Volatility, Momentum, and Time-Varying Skewness in Foreign Exchange Returns

Timothy C. JOHNSON
London Business School, Sussex Place, Regent's Park, London NW1 4SA, United Kingdom
(*tjohnson@london.edu*)

This article tests a stochastic volatility model of exchange rates that links both the level of volatility and its instantaneous covariance with returns to pathwise properties of the currency. In particular, the model implies that the return–volatility covariance behaves like a weighted average of recent returns and hence switches signs according to the direction of trends in the data. This implies that the skewness of the finite-horizon return distribution likewise switches sign, leading to time-varying implied volatility "smiles" in options prices. The model is fit and assessed using Bayesian techniques. Some previously reported volatility results are accounted for by the fitted models. The predicted pattern of skewness dynamics accords well with that found in historical options prices.

KEY WORDS: Conditional skewness; Exchange rate dynamics; Stochastic volatility.

1. INTRODUCTION

This article tests a new stochastic volatility model that captures some complex and unrecognized pathwise properties of higher moment dynamics in exchange rate returns. Motivation for the model comes from the observed behavior of volatility "smiles" implied in the prices of traded options. These curves—which may be viewed as transformations of an implicit probability distribution for returns to a given horizon—can change radically over time. Recently, Campa, Chang, and Reider (1998) documented an extraordinary and robust regularity of these changes: in the case of options on five different currencies, the skewness of the distribution increases in the direction of realized trends in the underlying exchange rate. That is, tails of the distribution appear to systematically grow and shrink in tandem with the currency's momentum.

In terms of a standard continuous-time model of stochastic volatility, the skewness of finite-horizon returns is governed entirely by the instantaneous correlation between volatility innovations and returns. A positive correlation implies that future positive returns will be associated with higher volatility than future negative returns, making large positive moves over, say, a month more likely than similar down moves—hence a fatter right tail. The evidence from exchange rates thus suggests that this correlation cannot be constant and is, in fact, itself correlated with returns.

There is an alternative preference-based explanation for the options evidence, however. The skewness implicit in options prices could result from hedging demands unconnected with the true distribution of returns. Intuitively, fat tails may simply represent a "crash insurance" premium in one direction. One could imagine that, for example, after a strong appreciation of the yen, real adjustment costs of a still further strengthening could be large but would be altogether avoided if the move were reversed. Technically, such a story requires that the feared outcome is uninsurable by trading in the currency alone, for example, due to jumps or crashes. But there is convincing evidence (Jackwerth and Rubenstein 2000) that the persistent left-skewness in equity index options is just such an insurance premium, unrelated to any skewness in stock returns. The currency analog is a somewhat more complicated but nevertheless plausible hypothesis.

Evidence against the risk hypothesis, however, can be found in the actual behavior of exchange rates. If there were no changes in the actual return–volatility correlation (hence in skewness) in response to trends, then the occurrence of such a trend would convey no information about current volatility. But if the correlation had been increasing in the direction of the momentum, then the successive shocks comprising the trend would be associated with increasingly positive shocks to volatility. Precisely this effect is found in Müller et al. (1997).

Those authors report strong support for a time-series model of daily volatility which allows long-term returns—that is, trends—to affect volatility levels, in addition to the effects of the individual lagged daily returns. For all currencies examined and in all time periods, long-term returns do, indeed, enter significantly and positively. In terms of predictive power, this specification dominates the corresponding generalized autoregressive conditional heteroscedasticity (GARCH) alternative in which the realized autocorrelation of returns plays no role (Dacorogna et al. 1998).

The message of that work may be summarized as the following stylized fact: volatility *increases* when trends *continue*, irrespective of the realized volatility along the path. This is equivalent to the dynamic observed in options prices: given that a trend has occurred, there is more skewness in the same direction because a continuation of the trend will be associated with greater volatility than a reversal. That this pattern is found in the returns themselves suggests that one need not invoke time-varying preference effects to account for the observed "smile" behavior.

Johnson (in press) presents a continuous-time model that allows volatility to depend on trends and other pathwise properties of returns. There the stochastic behavior is derived from a theory of evolving inferences by investors about the degree of persistence of exogenous shocks. Positively auto-

Journal of Business & Economic Statistics
July 2002, Vol. 20, No. 3
DOI 10.1198/073500102288618522

correlated shocks are more likely *ex post* to have been persistent and hence to have greater long-run impact in valuation, which translates into higher current volatility. As well as offering a potential structural explanation for the currency findings, the model provides a full theory of conditional dynamics for all moments. On a purely practical level, the continuous-time setting also affords a straightforward and consistent framework for valuing derivatives. Yet the model is tightly parameterized, because of the simplified economic mechanisms underlying it. Testing whether it can account for the observed behavior of exchange rate volatility despite its restrictions is the goal of this article.

Overall the model is quite successful. Structural parameters estimated from the implied return process do indeed deliver a trend–volatility relationship consistent with the evidence from options markets, while also reproducing the main features of first- and second-moment dynamics. It should be emphasized, however, that the specification examined here is not optimized for descriptive flexibility. The aim is not to introduce yet another competing class of conditional heteroscedasticity models.

Rather, the main contribution of the article is to highlight a curious and robust feature of exchange rate dynamics and to suggest an economic explanation. Options markets do tell us that stochastic volatility alone cannot account for the rich behavior of conditional distributions. In documenting and modeling regularities in higher moments, the article contributes to the increasing sophistication of models available to financial engineers. In testing a structural model, it links the stochastic volatility literature to the growing branch of asset pricing research that seeks to understand the underlying causes of observed patterns of higher-moment predictability.

The outline of the article is as follows. The next section describes the exchange rate model and the steps needed to operationalize it. In the third section the model is estimated with two samples of daily returns, and its goodness of fit is assessed. The trend effect in volatility is taken up in Section 4. The fitted models are shown to reproduce the previously reported phenomena, and a further predicted pattern, heretofore undocumented, is also found in historical options data. The article concludes in Section 5 with a discussion of some of the consequences of the volatility–momentum relationship.

2. THE MODEL

This article uses a continuous-time diffusion model of exchange rates, which may be written

$$de_t = \tilde{\mu}(X_t)\,dt + \tilde{h}(X_t)\,d\widetilde{W}_t, \tag{1}$$

where e_t is the log exchange rate, $\widetilde{W}_t$ is a standard Wiener process, and X_t is a five-dimensional vector of state variables. The specification comes from a structural model relating prices to underlying fundamentals when agents cannot fully observe the state of the economy. (The tilde notation, used throughout, denotes quantities estimated contemporaneously by investors.) For testing this specification, the economics of that model are in the background. However, before giving the governing equations for the state variables explicitly, this section first summarizes the intuition and assumptions of the derivation in order to facilitate interpretation. For details the interested reader is referred to Johnson (in press).

2.1 Unobservable Persistence

An initial assumption of the underlying model is that exchange rates are determined by an equation of the form

$$e_t = \theta_0 + \theta_1 \Delta_t + \theta_2 \widetilde{\mathrm{E}}_t \mu_t. \tag{2}$$

Here the θ are constants, Δ is a (scalar) exogenous driving variable, μ is the drift of Δ, and $\widetilde{\mathrm{E}}_t$ denotes expectation with respect to information contained in the history of Δ prior to t. For empirical purposes it turns out to be unnecessary to identify the exogenous "fundamental" process (as discussed in the following section), but under standard assumptions, one can derive (2) in a two-country monetary economy, Δ being the difference in log money supplies (per unit consumption) between the countries. As the notation suggests, the drift rate μ of this series is assumed to be time-varying and not directly observable. The primary implications of the model all emerge from the evolution of agents' inferences about this rate.

The fundamental variable Δ is assumed to be subject to two *types* of shocks: persistent or transient. This trait is determined by a first-order Markov process, S_t, which is also unobservable. Shocks that are persistent affect Δ_t in all future periods, as well as instantaneously, by altering the growth rate μ. Thus observers update their estimates of that quantity in accordance with their beliefs about the current state of S. If its two states are taken to be $\{0, 1\}$, with 1 corresponding to the persistent-shock state, then $\widetilde{S}_t$, the expectation of S at t, is also the probability that current shocks are persistent. Solving the model entails deriving the full joint conditional distribution of beliefs about μ and S given the history of observations about Δ.

Formally, the system and observation equations of the model are

$$\begin{aligned} d\Delta_t &= \mu_t\,dt + \sigma_0\,dW_t \\ d\mu_t &= S_t\sigma_0\,dW_t \\ dS_t &= (\lambda_0 + \lambda_1)(\overline{S} - S_t)\,dt + (1 - 2S_t)\,dN_t^{\lambda}, \end{aligned} \tag{3}$$

where N_t^{λ} is a compensated Poisson process with intensity $\lambda = \lambda(S) = \lambda_1 S + \lambda_0(1 - S)$ and $\overline{S} = \lambda_0/(\lambda_0 + \lambda_1)$. The third equation thus describes a two-state Markov process whose transitions are governed by λ_0 and λ_1. Heuristically, the probability of a switch from 0 to 1 in time dt is $\lambda_0\,dt$, and that from 1 to 0 is $\lambda_1\,dt$. More precisely, λ_i^{-1} is the expected duration of $S = i$ episodes.

Note that in the first equation Δ itself is taken to be homoscedastic: W is a Wiener process and σ_0 is a constant. All variations in exchange rate volatility will thus be endogenous in this model and not driven by a stochastic information flow.

Finally, the second equation implements the idea that fundamental news will have long-lasting effects when $S = 1$ but not otherwise. This equation is unnecessarily restrictive in the sense that allowing for additional random shocks to growth rates or allowing multipliers other than just 0 and 1

for the shocks does not alter the main dynamic properties of the model. The present formulation is simply the most parsimonious one that captures the desired dichotomous quality of innovations. Only three free parameters (σ_0, λ_0, and λ_1) characterize the behavior of the system and hence inferences about it.

The approach here is closely related to that of Detemple (1986, 1991), Feldman (1989), Wang (1993), Veronesi (2000, 1999), and David (1997). These authors also introduce inferential uncertainty into asset pricing problems, in the form of an unobservable growth rate of fundamentals, and thence derive endogenous volatility processes. The present model complements these works in describing the effects of a different type of parameter uncertainty, which corresponds to a realistic and economically important problem. Intuitively, it seems obvious that shocks to growth rates vary in their temporal impact. [And, in fact, Evans and Lewis (1995) report evidence of variable persistence of shocks to U.S. interest rates, and Evans (1998) finds support for a switching model of dividend growth rates that includes shifts in persistence between regimes.] In a monetary context, a natural interpretation of an $S > 0$ (persistent shock) regime is as a period of changing policy. A monetary contraction in this case is accompanied by a commitment to lower monetary growth in the future. Otherwise ($S = 0$), reserves are withdrawn but there is no perceived implication of continued tightening.

