. One way to formally test this is to estimate μ_J as a free parameter. **AFF1V-J** shows a very different picture: the frequency of jumps drops to 1.7 per year and the jump size has a significant negative mean, which is consistent with

---

[11] Because of our profiling approach to the estimation of λ_J (see Section 2), we cannot compute standard errors. However, for all of our jump models the objective function dramatically increased outside of the interval $\hat{\lambda}_J \pm 0.1$. Hence, this could be considered as an informal confidence interval.

negative skewness of returns. The volatility component becomes even more persistent. The improvement occurs not only in realism of the parameters estimates, but in statistic inference as well: t-ratios substantially improve and the model cannot be rejected at 5% confidence level, but is rejected at 10%. Moreover, the properties of the jump-diffusion model become intuitive: volatility is more persistent, jumps are seldom: a little bit less than two per year.

The **AFF1V-JJ** model has an appealing feature that during market stress, when returns jump down, spot volatility jumps up. [12] In combination these two components can produce a market move of a large size. As a result of a jump, volatility deviates from its long-run mean and then mean reversion pulls it back via diffusive movements. Such dynamics are very similar to the one generated by multiplicative volatility factors in **LL2VI**. Despite the intuitive appeal of **AFF1V-JJ**, we cannot distinguish it from a more parsimonious **AFF1V-J** based on the EMM diagnostics. Dramatic decrease in the objective function is not compensated by the degrees-of-freedom loss. As a result, asymptotic p-values are almost identical. The t-ratio diagnostics do not clarify the picture either: while the GARCH part of the auxiliary model seems to be fit worse by **AFF1V-JJ**, we have recently uncovered evidence that the t-ratios might not be reliably estimated when the degrees of freedom are small. Hence, based on the parsimony considerations, we select **AFF1V-J** as our preferred affine model.

3.3. Additional diagnostics via reprojection

Our results up to this point indicate that, intuitively, a pure diffusion model with multiplicative volatility factors, one of which is almost explosive, can generate dynamics similar to that of single volatility model with explicit jump component. The empirical success of the logarithmic specification can intuitively be explained by the fact that the second factor not only accommodates the tails of the (conditional) return distribution, but also accommodates the volatility dynamics during extreme market conditions, since the specification of the second factor is mean reverting with local persistence and state-dependent volatility of volatility. The potential explosiveness of the persistent volatility factor contributes to realistic modeling of the extreme behavior. This finding has potentially very important implications because, if the diffusion model turns out to be better than the jump diffusion previously advocated in the literature, this will have very important simplifying implications for hedging as well as complicating implications for option pricing.

For this reason, we further investigate the best models from each class in order to better understand the differences and commonalities between the two models. We turn therefore our attention to the time-series properties of the volatility factors. Since the factors are latent we use the reprojection method of Gallant and Tauchen (1998). Figs. 2–5 report time-series plots of the Dow Jones returns and volatility factors