In highlighting the importance of persistence in determining return volatility, the model builds on the work of Barsky and DeLong (1993). Those authors observed that a static version of essentially the same inferential problem—determining the persistence of growth rate changes—could explain the level of stock market volatility. In essence, the system above simply frames the same problem in an evolving context: the right answer varies through time. The result is a theory of volatility changes. A related problem also appears in Barberis et al. (1998). In that paper inferences about persistence also change through time, although there the observer is deluded: the true process does not switch. The authors are interested in the (first-moment) implications of inferential mistakes. By contrast, in the unobservable persistence model, agents are assumed to be fully rational.

2.2 Return Dynamics

The information structure described above, together with the exchange rate equation (2), determine the dynamics of returns in terms of agents' beliefs about the unobservable parameters. The key quantities are the conditional moments, $\widetilde{\mathrm{E}}_t(\mu^i S^j)$, of the joint distribution, each of which is updated according to a filtering equation driven by the "news,"

$$d\widetilde{W}_t \equiv (d\Delta_t - \tilde{\mu}_t\, dt)/\sigma_0, \tag{4}$$

which is just the unexpected change in the exogenous fundamental process.

The exact filter turns out to be infinite dimensional: investors' beliefs cannot be summarized by a few sufficient statistics, as in, for example, Kalman filtering. The equation governing $\tilde{\mu}_t$ (the conditional mean of μ) depends on the conditional variance of μ and on its covariance with S, each of which depends on still higher moments. However, as shown in Appendix A, the full system is well approximated by a five-dimensional subsystem in which beliefs about the parameters are fully summarized by $\widetilde{S}_t$ (the probability that $S = 1$ at t) and the means and variances for μ conditional on the S-state. This simply corresponds to describing current knowledge of the growth rate by a binomial mixture of two symmetric distributions, where the mixing weight is $\widetilde{S}_t$.

Under this auxiliary approximation, the moments evolve according to the following degenerate system of stochastic differential equations:

$$d\widetilde{S}_t = (\lambda_0 + \lambda_1)(\overline{S} - \widetilde{S}_t)\, dt + \widetilde{S}_t \widetilde{R}_t \big(\tilde{m}_t^{(1)} - \tilde{m}_t^{(0)}\big)\, \frac{d\widetilde{W}_t}{\sigma_0} \tag{5}$$

$$d\tilde{m}_t^{(1)} = -\widetilde{R}_t\big(\tilde{m}_t^{(1)} - \tilde{m}_t^{(0)}\big)\left[\frac{\lambda_0}{\widetilde{S}_t} + \frac{(\sigma_0^2 + \tilde{v}_t^{(1)})}{\sigma_0^2}\right] 1_{\{\widetilde{S}_t > 0\}}\, dt + \big(\sigma_0^2 + \tilde{v}_t^{(1)}\big)\frac{d\widetilde{W}_t}{\sigma_0} \tag{6}$$

$$d\tilde{m}_t^{(0)} = -\widetilde{S}_t\big(\tilde{m}_t^{(1)} - \tilde{m}_t^{(0)}\big)\left[\frac{\lambda_1}{\widetilde{R}_t} + \frac{\tilde{v}_t^{(0)}}{\sigma_0^2}\right] 1_{\{\widetilde{R}_t > 0\}}\, dt + \tilde{v}_t^{(0)}\frac{d\widetilde{W}_t}{\sigma_0} \tag{7}$$

$$d\tilde{v}_t^{(1)} = \left\{\sigma_0^2 - \frac{(\sigma_0^2 + \tilde{v}_t^{(1)})^2}{\sigma_0^2} + \lambda_0 \frac{\widetilde{R}_t}{\widetilde{S}_t}\big[\big(\tilde{v}_t^{(0)} - \tilde{v}_t^{(1)}\big) + \big(\tilde{m}_t^{(1)} - \tilde{m}_t^{(0)}\big)^2\big] 1_{\{\widetilde{S}_t > 0\}}\right\} dt \tag{8}$$

$$d\tilde{v}_t^{(0)} = \left\{-\frac{(\tilde{v}_t^{(0)})^2}{\sigma_0^2} + \lambda_1 \frac{\widetilde{S}_t}{\widetilde{R}_t}\big[\big(\tilde{v}_t^{(1)} - \tilde{v}_t^{(0)}\big) + \big(\tilde{m}_t^{(1)} - \tilde{m}_t^{(0)}\big)^2\big] 1_{\{\widetilde{R}_t > 0\}}\right\} dt. \tag{9}$$

Here $\tilde{m}_t^{(i)}$ and $\tilde{v}_t^{(i)}$ are the (approximate) mean and variance of μ_t conditional on $S_t = i$, $\widetilde{R}_t \equiv 1 - \widetilde{S}_t$, and $1_{\{\bullet\}}$ denotes the indicator function, which is 1 when its argument is true and 0 otherwise.

The five moments given by (5)–(9) comprise the state vector [denoted X in Eq. (1)] governing the economy. In terms of them, exchange rate dynamics are given by

$$de_t = \tilde{\mu}\, dt + \tilde{h}\, d\widetilde{W}_t \tag{10}$$

$$\tilde{\mu}_t = \widetilde{S}_t \tilde{m}_t^{(1)} + (1 - \widetilde{S}_t)\tilde{m}_t^{(0)} \tag{11}$$

$$\tilde{h}_t = \sigma_0 + \theta_2\big(\sigma_0^{-1}\widetilde{\mathrm{var}}_t(\mu) + \sigma_0 \widetilde{S}_t\big) \tag{12}$$

$$\widetilde{\mathrm{var}}_t(\mu) = \widetilde{S}_t \tilde{v}_t^{(1)} + (1 - \widetilde{S}_t)\tilde{v}_t^{(0)} + \widetilde{S}_t(1 - \widetilde{S}_t)\big(\tilde{m}_t^{(1)} - \tilde{m}_t^{(0)}\big)^2. \tag{13}$$

Equations (10)–(13) together with (5)–(9) constitute the empirical specification tested in this article. Notice that the model no longer makes any reference to the exogenous time series Δ_t. The innovation process $\widetilde{W}_t$, which drives the system, was defined via that series. However, from Equation (10),

$\widetilde{W}_t$ may equivalently be viewed as the process of unexpected changes in the exchange rate itself.

This has the important consequence that we, the econometricians, can completely recover the entire path of the agents' beliefs about fundamentals (conditional on particular parameter values) without ever observing—or even identifying—that series.

Since those beliefs make up the latent state variables of our model, this affords a simple recursive method for computation of the discretized likelihood function. Given a parameter vector (including start-up values for $\tilde{m}^{(i)}$, $\tilde{v}^{(i)}$, and $\widetilde{S}$), we loop through the data, performing the following steps at each time t:

1. Calculate $\tilde{\mu}_t$ and $\tilde{h}_t$ from the state variables according to (11)–(13).
2. Impute the unexpected innovation $\Delta\widetilde{W}_t$ from the return Δe_t as $(\Delta e_t - \tilde{\mu}_t \Delta t)/\tilde{h}_t$.
3. Compute the addition to the log likelihood: $-\log \tilde{h}_t - \frac{1}{2}(\Delta \widetilde{W}_t)^2$.
4. Update all of the state variables simultaneously via (5)–(9).

Note that other than as auxiliary state variables in this evaluation, the agents' conditional moments for μ_t and S_t play no further role in our analysis. Indeed, it is important now to distinguish between their inference problem and ours, which is to estimate the model's constant parameters.

Besides the start-up values of the state variables, there are now only four free parameters: θ_2 plus the three structural variables. The parameter θ_1 is restricted to unity, which is merely a normalization because that parameter is not identified. Taking it to be positive corresponds to defining increases in Δ_t to be "good news." Because increases in the fundamental growth rate should also then be good, it is natural to impose $\theta_2 > 0$. [This is also suggested by the monetary interpretation of the exchange rate Equation (2), under which $-\theta_2$ is the semielasticity of money demand with respect to nominal interest rates, which should be negative.] The structural variables σ_0, λ_0, and λ_1 are also restricted to positive values.

2.3 The Role of Trends

Before proceeding to the data, it is worthwhile to explain some key properties of the system on an intuitive level. To start, consider (12), which describes exchange rate volatility. The process $\tilde{h}$ has three components. The first is just the constant, σ_0, inherited from fundamentals. The next two account for all of the heteroscedasticity. One component is proportional to the current level of uncertainty about true returns: $\widetilde{\text{var}}_t(\mu)$. This term generates GARCH effects: it goes up or down, depending on the recent history of fundamental variability. Big shocks have long-lasting effects on posterior uncertainty, leading to persistence in volatility. Any model with growth rate uncertainty will contain such a term.

The last component of return volatility is, however, unique to this system. It is proportional to $\widetilde{S}$, the likelihood that current shocks are of the permanent variety. This simply reflects the fact that when shocks are thought to have long-lasting effects they move prices more. This component would be present even if there were no parameter uncertainty in the model. When S is not observed, however, changes in this term respond to the information contained in the path of returns.

Consider now how this pathwise response of beliefs about persistence works. The issue is: when are we likely to think fundamentals are undergoing permanent change? The answer, which is the key to the volatility–momentum relation discussed in the Introduction, is that we raise $\widetilde{S}$ following successive surprises in one direction. Intuitively, if a persistent positive shock occurs, the next period's actual growth rate, μ, is higher, but we do not yet know it. So we are more likely to be surprised in the positive direction. Or, *ex post*, after several positive shocks in a row, it is increasingly likely that our growth rate estimator, $\tilde{\mu}$, was too low. So μ is likely to have been perturbed recently; that is, some recent shocks have been permanent. But permanent shocks tend to follow other permanent shocks (regimes change infrequently). Hence current innovations are probably also permanent.

This intuitive argument can be formalized by analysis of the stochastic differential equation describing the evolution of $\tilde{h}$ in terms of the innovation process. If this representation is written as

$$d\tilde{h}_t = \tilde{f}(X_t)\,dt + \tilde{g}(X_t)\,d\widetilde{W}_t, \tag{14}$$

then the quantity $\tilde{g}$ describes the volatility of volatility (which may have either sign). The product $\tilde{g} \cdot \tilde{h}$ is the covariance between return and volatility innovations, and since it is easy to show that $\tilde{h} > 0$, the corresponding instantaneous correlation coefficient is just the sign of $\tilde{g}$.

It turns out that $\tilde{g}_t$ can be approximated by a process of the form

$$\int_0^t \psi(t-u)\,d\widetilde{W}_u,$$

where $\psi()$ is a nonnegative, decreasing kernel determined by the switching parameters λ_0 and λ_1. Thus the volatility of volatility behaves like a local momentum indicator. In particular, under mild conditions, the sign of this indicator will equal the sign of $\tilde{g}_t$. This immediately implies the trend effect described in the Introduction. If the two signs are positive, say, then recent returns have been positive, that is, the local trend is up. That trend continues if (and only if) the next shock is positive. But in that case, since $\tilde{g}$ is positive, Equation (14) says that volatility goes up. Volatility increases when trends continue.

Moreover, the approximation argument implies that the strength of recent trends drives the magnitude of $\tilde{g}$ and hence the return–volatility covariance. As noted above, the strength of this covariance determines the expected skewness of finite-horizon returns. This covariance updating accounts for the relationship between momentum and levels of expected skewness.

This has been an overview of the principal features of the model. The remainder of the article employs a discretized version of Equations (5)–(13). Further details of the filter, including all of the discretized equations and the likelihood function, may be found in Appendix A.

3. FITTING THE MODEL

The theory of unobservable persistence provides a plausible account of the causes of stochastic volatility in returns. But it is a tightly parameterized and restrictive theory, demanding specific relationships between different moments through time. It is far from clear that these demands can be satisfied in practice. This section fits the model to each of two currency series. The sample consists of 15 years of daily returns for the U.S. dollar-Deutschemark (DEM) and dollar-yen (JPY). (A further description of the data is given in Appendix B.) At issue is whether the implied values of the structural parameters can describe the main time series properties of the exchange rates.

Overall the results are satisfactory. The model seems fully capable of capturing the primary features of the data. With these specification tests passed, the implications of the model for the volatility–momentum relationship are then taken up in Section 4.

3.1 Estimating the Parameters

The four free parameters of the model of Section 2 (called σ_0, θ_2, λ_0, and λ_1) each play multiple roles in determining the behavior of returns. There is no simple mapping between their values and, for example, the implied moments of a time series realization. (Indeed, it is not even clear that such moments exist. Neither stationarity nor ergodicity of the model defined by (5)–(13) has been established.) In this situation the likelihood function may be expected to be irregular, and asymptotic methods may be suspect. This article adopts a Bayesian approach to inference, which delivers the full posterior joint distribution of the parameters given the data.

Recently developed Markov chain Monte Carlo techniques permit sampling of high-dimensional posterior distributions even when, as here, the parameters enter the likelihood function through the interaction of multiple auxiliary processes. The posterior draws can then be used to perform exact finite-sample analysis of complex functionals of the parameters.

The sampling method implemented here is a regenerative Metropolis–Hastings algorithm of the type proposed by Tierney (1994, 1996). The algorithm, though computationally intensive, involves no special assumptions. [A cogent description is given in the article of Chib and Greenberg (1995).] The full procedure, including details of the methods used to monitor convergence, is described in Appendix B. The end result is, for each currency, a sample of 18,000 draws from the joint distribution of the parameters given the data.

Table 1 summarizes the information in the sample about the location of the unknown parameters. (Histograms of the marginals are shown in Fig. 1.) It is difficult to tell what the numbers in the table mean in terms of familiar time series properties. From the plots of the individual marginal distributions, it is apparent that all of the parameters are "significant" in the sense of having no probability of being zero. At the same time, the plots reveal extreme dispersion in some of the estimates. From this, and also from the large differences between the two currencies' mean parameters, the informativeness of the data appears weak.

Table 1. Fitted Parameters for the Unobservable Persistence Model

	DEM		JPY	
Parameter	*Mean*	*Mode*	*Mean*	*Mode*
σ_0	.0042	.0054	.0707	.0682
θ_2	305.7	105.8	.8182	.9984
λ_0	.2694	.2312	.5225	.4468
λ_1	8.826	8.125	2.270	2.647

NOTE: The table shows the parameter location estimates for the model of Section 2 derived from the joint posterior distributions, using 15 years of daily returns. Log exchange rates are modeled as $de_t = \tilde{E}_t(\mu)\,dt + \tilde{h}_t\,d\tilde{W}_t$ with $\tilde{h}_t \equiv \sigma_0(1+\theta_2(\tilde{S}_t + \widetilde{\text{var}}_t(\mu)/\sigma_0^2))$. $\tilde{E}_t(\mu)$, $\widetilde{\text{var}}_t(\mu)$, and $\tilde{S}_t$ are state variables describing beliefs about the unobservable fundamental growth rate μ and the persistence regime S. λ_0 and λ_1 are the switching intensities governing that regime, σ_0 is the exogenous component of volatility (in annual units), and θ_2 is the sensitivity of the exchange rate to changes in expected growth rates. The Metropolis–Hastings algorithm for generating the posterior distribution is described in Appendix B. Histograms of the marginals are shown in Figure 1.

To properly assess the precision of the estimation, however, it is important to account for dependencies in the posterior. The functions of the parameters that determine return dynamics may, after all, be quite accurately estimated. For example, by examining the evolution equation for $\tilde{h}$, one can deduce that its drift is zero when $\tilde{S}_t = \overline{S} \equiv \lambda_0/(\lambda_0+\lambda_1)$ and $\tilde{h}_t = \sigma_0(1+\theta_2\sqrt{\overline{S}})$. The latter quantity thus defines the steady-state volatility of the model. Figure 2 shows the marginal posterior distribution of this quantity for the two currencies. The figure indicates that this combination of the parameters is estimated quite precisely, with the posterior means being almost exactly at the unconditional sample standard deviations of the data series. Despite the seeming disparities of parameters for the two currencies, both imply return processes with about the same average volatility level, approximately 10%.

While the figure shows that the precision of the estimates in properly transformed units may be good, it leaves open the question of assessing the magnitudes of the parameters in terms of their dynamic implications. In particular, it is not clear if they are large in the sense of being significantly different from whatever region in the parameter space corresponds to homoscedasticity.

To assess the homoscedastic alternative, first consider the implied temporal duration of shocks to volatility. One way in which the parameters could be insignificant is for this quantity to be small. The drift of $\tilde{h}$ turns out to be damped quadratically toward its steady state with a half-life of

$$\log(\sqrt{2})/\sqrt{\overline{S}} \tag{15}$$

years. Figure 3 plots the marginal posterior distribution of this quantity for the two currencies.

For the yen, the data are highly informative, indicating a value reliably close to 9 months. Inferences about the DEM value are less precise, but apparently volatility shocks last at least twice as long. So the data support the usual conclusions from the GARCH literature of highly autocorrelated but probably not integrated volatility. More immediately relevant is the fact that neither posterior places any mass at the lower bound of $\log(\sqrt{2}) = 0.347$, indicating no attempt to dampen the models' volatility shocks.

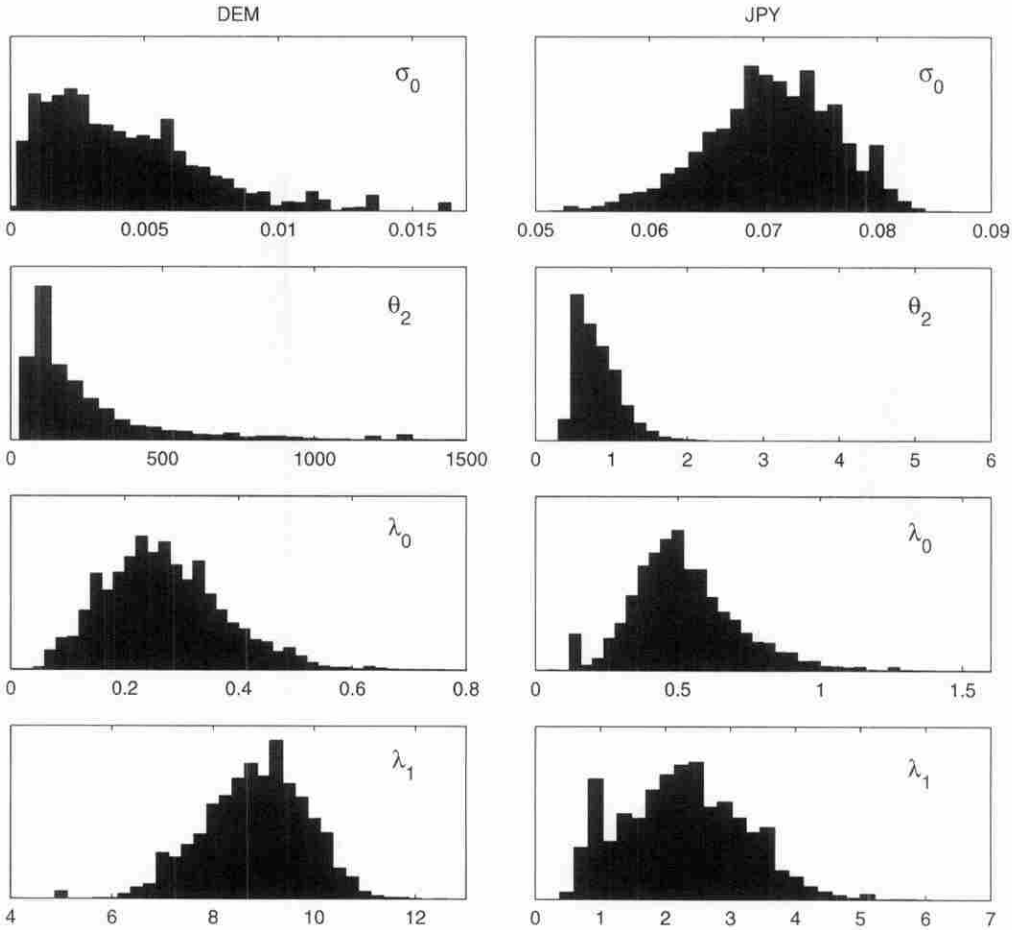

Figure 1. Posterior Distribution of Model Parameters. Log exchange rates are modeled as $de_t = \tilde{E}_t(\mu)\,dt + \tilde{h}_t\,d\tilde{W}_t$ *with* $\tilde{h}_t \equiv \sigma_0(1 + \theta_2(\tilde{S}_t + \widetilde{var}_t(\mu)/\sigma_0^2))$. $\tilde{E}_t(\mu)$, $\widetilde{var}_t(\mu)$, *and* $\tilde{S}_t$ *are state variables describing beliefs about the unobservable fundamental growth rate* μ *and the persistence regime* S. λ_0 *and* λ_1 *are the switching intensities governing that regime,* σ_0 *is the exogenous component of volatility (in annual units), and* θ_2 *is the sensitivity of the exchange rate to changes in expected growth rates. Using 15 years of daily currency data, the joint posterior distribution is sampled, using the two-stage acceptance/rejection regenerative Metropolis–Hastings algorithm of Tierney (1994, 1996). Chains of length 20,000 are drawn for each currency, of which the first 2,000 are discarded. The plots show the marginal posterior densities for* σ_0 *(first row),* θ_2 *(second row),* λ_0 *(third row), and* λ_1 *(fourth row).*