---

[12] This behavior is modeled by negative correlation ψ_{lJ}. It is very hard to estimate this parameter with high precision because it measures the correlation of two seldom and unobservable events, therefore the large standard error reported in the table should not be surprising.
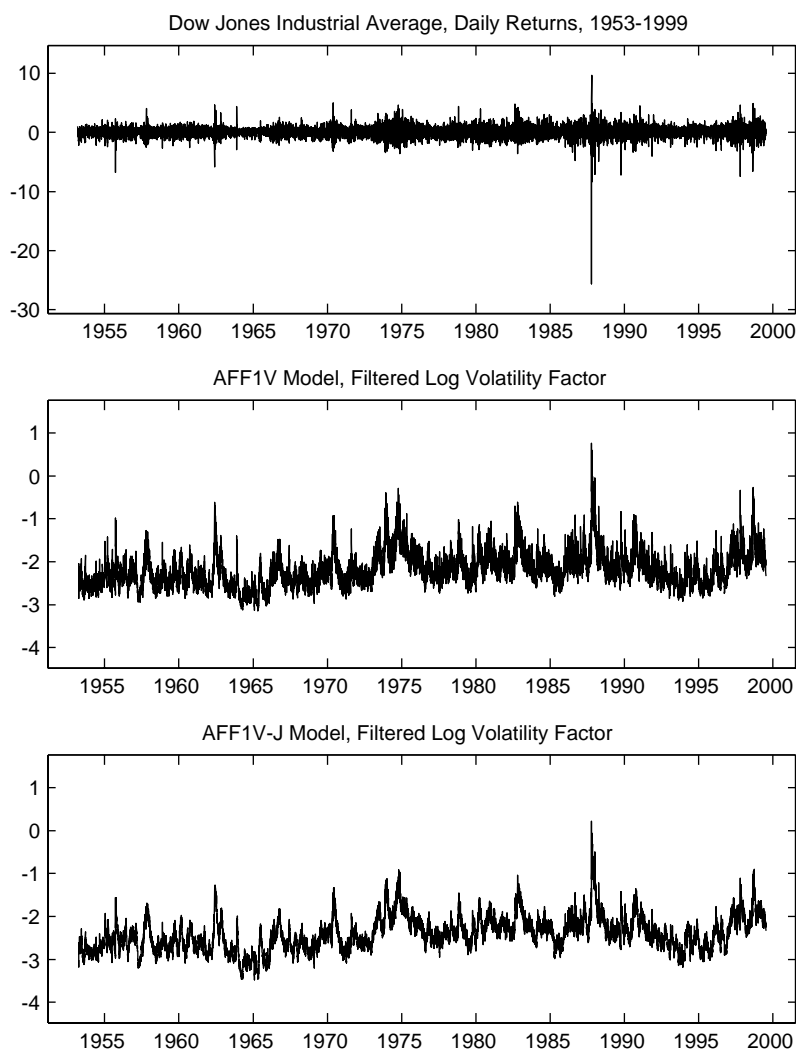
Fig. 2. Reprojection of volatility factors from AFF1V and AFF1V-J models—1953–1999.

reprojected from the single volatility (benchmark) models in affine and logarithmic class as well as from the most successful models in each class.

The first two plots pertain to **AFF1V** and **AFF1V-J** covering sample 1953–1999 (Fig. 2) and a single year, namely 1998 (Fig. 3). Likewise, Figs. 4 and 5 cover **LL1VF** and **LL2VI** for the same samples, i.e. 1953–1999 and 1998. Note that we report the estimates of the logarithm of the affine volatility factors to make them comparable to the volatility factors from the logarithmic models. It is worth noting that when one looks at Figs. 3 and 4 it appears that the model without jumps (AFF1V) creates a more volatile reprojected factor than the model with jumps (AFF1V-J). This is as expected
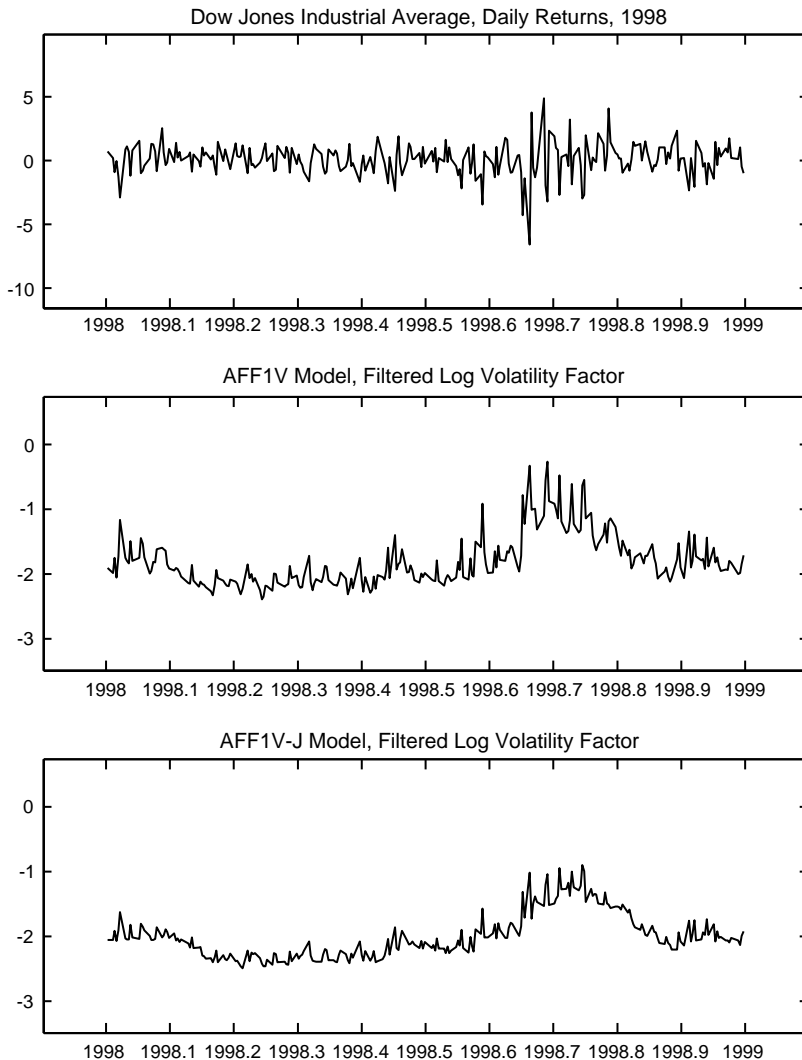
Fig. 3. Reprojection of volatility factors from AFF1V and AFF1V models—1998.