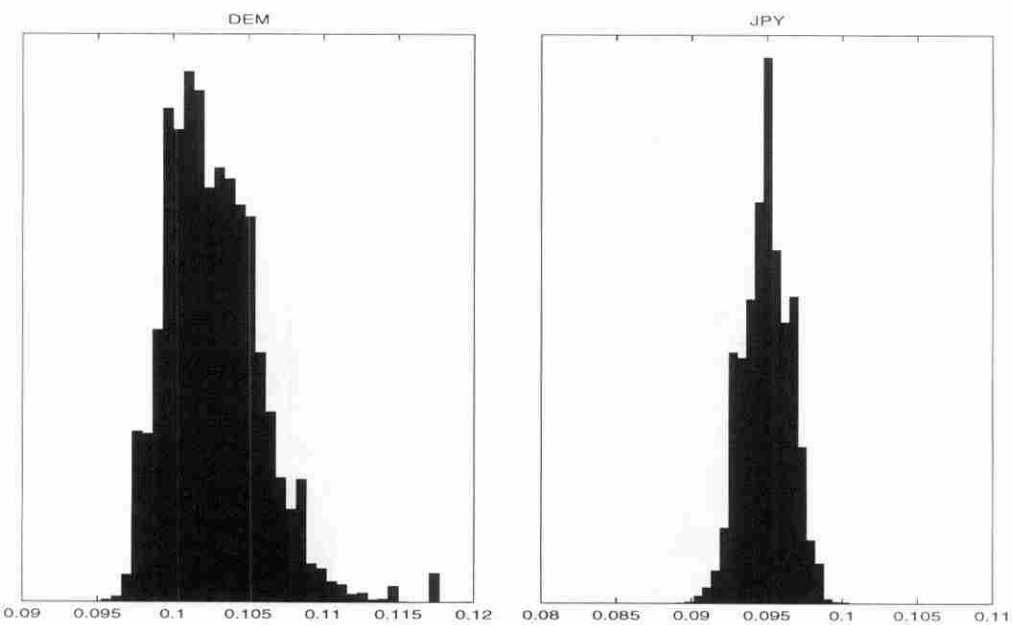

Figure 2. Posterior Distribution of Steady-State Volatility. Samples from the posterior distribution of the parameters of the model of Section 2 are drawn for two currencies, using the data series and methodology described in Appendix B. The histograms show the marginal posterior distributions of the long-run average volatility, $\sigma_0(1 + \theta_2\sqrt{S})$, *in annualized units, for each currency.*

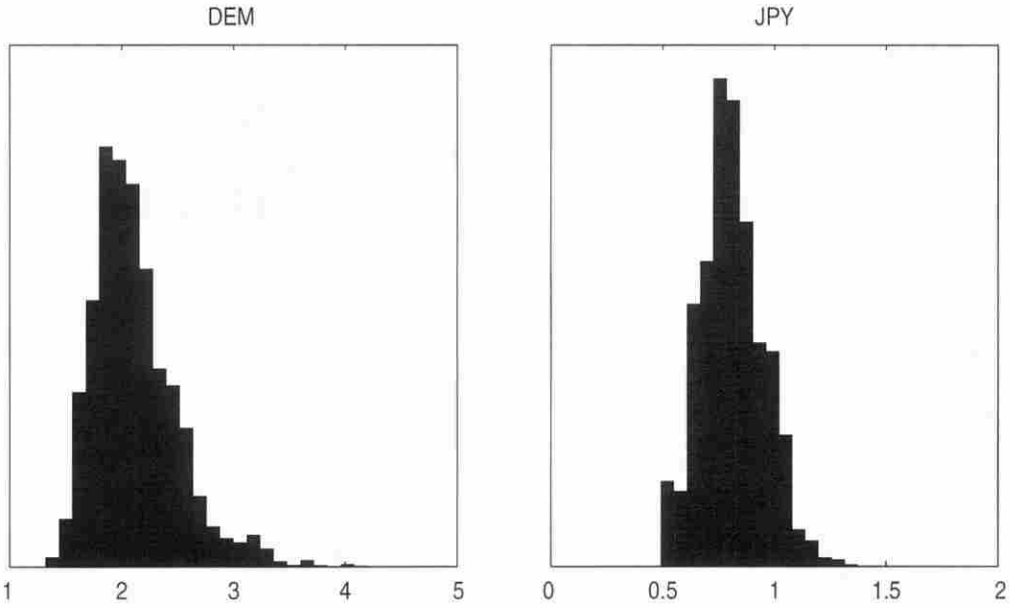

Figure 3. Posterior Distribution of the Volatility Persistence Parameter. Samples from the posterior distribution of the parameters of the model of Section 2 are drawn for two currencies, using the data series and methodology described in Appendix B. The histograms show the marginal posterior distributions of $\log(\sqrt{2})/\sqrt{\bar{S}}$ *in units of years.* $\bar{S}$ *is the unconditional mean of the persistence regime indicator S, and the plotted quantity corresponds to the half-life of a volatility shock away from the steady state.*

Next, if the data did not support the theory's volatility specification, they could also collapse the system to homoscedasticity by ensuring that the process $\tilde{g}_t$, the volatility of volatility, never became large in absolute value. It can be shown that the average level of this coefficient is given by the functional

$$\sigma_0\theta_2\frac{2.4\lambda_0\lambda_1^{3/2}}{(1+\lambda_0+\lambda_1)^{1/2}(\lambda_0+\lambda_1)^{5/2}}. \tag{16}$$

(Specifically, the drift of $\tilde{g}_t^2$ is zero at approximately this level of $|\tilde{g}_t|$.) Notice that the data could force this to be small, thus "turning off" the model, by making either of the regime parameters, λ_0 or λ_1, very small or very big. Alternatively, the same could be accomplished by making either one big or small relative to the other. Last, the data could simply put $\theta_2\sigma_0$ near zero.

Despite the complicated functional form of the quantity in (16), its finite sample distribution may again be immediately computed from the posterior draws. Figure 4 plots this for the two exchange rates. The horizontal axis is in units of annualized return volatility percentage. (So if yen volatility is 10% and $\tilde{g}_t$ is 1.0, this corresponds to swings from 9% to 11%.) Inferences about this quantity are remarkably precise, with

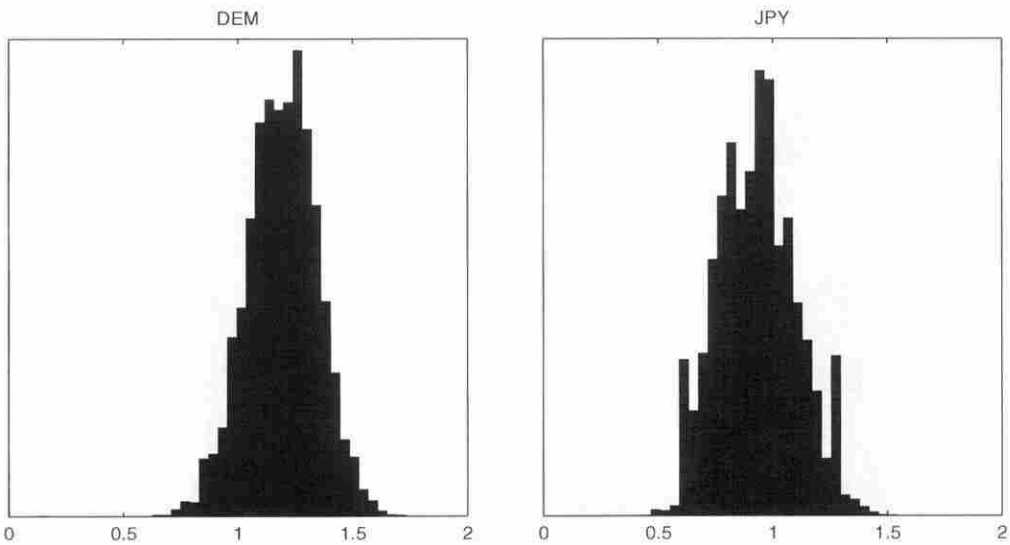

Figure 4. Posterior Distribution of the Heteroscedasticity Parameter. Samples from the posterior distribution of the parameters of the model of Section 2 are drawn for two currencies, using the data series and methodology described in Appendix B. The histograms show the marginal posterior distributions of the average volatility of volatility [Eq. (16)] in units of annualized volatility percentage.

posterior mass for both entirely contained in an interval of about a unit length. There is no probability that the statistic is near zero. The conclusion is, again, that the data clearly reject homoscedasticity in favor of the model.

That is about as far as one can go "testing down" with this model. Its unusual construction means that none of the familiar GARCH-type alternatives is nested. As the last three figures show, the data are quite informative about the implied dynamics of the models. Virtually all the posterior mass for the quantities examined lies within an interval of ±50% of their central location. These intervals define models with similar long-run volatility, volatility of volatility, and volatility persistence. The next step is to assess whether the fit that these estimated parameters achieve is acceptable.

3.2 Discrepancy Statistics

To check whether the fitted models provide a reasonable description of the data, this section implements several specification tests which again exploit the information contained in the draws from the joint posterior parameter distribution.

The idea of posterior predictive tests—introduced by Meng (1994) and Gelman, Meng, and Stern (1996)—is straightforward and intuitive. It addresses the question of whether the fitted models "look like" the data. For any given test statistic, the exact distribution is generated by (i) drawing m parameter values randomly from the posterior; (ii) for each draw, simulating a time series from the model it describes which is of the same length as the data; then (iii) tabulating the values of the test statistic for each of these series. Finally, the actual sample value is assessed against the resulting distribution.

Specifically, the method delivers Bayesian "p values" defined as the percentage of the m replicated samples producing a test statistic greater than the observed sample test statistic. As in classical inference, p values close to 0 or 1 constitute evidence against the specification. Such low tail probabilities are usually quite easy to detect [Gelfand (1996, p. 153) suggests that "in practice m need not be too large, say 100 to 200"], making the technique attractive computationally. The tests below are performed on 500 replicated samples. In addition to p-values, the full simulated distribution of each test statistic is presented graphically.

Also as in classical inference, the usefulness of the technique depends on appropriate design of the test statistics. Just as with the familiar generalized method of moments (GMM) tests of overidentifying restrictions, power is maximized by assessing discrepancy along dimensions about which the data speak most prominently (i.e., the principal moments). At the same time, these are typically the dimensions that are explicitly matched by the estimation technique, rendering them useless for testing goodness of fit.

Here, however, the model has not actually been permitted to explicitly fit any of the most obvious features. As already noted, the free parameters derive from a particular structural theory and do not directly encode an average mean, variance, or autocorrelation function for the implied return process. Thus it is not at all a foregone conclusion that the model's sample paths look like the data in these respects. So the fit can be assessed with the best possible test statistics: the first, second, and fourth moments of returns.

The results are shown in Figure 5. To see how to interpret these, consider the first panel (upper left), which shows the distribution of the mean return for the Deutschemark. The in-sample value was −.00012 (marked on the horizontal axis with the symbol ∧). Based on the pseudosamples, the probability of seeing a statistic at least this large (right tail) in a sample this length is 59.2%, and that of a value at least this small (left tail), 40.8%. So the true sample is squarely in the center: test passed. The yen sample mean, −.00026, is actually higher than that of the pseudosamples. Its one-sided p value is 7%, which is low but is not strong evidence of misspecification. The plots on the second row show the analogous tests for the mean absolute value. The realized values are comfortably in the middle of the histograms for both currencies. The fourth moment densities (bottom figures) contain some surprises. Typically for models with normal innovations, the yen kurtosis, 5.33, exceeds expectations: most of the distribution's mass is around 3. But, atypically, one cannot consequently reject the model: the p value is an unalarming 6.8%. The mark sample value 4.78 is almost exactly in the middle (p value 46.8%). And note the horizontal scale. The pseudosamples quite regularly exhibit enormous kurtosis. (Indeed, 6% of the mass is off the scale to the right.) Accounting for parameter uncertainty, the fat tails of the data are not unusual. This result is striking. Readers familiar with these sorts of data sets will readily appreciate how difficult it is for a volatility model with normal innovations to match the kurtosis of daily returns.