given that the parameter estimates of the model will jumps yielded a more persistent U_3 process. Note that in logarithmic models the volatility level is controlled through one parameter β_{10}, while in affine models it is controlled through α_{i0}/α_{ii}. Since we can not allocate β_{10} between the two volatility factors in **LL2VF**, we report factors scaled by the respective weights, β_{i0}, without regard to the level on all plots. As a result, one can compare the shape and relative size of the volatility, but not the level.

The one-factor models yield reprojected volatilities which look quite different. The **AFF1V** appears to be more erratic, which is consistent with a much lower persistence
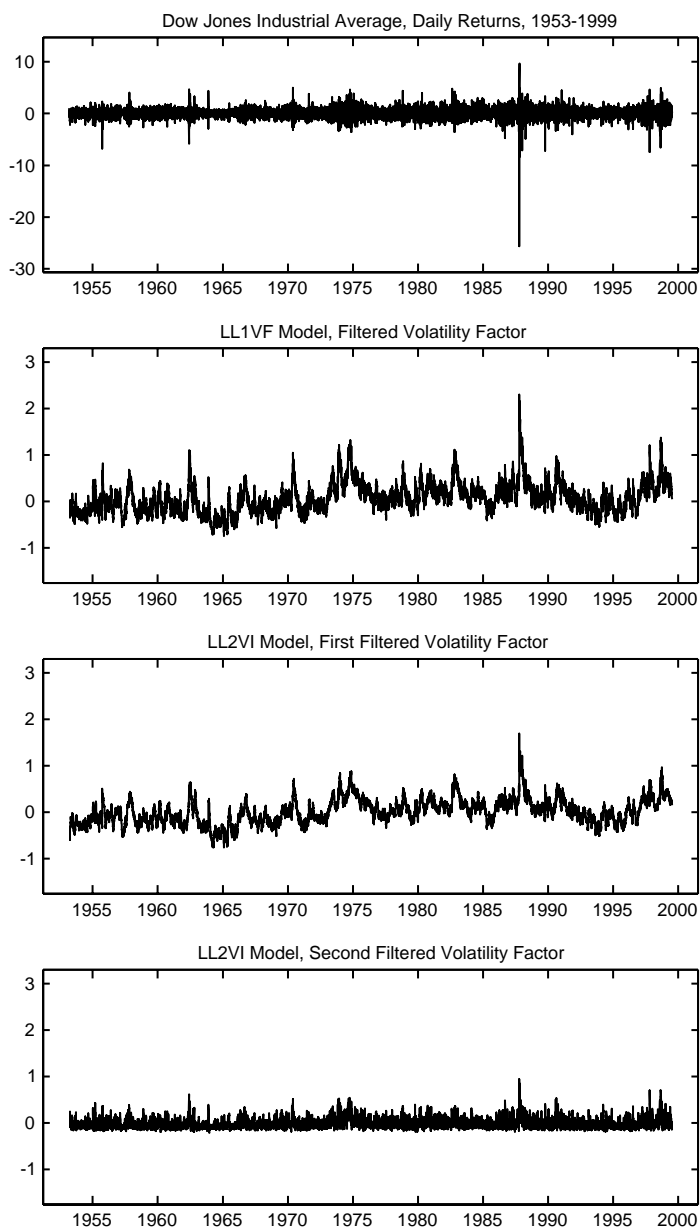
Fig. 4. Reprojection of volatility factors from LL1VF and LL2VI models—1953–1999.

measured by α_{33}. Nonetheless, the overall pattern and the volatility range seems to be close for both models. These plots partially confirm the findings of ABL, who compare the empirical fit of logarithmic and affine volatility models.
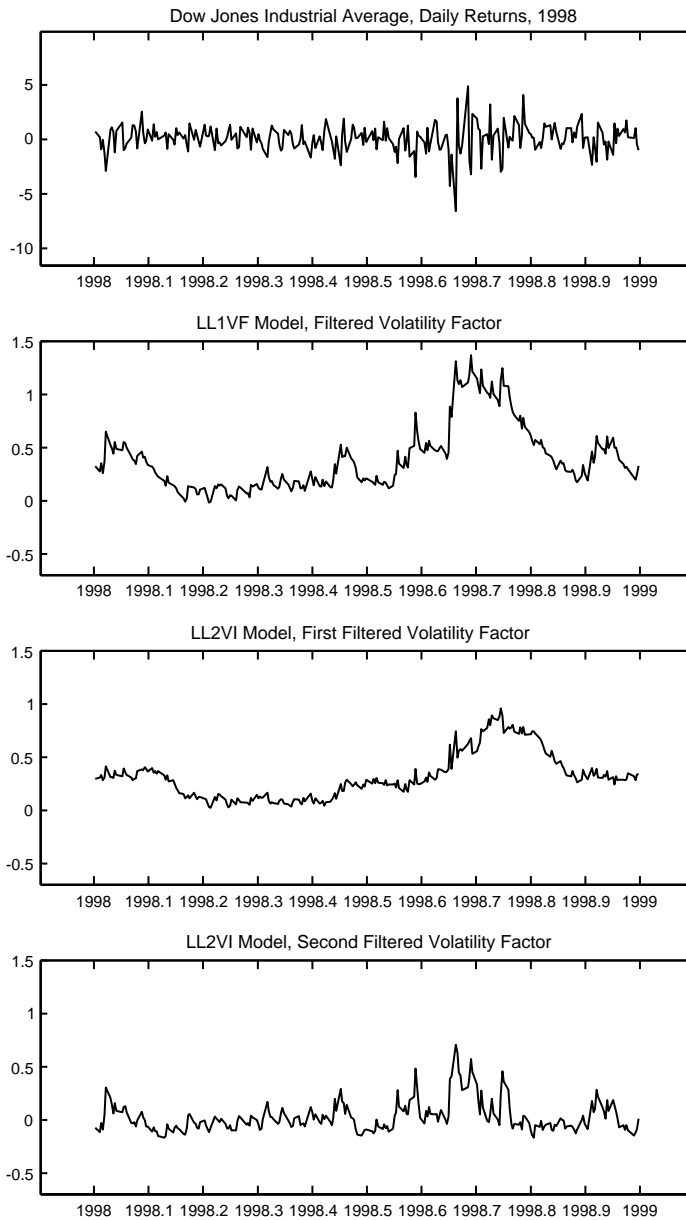
Fig. 5. Reprojection of volatility factors from LL1VF and LL2VI models—1998.