So far, the model seems well specified in terms of unconditional sample moments. The same techniques can be used to assess the main dynamic properties of moments as well. Here the natural quantities to use as test statistics are the autocorrelation coefficients of returns and absolute returns.

Figure 6 shows the first 20 days' lag of the raw returns' autocorrelation (with the Deutschemark on top and the yen below). The same function is then calculated for each of the 500 pseudosamples. At each lag length this then produces a distribution like those in the preceding figure. For ease of presentation, the 5th and 95th percentiles of these are plotted around the true sample function to suggest a confidence band. (It is not a joint confidence region, however, since the autocorrelation ordinates are mutually dependent.)

Here, too, the conclusion seems to be that the data look as expected, given the model and posterior parameter uncertainty. A few JPY values fall slightly outside the 90% range, but this is to be expected. The important point is that the pseudosamples, like the data, have trivial levels of autocorrelation in returns. Once more, this was not to be taken for granted. The model does contain a dynamic specification of the mean term, which is tightly linked to the other moments. The structure could well have generated an autocorrelation seriously awry in some respect.

The volatility autocorrelation is perhaps the most important functional of all to match. The unobserved persistence model was motivated as an attempt to account for patterns of conditional heteroscedasticity. So certainly it must be regarded as inadequate if it cannot describe this property of the data. Figure 7 is analogous to Figure 6. For each currency a 90%

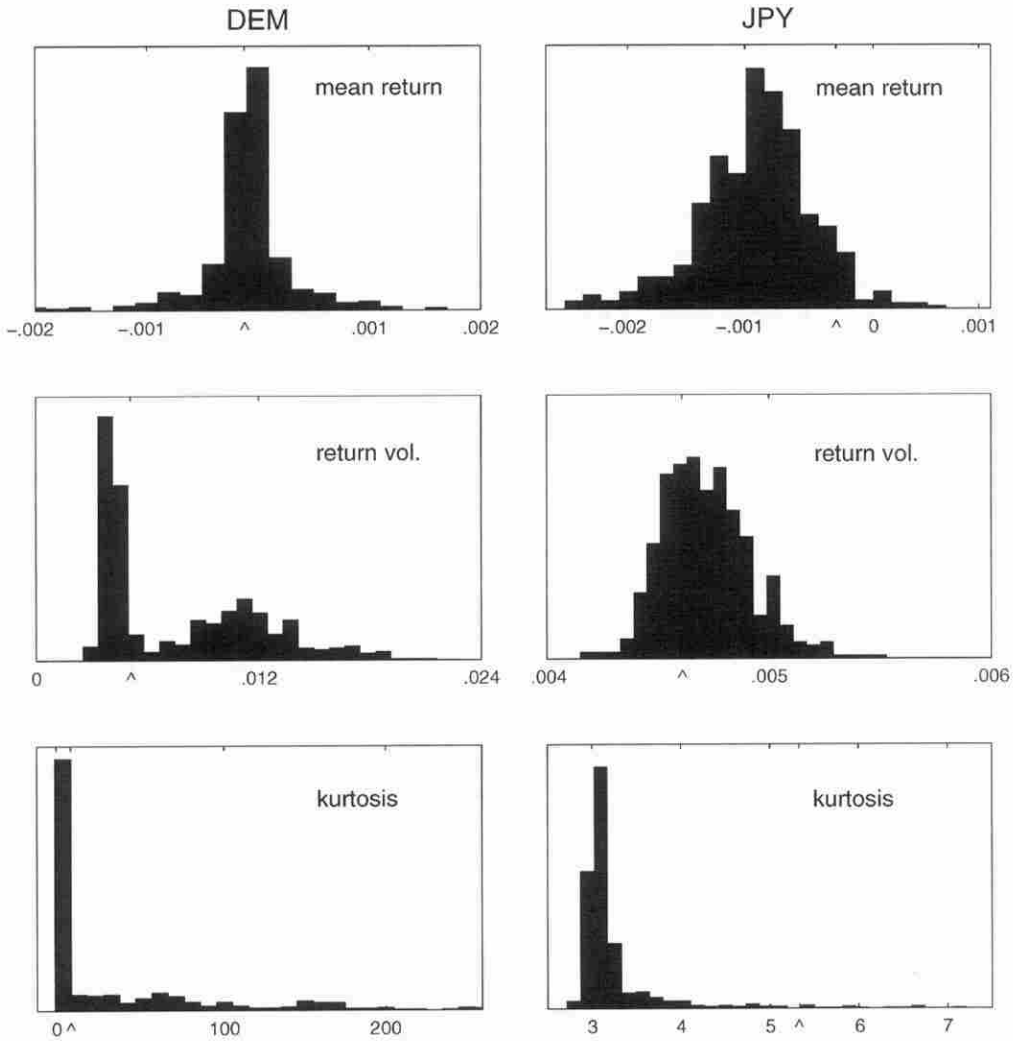

Figure 5. Posterior Distribution of Sample Statistics. The figure compares actual values of sample statistics with those generated from 500 pseudosamples of the same length. These are simulations of the model determined by a random draw from the posterior distribution of the parameters. The value from the actual data is marked on each horizontal axis with the symbol ∧. The top two panels are the daily mean returns. The middle panels are the mean absolute values of the daily returns. The bottom panels are the standardized fourth moment of the returns. The returns are unannualized.

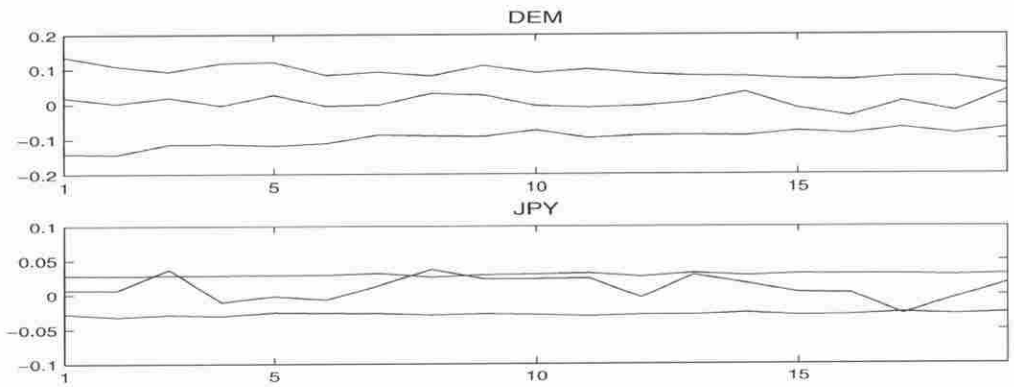

Figure 6. Posterior Distribution of Return Autocorrelations. The figure compares actual values of return autocorrelations with those generated from 500 pseudosamples of the same length. These are simulations of the model determined by a random draw from the posterior distribution of the parameters. The horizontal axis is the lag in days. The top and bottom lines in each panel represent 5th and 95th quantiles from the simulated distribution of each ordinate. The middle series is the actual return autocorrelation from the data.

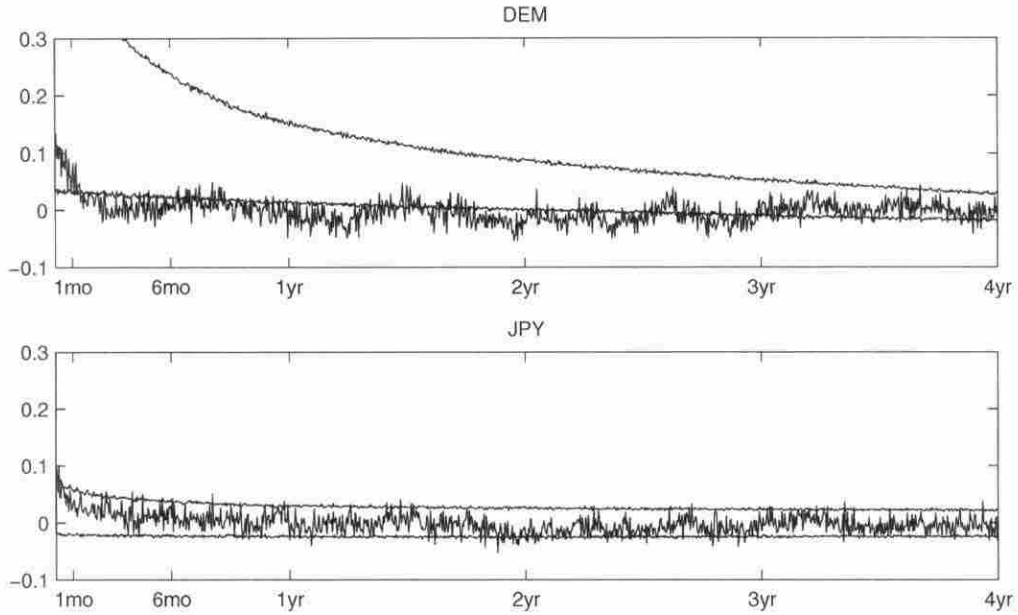

Figure 7. Posterior Distribution of Volatility Autocorrelations. The figure compares actual values of autocorrelations of absolute returns with those generated from 500 pseudosamples of the same length. These are simulations of the model determined by a random draw from the posterior distribution of the parameters. The top and bottom lines represent 5th and 95th quantiles from the distribution of each ordinate. The middle series is the value from the data.

probability band is shown for the autocorrelation function of the returns' absolute value extending out to a lag of 4 years. The DEM graph indicates that the posterior density conveys little information about the level of the autocorrelation function. The bands are extremely wide. Still, they show that the pseudosamples are on average much more GARCH-like than the currency. Although at long lags (over 3 years) and short lags (under a month) nothing appears amiss, the intermediate lags indicate that model shocks to volatility are too persistent. The discrepancy statistics here highlight a potential direction for model improvement. This is an indication that the specification might benefit from increased flexibility. The JPY series presents no such difficulties. Here, the data define quite precisely the expected band, and the true series falls squarely inside it nearly everywhere. For both currencies, it is clear that the model has no difficulty in accommodating GARCH effects and long memory in volatility.

The analysis here has produced evidence of an acceptable fit between the model and real currency returns. The claim thus far is not that the model dominates or outperforms any of the myriad competing volatility specifications, but only that the restrictions it imposes are not in conflict with the fundamental features of the data.

4. MOMENTUM, VOLATILITY, AND SKEWNESS

This section examines both of the findings discussed in the Introduction: the role of trends in predicting instantaneous volatility and the comovement of returns and expected skewness. The scale of the effects found in the data is shown to be consistent with that predicted by the model. That this is achieved without the benefit of a specification search (only the parameters fit in the last section are used) supports the assertion that the unobservable persistence model provides a structural explanation for the empirical facts.

This assertion is then further strengthened by a final test. The model predicts a relationship between levels of expected skewness and trends, in addition to that between their changes. This effect, previously undocumented, is shown to be present in historical options data. And, again, the size of the response is commensurate with that implied by the model.