The volatility factors in the preferred specifications resemble each other much better. The persistence, range and level (after taking into an account the value of $\beta_{10} \approx -2.2$) of the first volatility factor, U_{3t}, of the **LL2VI** model is very close to the

only volatility factor of **AFF1V-J**. We plotted a single year from the sample, namely 1998, to highlight that the first factor in **LL2VI** picks up persistence, as does the volatility factor in **AFF1V-J**, (see Figs. 3 and 5), and in fact resembles very much the single factor model reprojected volatilities. The second factor of **LL2VI** looks very different. This is apparent from both the entire sample plot in Fig. 4 as well as the 1998 reprojected volatilities. The second volatility factor U_{4t} clearly behaves like white noise allowing to generate observations in the tails similarly to the jump component of **AFF1V-J**.[13]

The local behavior of **LL2VI** is very different from **AFF1V-J** model, even from a theoretical point of view. For the jump-diffusion model extreme events are represented by the i.i.d. jump process, while the second volatility factor in the **LL2VI** model has a half-life of three and a half-days, meaning that extreme events taper off over several days. This is consistent with Das and Sundaram (1999) who find downward sloping and hump-shaped term structures of higher moments for jump diffusions and pure diffusions, respectively. The model diagnostics that we have used do not pick up the subtle features of the data generating processes. The statistical tools appear unable to discriminate between such features pertaining to events rarely occuring over the entire sample (despite its length).

4. Conclusion

In this paper, we examine various generalizations of SV models via the EMM estimation procedure applied to a sample of post-war Dow Jones daily return series.

We explored and compared the following two-factor specifications (1) a continuous path affine diffusion factor process augmented with a jump component to better fit the tail behavior, (2) a two-factor logarithmic SV specification with possible feedback, the latter causing volatility of volatility to increase, and (3) the two factor affine SV model.

We find that none of the one-factor stochastic volatility specifications fit the data, which confirms previous findings. We note that the asymptotic *p*-value for the affine model is much higher than for logarithmic specification.

The two-factor affine model improves dramatically upon the single volatility model in terms of the EMM objective function value. However, the associated loss of degrees of freedom does not compensate enough: *p*-values of one and two factor affine models are roughly the same. The same conclusion applies to affine model with jumps to returns which have zero mean. However, the model cannot be rejected at 10% confidence level when the constraint on the mean of the jump size is relaxed. Moreover, this simple modification dramatically changes the behavior of both volatility, which becomes more persistent, and jumps, which become less frequent. The jumps to volatility further reduce the objective function value.

---

[13] We notice that the reprojected path of U_{4t} features the local exuberance around the summer of 1998 when LTCM and the Russian financial crisis shook financial markets.

Based on the *p*-values we find that the logarithmic two-factor model specification without feedback is rejected at the 5% significance level, though it dominates all rejected affine models. The most important new finding is that two factor logarithmic specification involving at least one volatility factor with feedback fit the data with the *p*-values of over 50%. Thus, we find that logarithmic factor model with feedback, which has rapidly moving stochastic volatility of volatility, is at par with the affine jump model.

All two-factor specifications feature one factor which accounts for the persistence in volatility and the second determines the tail behavior. The empirical success of the logarithmic specification can intuitively be explained by the fact that the second factor not only accommodates the tails of the (conditional) return distribution, but also accommodates the volatility dynamics during extreme market conditions, since the specification of the second factor is mean-reverting with local persistence and state-dependent volatility of volatility. The near explosiveness of the short memory volatility factor contributes to realistic modeling of the extreme behavior.

Casual observations of the data reveal that abrupt changes in the volatility are an essential ingredient of a successful model. Jumps in returns and volatility simultaneously appear to be the ideal model. Yet, the improvement in statistical fit is not strong enough to justify this conclusion. The statistical tools combined with the data do not allow us to fine-tune the diagnostic any further to declare an overall winner. Additional data, e.g. options data, would help in discriminating the remaining competing models.