4.1 Trends as Volatility Predictors

Müller et al. (1997) provide direct evidence of the influence of long-term returns (trends) on short-term volatility. The authors estimate a volatility specification, dubbed HARCH, containing standard autoregressive conditional heteroscedasticity (ARCH) terms as well as terms measuring the strength (in absolute value) of trailing trends of various lags. These multiperiod return terms are positive and statistically large across all currencies and all time periods studied. It is important to realize that the individual returns comprising the trend also appear as predictors (via the ARCH terms). Thus the effect is not merely a mechanical consequence of autocorrelated volatility. The occurrence of a trend contributes to volatility over and above that expected from any concurrent pathwise variability.

To assess whether the magnitude of the HARCH effect is in line with that implied by this article's model, the technique of posterior predictive testing is employed as in the preceding section. A discrepancy statistic capturing the influence of longer term returns on short-term volatility is used to evaluate the similarity of the model's expected realizations to the data. With the pseudosamples generated above, the finite-sample

distribution of this statistic is then tabulated and compared with the observed value in the data.

The results of Müller et al. (1997) suggest using the coefficient on long-term absolute returns from a fitted HARCH model as such a statistic. To avoid the computational burden this would involve, the same sensitivity can be measured by a simpler quantity. One choice is simply β_2 in the following regression:

$$\log \hat{h}_t = \beta_0 + \beta_1 \log \hat{h}_{t-1} + \beta_2 \left(\log \frac{|R5|_{t-1}}{\sqrt{5}} - \log \hat{h}_{t-1} \right) + \epsilon_t.$$

Here $\hat{h}$ is the 5-day standard deviation, and $|R5|$ is the 5-day absolute return. No particular statistical properties are required of this regression or of the ordinary least squares (OLS) estimates of the coefficient. β_2 is simply being used as a descriptor of the association between volatility and lagged trends, suitably orthogonalized. Computing this statistic with the two currency samples used in the last section produces positive values for both (.064 for the mark and .029 for the yen), confirming the presence of the trend effect in returns.

When the same regression is performed with each of the 500 pseudosamples, the result is the pair of distributions depicted in Figure 8. (The values from the data are indicated on the horizontal axis.)

Both models do, in fact, predict a positive effect. For the DEM, the model achieves striking fidelity. The true sample value is almost exactly the distribution's median, and there is almost no mass on the negative half-axis. For the yen models, the effect is not as extreme as in the real returns. Only 7% of the replications attain the level seen in the data. The specification, while not as responsive to trends as desired, would still not be regarded as critically improbable at conventional levels. Moreover, the distribution still places 64% of its mass on positive values and has a positive mean (.007). It is also encouraging that the model does a better job matching this feature of the data for the currency series in which it is more pronounced. Clearly its structural restrictions do not preclude strong trend effects.

The findings of Müller et al. deserve attention. As demonstrated in their article, it is impossible to produce the volatility–trend relation with a standard GARCH-type model. Their HARCH specification [like the earlier QARCH specification of Sentana (1995), of which it is a restricted version] thus constitutes a valuable addition to the family of discrete-time volatility models. The model introduced here complements it in providing a parsimonious, structural explanation for its characteristic feature. In addition, the complete-markets, continuous-time framework enables the derivative pricing implications of this new effect to be readily analyzed. This is the next topic.

4.2 Risk Reversal Dynamics

Currency options traded in the interbank market are quoted in units of Black–Scholes implied volatility, and their strike price (relative to the current spot value) is commonly expressed by the implied Black–Scholes delta, or hedge ratio. When the former is plotted against the latter for options of the same expiration, the resulting curve is referred to generically as the "smile"—its typical shape being convex with a minimum near the 50-delta strike, its midpoint. The asymmetry of the smile can be measured by the difference in heights of this curve at points equidistant from the middle to the left and right. Traders call this difference the "risk reversal."

Risk reversals are directly quoted and traded by market participants, and bid–ask spreads are narrow. Hence changes in this degree of asymmetry can be measured extremely accurately. Recently, Campa et al. (1998) (hereafter CCR) documented a strong correlation between changes in spot exchange rates and changes in risk reversals. Using interbank data on five

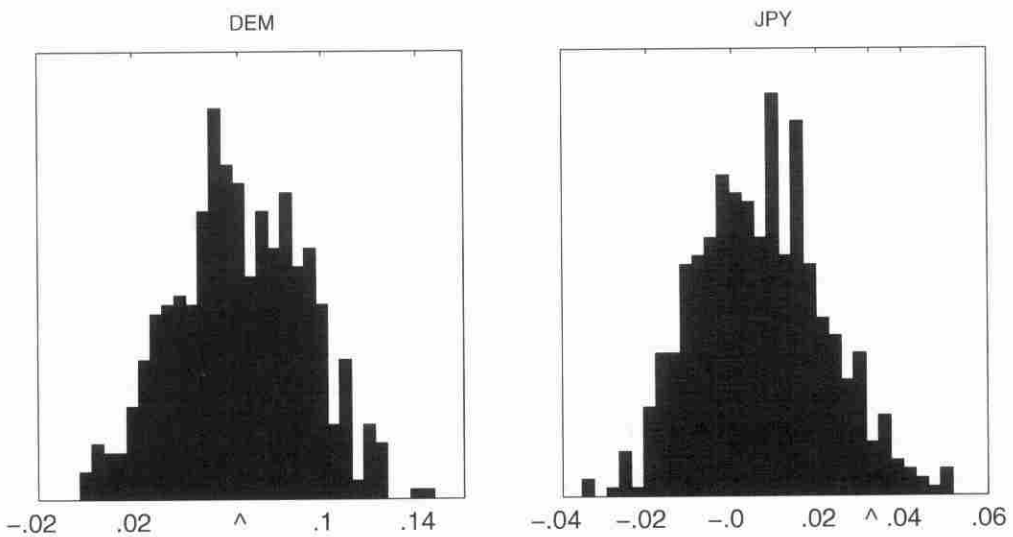

Figure 8. Strength of the Trend Effect. For each of the pseudosamples generated in Section 3, a 5-day volatility series is regressed on its own lag and the lag of the 5-day trend (orthogonalized) to compare the strength of the last term in the data with its distribution under the model. So, for each simulated series, the specification is $\log \hat{h}_t = \beta_0 + \beta_1 \log \hat{h}_{t-1} + \beta_2 (\log(|R5|_{t-1}/\sqrt{5}) - \log \hat{h}_{t-1})$, *where* $\hat{h}$ *is the 5-day standard deviation and* $|R5|$ *is the 5-day absolute return. The figure shows the distribution of* β_2. *The values computed from the real data are marked by the symbol* $\wedge$ *on the horizontal axis.*

currency pairs, they report that, for every series examined, the difference between the implied volatility of a 25-delta call and a 25-delta put tended to increase as the price of the underlying currency went up. (That is, calls on the yen became relatively more expensive as the yen strengthened.) The effect was not sensitive to the choice of delta or expiration of the options.

In the language of finance, this phenomenon can be equivalently described in terms of the state price density or risk-neutral distribution. The options data impounded in the smile may also be used to extract a density, with respect to which discounted expected pay-offs correspond to prices. The asymmetry of the smile may then be viewed as arising from the relative thickness of the left and right tails of this density. For ordinary, unimodal distributions, in fact, there is a monotonic relationship between the risk reversal and the central third moment, or skewness, of the distribution. The CCR result is thus that the skewness of the risk-neutral distribution covaries positively with weekly returns.

As noted in the Introduction, this would be an expected property of the true, statistical distribution in the presence of the trend effect. From the last section, the model does capture this effect to the degree implied by the movements of exchange rates themselves. What is far from obvious is whether parameters estimated only from these data also imply a responsiveness of forecast distributions to momentum that is anything like that observed in options markets.

To find out, theoretical options prices derived from the fitted models are generated, and the behavior of their implied risk reversals is compared to that found by CCR. For this calculation, it is necessary to abstract from the parameter uncertainty of the fitted models due to the computational requirements. Posterior modes are thus used as point estimates for the model unknowns. Since noninformative priors were used to generate the posterior, these are nearly identical to the maximum likelihood estimates. (The findings of this section are not altered by using posterior means as point estimates. The informativeness of the prior is discussed further in Appendix B.3.)

Now, under the null hypothesis that the actual currency samples were generated by the models parameterized by these values, histories of all of the latent state variables may be computed. From these, theoretical values of options to any horizon can then be evaluated at each point in time. [This is done by computing discounted expectations via Monte Carlo integration: the model's system of equations is evolved forward by drawing 20,000 random paths of $\Delta\widetilde{W}_t$. The dynamics must be transformed to the risk-neutral measure. But there are no market prices of risk involved because the same innovation series drives all of the state variables. Technical details about the equivalent martingale measure and the interest rate assumptions may be found in Johnson (1999).]

These theoretical options prices are generated at weekly intervals through the course of the 15-year sample for each currency. Evaluating a range of strikes, converting to Black–Scholes implied volatility units, and interpolating with the use of cubic splines gives the risk reversal at any desired moneyness. The results below use the 1-month 25-delta value.

Figure 9 plots changes to the theoretical risk reversal against weekly returns, the relationship studied by CCR. Both fitted models exhibit strong positive associations. Replicating the regression of the above authors yields robust t statistics of 31 and 39 for the coefficient of the spot difference for the Deutschemark and yen, respectively. Their values were 8.3 and 10.9. The model's coefficients are somewhat lower than CCR's, but its adjusted standard errors are much lower. Compellingly, the regressions explain 58% of the variance for the

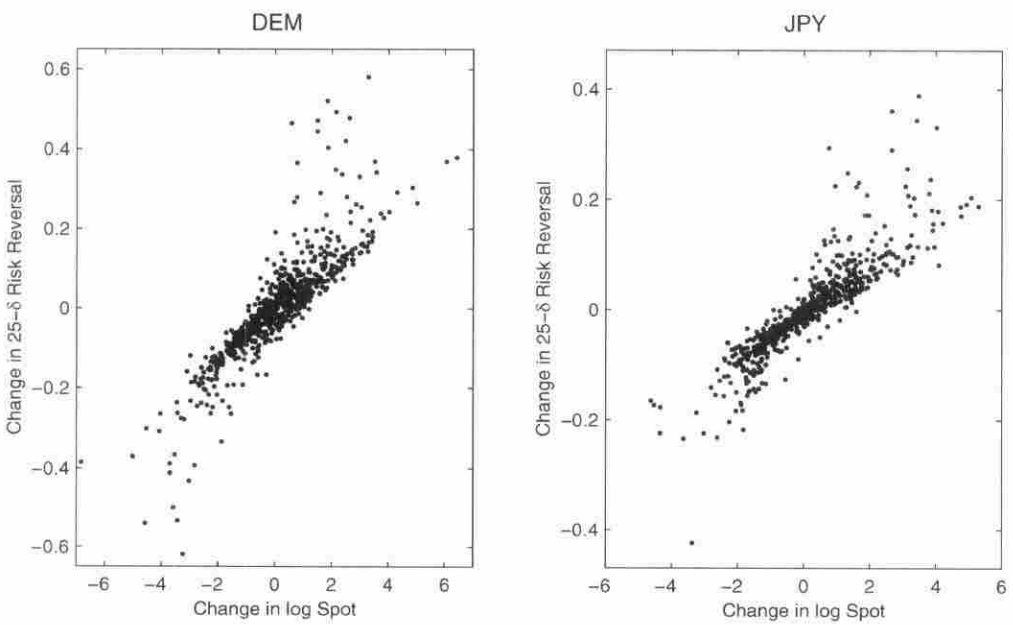

Figure 9. Risk-Reversal Relationship. Theoretical 1-month options prices are calculated via Monte Carlo integration for every week from 1/1/80 to 3/31/94 under the models fitted in Section 2. The graphs relate the difference in Black–Scholes implied annual volatilities between a 25-delta call and a 25-delta put (the "risk reversal") to the weekly change in the log spot rate. Both are expressed in percentage terms.

DEM and 69% for the JPY. These very nearly match the R^2's CCR report (59% and 71%) for the period 4/3/96 to 3/5/97. Measured on this basis, the strength of the relationship is the same under the model and in the data.

The conclusion of this exercise is that neither market irrationality nor rapidly time-varying risk premia need be invoked to explain the observed behavior of options smiles. This is not to say that either (or both) might not be at work. What has been established, however, is that the findings in CCR are perfectly consistent with the hypothesis that the shape of the true distribution of returns changes through time and that markets incorporate those changes rationally in their forecasts. The models fit in Section 3 describe not only the return process itself, but also the evolution of those forecasts.

4.3 Levels of Skewness

Campa et al. (1998) document the positive correlation between changes in expected skewness and changes in spot exchange rates. The model of Section 2, in fact, makes a more precise and somewhat stronger prediction. Recall that the volatility of volatility, $\tilde{g}$, is well approximated by the process

$$\Phi_t = \text{const} \cdot \int_0^t \psi(t-u)\, d\widetilde{W}_u, \tag{17}$$

where $\psi(x)$ is effectively an exponential smoother. The decay rate of this kernel turns out to be $(1+\lambda_0+\lambda_1)$, which, for the two currency models used here, is 9.35 and 4.09 in inverse years (for the mark and yen, respectively), corresponding to half-lives of about 4 and 9 weeks. Now $\tilde{g}$ also determines the volatility–spot covariance and hence the degree of asymmetry of finite-horizon returns. Thus the model implies a relationship between a specific moving average of returns and levels of expected skewness. This relationship implies the one reported by CCR (because weekly differences of a longer-term moving average of returns are essentially the same as weekly returns) but is not implied by it.

Figure 10 illustrates the steps of this argument, using the theoretical options price histories computed in the last section.

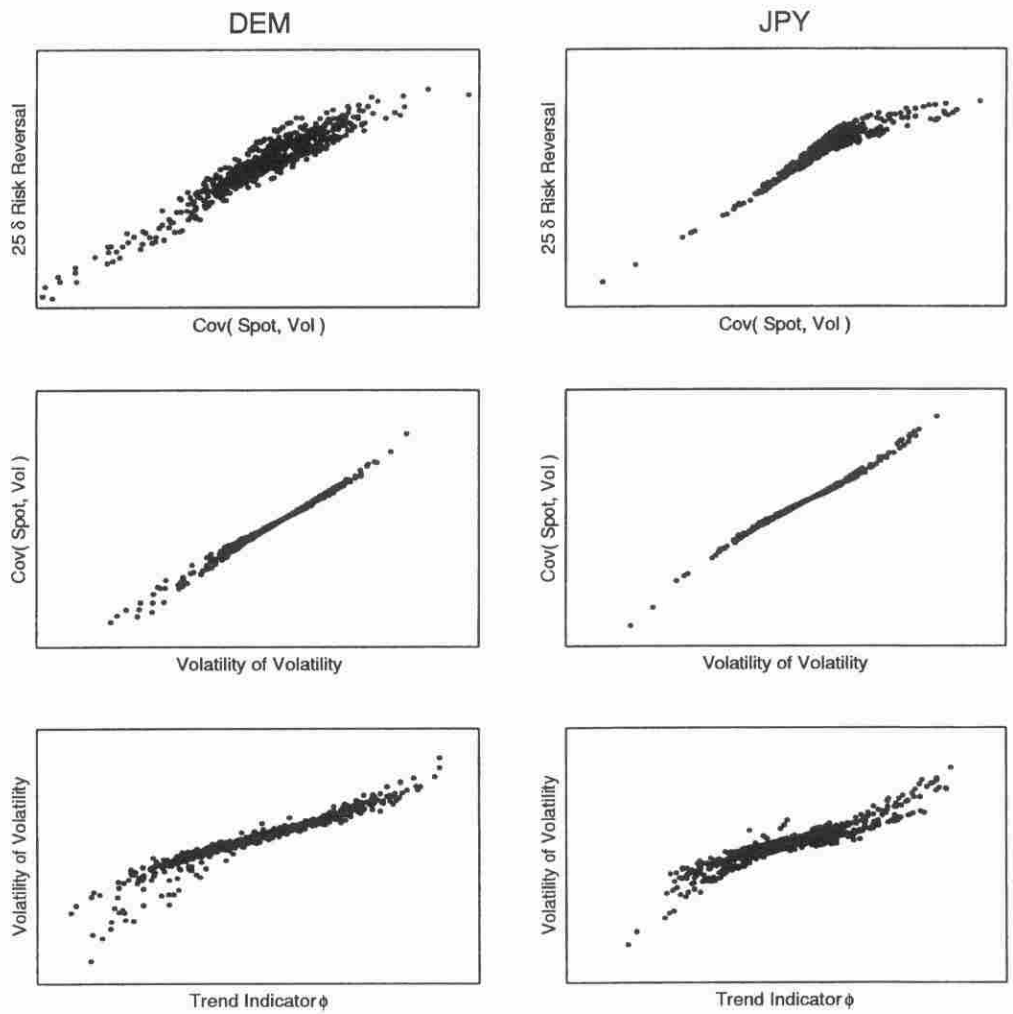

Figure 10. Anatomy of the Trend Effect. Theoretical 1-month options prices are calculated via Monte Carlo integration for every week from 1/1/80 to 3/31/94 under the maximum likelihood models fitted in Section 2. The top graphs relate the difference in Black–Scholes implied volatilities between a 25-delta call and a 25-delta put (the "risk reversal") to the instantaneous covariance under the model between spot and volatility innovations. The second row plots the same covariance against the volatility of volatility. The bottom graphs show the volatility of volatility against a trailing weighted average of spot returns. The weights are given by an approximate exponential smoother of half-lives of 19 and 44 trading days for the mark and yen, respectively.

The top pair of plots demonstrates the mechanical dependence of the 1-month horizon expected skewness (as measured by the risk reversal) and the instantaneous covariance between spot and volatility changes. This is a property of any stochastic volatility model. The middle row shows that the latter covariance is effectively determined by the (signed) volatility of volatility because the level of volatility is relatively stationary. Finally, the third row graphs the relationship between $\tilde{g}$ and the momentum indicator described above. Units of the graphs are omitted for readability. The point is simply to indicate the expected positive association between the y variable on the top row and the x variable on the bottom.

This prediction about levels of skewness may be investigated by means of the distributional histories implicit in foreign exchange futures options listed on the Chicago Mercantile Exchange (CME).

By matching traded options prices with simultaneous transactions in the underlying futures contract, the implied risk-neutral distribution can be estimated either parametrically or nonparametrically. Bates (1996) takes the former approach, fitting a four-parameter family of distributions derived from a continuous-time jump diffusion model and the associated options pricing formula. The parametric framework permits the extraction of the key features of the distribution (the moments) without the data requirements of nonparametric methods. Yet the class is still flexible enough to fit Deutschemark and yen options prices as well, on average, as cubic splines (Bates 1996, p. 80). Although the resulting densities are to some degree model dependent, the moment estimates they imply are likely to be far more robust than those derived from the (poorly measured) tail behavior of nonparametric densities.

For each day since the inception of CME trading through December 31, 1992, Bates collects all transactions in options with the same expiration as the front-month futures contract (defined as the nearest settlement greater than 28 days ahead). He then estimates the four-parameter density via nonlinear least squares for each day for which there are at least 40 matched transactions over at least eight strike prices. This results in estimates for the first four moments of the state price density on 2,024 different days for the mark and 1,191 days for the yen. These provide a detailed, high-frequency record of traders' expectations over a significant subperiod of the sample employed in the preceding section.

The hypothesis at hand concerns the third moment, or (standardized) skewness of the exchange rate distributions:

$$\mathrm{E}(e_T - \mathrm{E}e_T)^3/[\mathrm{E}(e_T - \mathrm{E}e_T)^2]^{3/2}.$$

Bates' histories clearly show time variation in these. The question is, does this variation coincide with that predicted by the model? Or, omitting the model as intermediary, does momentum explain it? Note that measurement error in the extracted moments would bias against finding support for either proposition.

To conduct the tests, the model's forecast distribution is generated for each currency twice per week during the time period of the CME sample. (Note that the horizon of this forecast is not constant since it corresponds to the fixed calendar dates of CME expirations. It thus declines from about 16 weeks to 4 weeks and then jumps up again. Hence the moments are not strictly comparable from one day to the next.) As in the last section, the distribution is generated via simulation of the state variable processes, after transformation to the risk-neutral measure. Again, the unknown parameters are fixed at the same point estimates throughout. The skewness of each forecast distribution is then evaluated by Monte Carlo integration.

Figures 11 and 12 plot the estimated moments against each other for the two currencies. Although the series are remarkably similar, they are not identical. There is variability in the options' skew beyond that predicted by the model. In the case of the yen, that variability primarily consists of a handful of outliers that are immediately reversed. These suggest one-time, exogenous information events superimposed on a systematic component, which is quite well approximated by the model.

For the longer Deutschemark series, the performance is worse. Besides the nonpersistent outliers, there is now a persistent positive bias to the model after the middle of 1989. The numeraire here is the dollar [conforming with Bates (1996)]. So the model predicts excessive skew to the dollar downside. The options series also displays systematically high variability of skewness in 1984, which the model does not reflect. This could be a period of parameter instability or, as Bates concludes, evidence of market mispricing. As with the yen, the model fits especially well from the second quarter of 1985 through the middle of 1989.

The relationship between the two skewness measures is summarized in Table 2. The observed and theoretical distributions display the same overall variability in asymmetry, as can be seen from the mean absolute levels of their skewness. The regression of the implicit skews on the predicted values leaves no doubt that the model identifies a highly significant component of market expectations. The findings in the table—the significant slope and insignificant intercept—are robust to the use of raw (as opposed to standardized) third moments and to the identification of outliers. (Without trimming, the t statistic on the model skew is 5.45 for the yen and 3.10 for the mark.)

Formally, the hypothesis that the model explains all of the derived skewness throughout both samples is rejected. Even excluding the outliers, the R^2's are not close to 1, the errors are persistent, and the coefficients on the model skews are less than unity (although this could be caused by measurement errors in the extracted moments). Visually, the model's prediction is that all of the points in the scatterplots (the bottom panels of Figs. 11 and 12) should lie close to the 45° line. They do not.