An alternative approach is, rather than trying to establish the "ideal" model, to realize that there will always be uncertainty regarding model misspecification and to be robust with respect to it in the spirit of Anderson et al. (2002). However, our model uncertainty does not directly fall into their framework because the model perturbations are not absolutely continuous with respect to each other. Hence, more theoretical work is required before this approach can be implemented in practice.

Appendix A. Regularity conditions for logarithmic models

The use of logarithmic volatility models raises several issues regarding regularity conditions which ensure existence of moments, strong solutions to stochastic difference equations (SDEs), and convergence of discretization schemes. In particular, the stochastic integrals associated with the SDEs of the logarithmic SV models with feedback are not defined in the usual sense (the integrand has to be in L^2, e.g. Kloeden and Platen, 1995, pp. 81–82). [14] The exponential transformation of the volatility factors results in explosive behavior. The explosiveness of the logarithmic SV process has been recognized for a while in the term structure literature. For instance, Brace et al. (1997) replace the continuously compounded rate by the effective annual rate. This removes the exponentiation of a lognormal variable, which in its turn removes fatness in the tail, so that the moments exist.

We have to ensure that solutions to the specified logarithmic SV processes exist and are unique. [15] The processes we consider do satisfy the local Lipshitz conditions, but violate the usual growth conditions in Itô's theorem (Kloeden and Platen, 1995, p. 128). To resolve this problem we splice the exponential volatility function in (8) with appropriate growth conditions at the point, which corresponds to the volatility level unlikely to occur in the U.S. markets, i.e. 150% annualized. [16]

Formally, instead of the volatility specification (8), we estimated

$$\sigma(u) = \begin{cases} \exp(u) & \text{if } u \leqslant u_0 = \log(1.5) \\ \frac{e^{u_0}}{\sqrt{u_0}} \sqrt{u_0 - u_0^2 + u^2} & \text{otherwise} \end{cases} \tag{A.1}$$

where at least one of the γ_i in (3) is equal to one. Now it is clear that we can find a constant K, such that

$$\sigma^2(u) \leqslant K^2(1 + u^2). \tag{A.2}$$

The particular functional form of σ in (A.1) is selected to ensure a smooth splicing, i.e. the two functions and their first derivatives coincide at u_0.

This modification is adequate to ensure the existence of stochastic integrals, SDE solutions, convergence of discretization schemes, and EMM applicability. From the practical perspective, we are effectively considering the exponential form of the volatility function. Fig. 6 compares both specifications.

---

[14] We are greatful to Nour Meddahi for pointing this out to us.

[15] It should be noted that we are discussing sufficient conditions. See Chen et al. (2001) for further discussion.

[16] For comparison, on October 19, 1987 implied 1 month volatility on S& P 100, which is approximately equal to integrated volatility over the whole month, was equal to 150%.
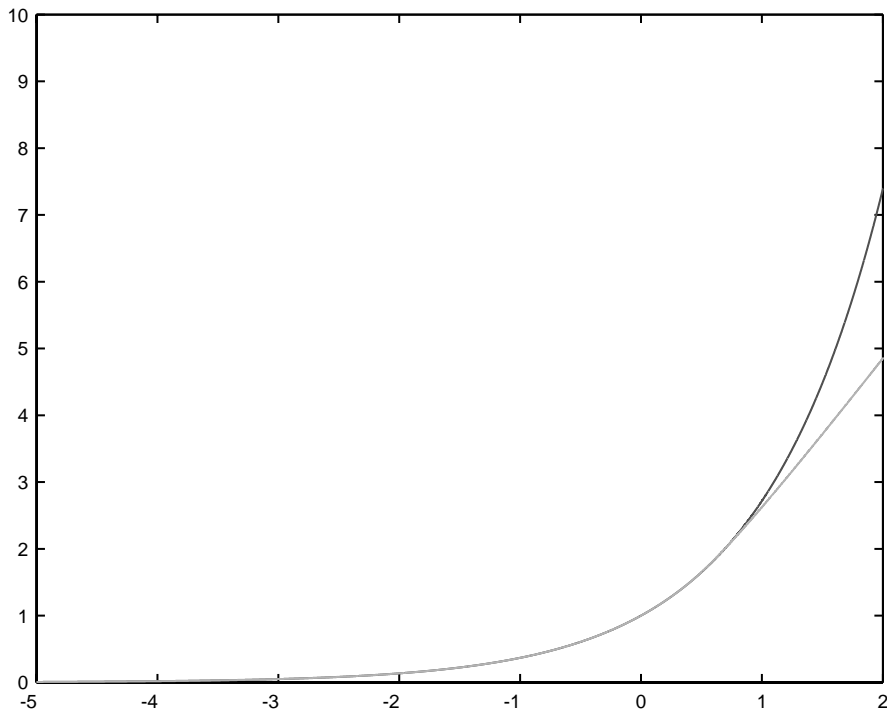
Fig. 6. The functional form of the diffusion coefficient in logarithmic models: spline vs. pure exponential form.

References

Aït-Sahalia, Y., 2002. Maximum likelihood estimation of discretely sampled diffusions: a closed-form approximation approach. Econometrica 70, 223–262.

Aït-Sahalia, Y., Brandt, M., 2001. Variable selection for portfolio choice. Journal of Finance 56, 1297–1351.

Alizadeh, S., Brandt, M.W., Diebold, F.X., 2002. Range-based estimation of stochastic volatility models. Journal of Finance 57, 1047–1091.

Andersen, T., Benzoni, L., Lund, J., 2002. An empirical investigation of continuous-time equity return models. Journal of Finance 57, 1239–1284.

Anderson, E., Hansen, L.P., Sargent, T., 2002. A quartet of semi-groups for model specification, detection, robustness, and the price of risk. Working Paper, University of North Carolina.

Bakshi, G., Cao, C., Chen, Z., 1997. Empirical performance of alternative option pricing models. Journal of Finance 52, 2003–2049.

Bates, D., 2000. Post-'87 crash fears in the S&P 500 futures option market. Journal of Econometrics 94, 181–238.

Black, F., Scholes, M., 1973. The pricing of options and corporate liabilities. Journal of Political Economy 81, 637–654.

Brace, A., Gatarek, D., Musiela, M., 1997. The market model of interest rate dynamics. Mathematical Finance 7, 127–155.

Chacko, G., Viceira, L., 1999. Spectral GMM estimation of continuous-time processes. Discussion Paper, Harvard Business School.

Chen, X., Hansen, L.P., Carrasco, M., 2001. Nonlinearity and temporal dependence. Discussion Paper, University of Chicago.

Chernov, M., Ghysels, E., 2000. A study towards a unified approach to the joint estimation of objective and risk neutral measures for the purpose of options valuation. Journal of Financial Economics 56, 407–458.

Dai, Q., Singleton, K., 2000. Specification analysis of affine term structure models. Journal of Finance 55, 1943–1978.

Das, S., Sundaram, R., 1999. Of smiles and smirks: a term-structure perspective. Journal of Financial and Quantitative Analysis 34, 211–240.

Detemple, J., Garcia, R., Rindisbacher, M., 2002. Asymptotic properties of Monte Carlo estimators of diffusion processes. Working Paper, CIRANO.

Duffie, D., Pan, J., Singleton, K., 2000. Transform analysis and option pricing for affine jump-diffusions. Econometrica 68, 1343–1377.

Durham, G., Gallant, A.R., 2000. Numerical techniques for maximum likelihood estimation of continuous-time diffusion processes. Working Paper, University of North Carolina.

Elerian, O., Chib, S., Shephard, N., 2001. Likelihood inference for discretely observed non-linear diffusions. Econometrica 69, 959–993.

Engle, R.F., Gonzales-Rivera, G., 1991. Semi-parametric ARCH models. Journal of Business and Economic Statistics 9, 345–359.

Engle, R.F., Lee, G., 1999. A long-run and short-run component model of stock return volatility. In: Engle, R.F., White, H. (Eds.) Cointegration, Causality and Forecasting—A Festschrift in Honour of Clive W.J. Granger, Oxford University Press, Oxford.

Eraker, B., Johannes, M., Polson, N., 2003. The impact of jumps in equity index volatility and returns. Journal of Finance, forthcoming.

Feller, W., 1971. An Introduction to Probability Theory and its Applications, Vol. II. Wiley, New York.

Gallant, A.R., Hsieh, D., Tauchen, G., 1997. Estimation of stochastic volatility models with diagnostics. Journal of Econometrics 81, 159–192.

Gallant, A.R., Hsu, C.-T., Tauchen, G., 1999. Using daily range data to calibrate volatility diffusions and extract the forward integrated variance. Review of Economics and Statistics, forthcoming.

Gallant, A.R., Long, J.R., 1997. Estimating stochastic differential equations efficiently by minimum chi-square. Biometrika 84, 125–141.

Gallant, A.R., Tauchen, G., 1989. Seminonparametric estimation of conditionally constrained heterogeneous processes: asset pricing applications. Econometrica 57, 1091–1120.

Gallant, A., Tauchen, G., 1992. A nonparametric approach to nonlinear time series analysis: estimation and simulation, in: Parzen, E., Brillinger, D., Rosenblatt, M., Taqqu, M., Geweke, J., Caines, P. (eds.), New Dimensions in Time Series Analysis. Springer-Verlag, New York.

Gallant, A.R., Tauchen, G., 1993. SNP: a program for nonparametric time series analysis, version 8.3, User's Guide. Discussion Paper, University of North Carolina at Chapel Hill.

Gallant, A.R., Tauchen, G., 1996. Which moments to match? Econometric Theory 12, 657–681.

Gallant, A.R., Tauchen, G., 1997. EMM: a program for efficient method of moments estimation, Version 1.4, User's guide. Discussion Paper, University of North Carolina at Chapel Hill.

Gallant, A.R., Tauchen, G., 1998. Reprojecting partially observed systems with application to interest rate diffusions. Journal of American Statistical Association 93, 10–24.

Gallant, A.R., Tauchen, G., 2003. SNP: a program for nonparametric time series analysis, a user's guide. Manuscript, University of North Carolina. (Available along with code via http://www.econ.duke.edu/~get./snp.html).

Ghysels, E., Ng, S., 1998. A semiparametric factor model of interest rates and tests of the affine term structure. Review of Economics and Statistics 80, 535–548.

Heston, S.L., 1993. A closed-form solution for options with stochastic volatility with applications to bond and currency options. Review of Financial Studies 6, 327–343.

Jones, C., 2003. The dynamics of stochastic volatility. Journal of Econometrics, this issue.

Kloeden, P.E., Platen, E., 1995. Numerical Solution of Stochastic Differential Equations. Springer, Berlin.

Meddahi, N., 2001. An eigenfunction approach for volatility modeling. Working Paper, University of Montreal.

Nelson, D., 1990. ARCH models as diffusion approximations. Journal of Econometrics 45, 7–38.

Pan, J., 2002. The jump-risk premia implicit in options: evidence from an integrated time-series study. Journal of Financial Economics 63, 3–50.

Platen, E., Rebolledo, R., 1985. Weak convergence of semimartingales and discretisation methods. Stochastic Processes and their Applications 20, 41–58.

Powell, J., Stock, J., Stoker, T., 1989. Semiparametric estimation of index models. Econometrica 57, 1403–1430.

Scott, L., 1987. Option pricing when the variance changes randomly: theory, estimation and an application. Journal of Financial and Quantitative Analysis 22, 419–438.

Stoker, T., 1993. Lectures on Semiparametric Econometrics. CORE Lecture Series.

van der Vaart, A., 2000. Asymptotic Statistics. Cambridge University Press, Cambridge, MA.