On the other hand, an extraordinary number of them do. Put differently, the regressions strongly reject the null of no relationship. Indeed, noting again that (a) observed skewness is measured with error, and (b) parameter uncertainty has not been taken into account in evaluating the model's skewness, it remains plausible that the two, in fact, coincide during the entire sample period.

To put the explanatory power here in perspective, it must be appreciated that the information in the model forecasts derives entirely from the time series of returns themselves. This information is related to options prices via a structural

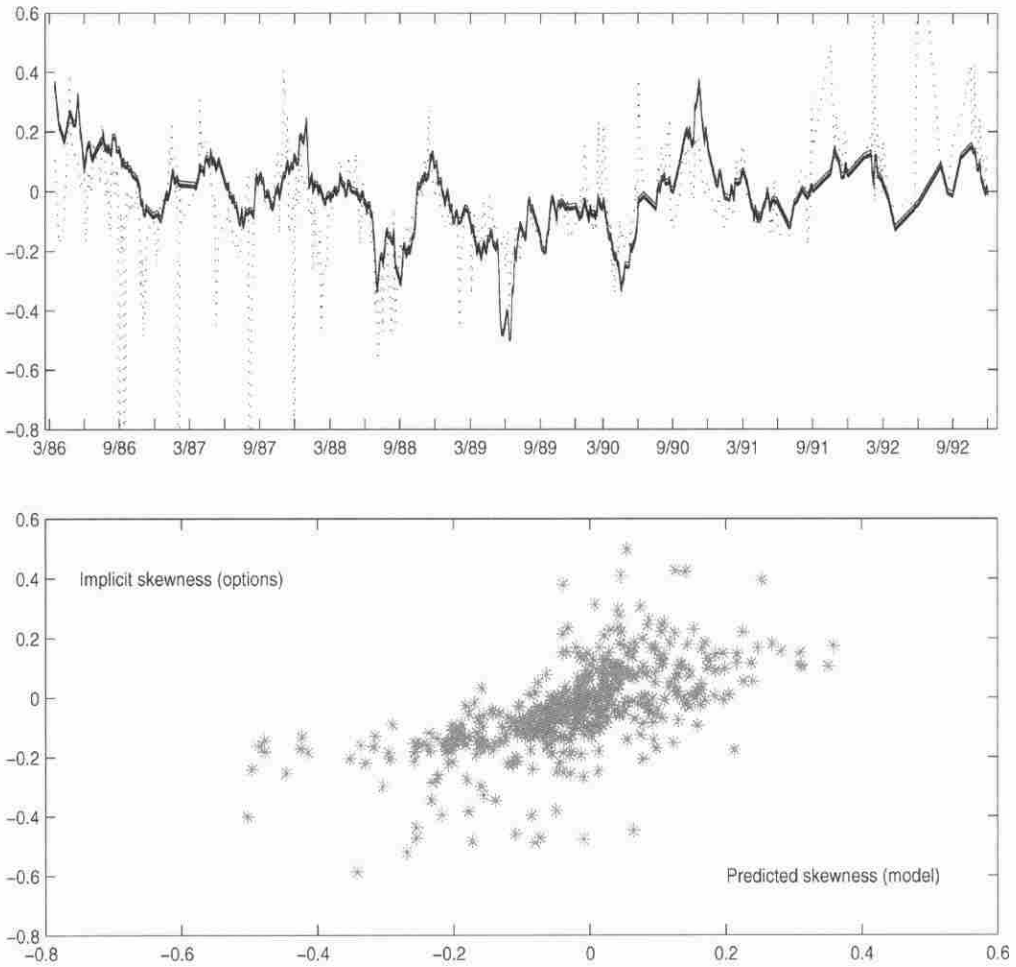

Figure 11. Implicit vs. Predicted Skewness: JPY. The top panel plots the standardized third moment of the exchange rate distribution at a horizon of between 30 and 90 days implied by listed CME futures options [as calculated by Bates (1996)] and the corresponding predicted value of the unobserved persistence model (using maximum likelihood parameters). The solid line is the model and the dashed line is the implicit series. The bottom panel is the same data in a scatter plot.

model, which must include a specification of the required preference adjustments. This is a challenging task for any asset pricing model. To the author's knowledge, this study may be the first to test the higher-moment implications of a stochastic volatility model against the information in options smiles in this way. Apparently the return dynamics, as captured by the model, reveal considerable information about the changing shape of the state price density.

Whatever the verdict on the model, the table does confirm the existence of the predicted relationship between momentum and expected skewness. The bottom section of the table shows that the regression results are basically unchanged when lagged trends [as given by the kernel weighting of Equation (17)] are used in place of the theoretical asymmetry as a predictor. This is not surprising, in light of the bottom graphs of Figure 10: the two candidate right-side variables have correlations 93% and 89% for the DEM and JPY, respectively. However, it is notable in that it both confirms and greatly strengthens the findings of CCR. Not only does it imply their result (by differencing), but it does so in a longer, nonoverlapping sample. There truly is a consistent relationship between trends and time-varying third moments, which deserves to be included in the list of stylized characteristics of currency returns.

5. CONCLUSION

This article has fit and tested a rather cumbersome and restrictive specification of stochastic volatility and returns. The model has the beneficial features of (i) deriving from a plausible theory of investor inferences about changing fundamental processes, and (ii) expressing all of its state variables as path functionals of returns and hence preserving market completeness.

The model was introduced to explore the connection between realized trends and changes in volatility. Foreign exchange returns exhibit the surprising and consistent property that volatility increases when trends continue and decreases when they reverse. Equivalently, the volatility–spot covariance, and hence finite-horizon skewness, behaves like a lagged momentum indicator.

Despite its restrictions, the model describes the data-generating process—including this dynamic—reasonably accurately. Moreover, fitted models lead to predictive distributions whose asymmetries accord well with those observed in cross-sectional histories of options prices.

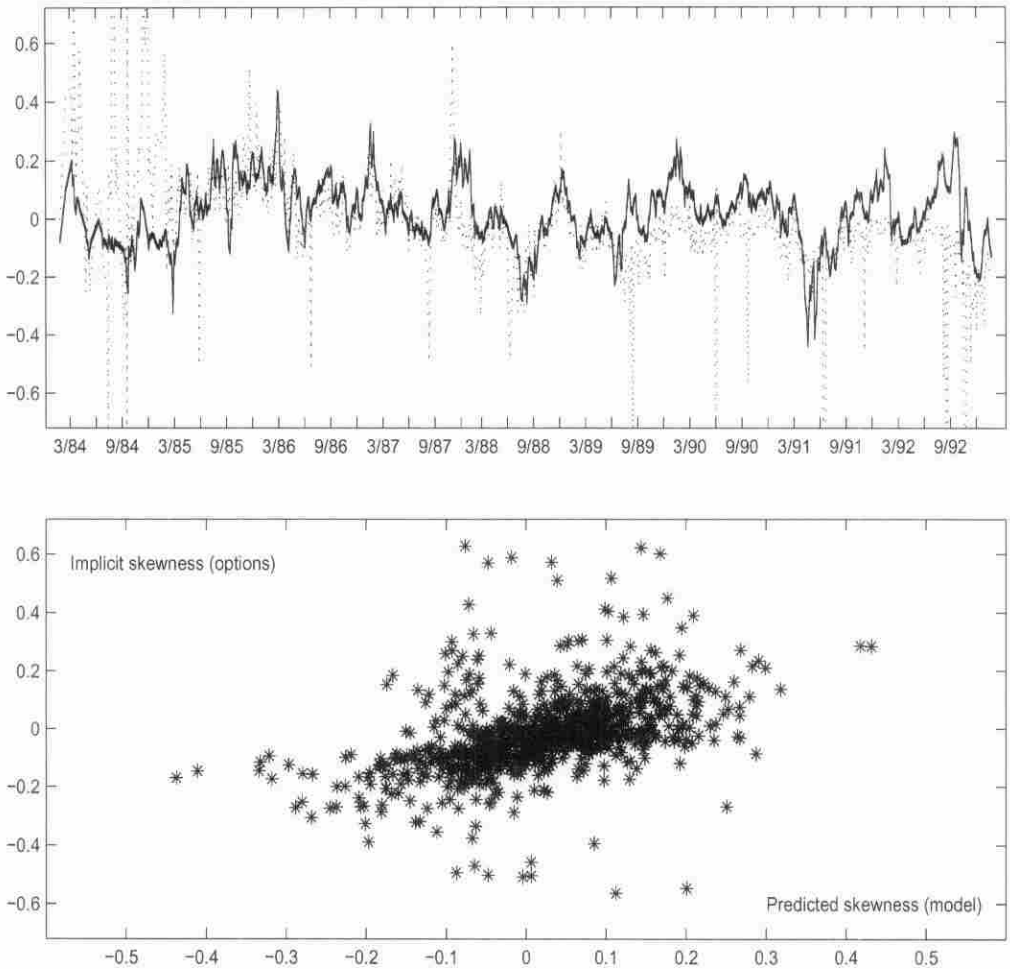

Figure 12. Implicit vs. Predicted Skewness: DEM. The top panel plots the standardized third moment of the exchange rate distribution at a horizon between 30 and 90 days implied by listed CME futures options [as calculated by Bates (1996)] and the corresponding predicted value of the unobserved persistence model (using maximum likelihood parameters). The solid line is the model and the dashed line is the implicit series. The bottom panel is the same data in a scatter plot.

The higher-moment regularities reported here should be of interest to both time series econometricians and financial engineers.

From a forecasting perspective, better understanding of the precise conditioning information that determines volatility can lead to significant model improvement. A suggestive illustration can be found in Dacorogna et al. (1998). There, a six-term HARCH model, which also captures the trend effect, is tested against a GARCH alternative for four currencies in a large high-frequency data set. The GARCH model is dominated across the board (both in and out of sample), offering direct evidence that ignoring this relationship entails significant costs in accuracy.

For financial practitioners, the ability to accurately model the evolution of implied volatility surfaces (or risk-neutral distributions) is also extremely important. The economic consequences of the trend effect can be seen most sharply in the values of knock-in and knock-out options. Intuitively, these derivatives condition their payoffs on the occurrence of a trend (from today's spot to the knock-in price). The last section showed that such trends produce systematic variation in expected skewness, which translates into a systematic asymmetry in option prices (the risk reversal). This implies a systematic relationship between the options' values and the location of the trigger price relative to the spot and strike prices. A standard stochastic volatility specification, where the correlation between returns and volatility is fixed, implies no such relationship. Even when such a model is calibrated to agree with the unobservable persistence model for all ordinary European options, the competing specifications can still produce exotic option prices that are significantly at odds. [Some detailed calculations appear in Johnson (1999).] In general, prices and hedge ratios of all contingent claims computed with misspecified volatility processes will be to some degree distorted.

This discussion underlines the conclusion that the trend effect is an important phenomenon in its own right. But it is also significant for the challenge it poses for asset pricing theories. This article has suggested one information-based explanation. Alternative behavioral or preference-driven explanations may perhaps be workable as well. Bringing such structural models of higher moment predictability to the data, as done here, promises to be an active area of future research.

Table 2. Comparison of Skewness Measures

	DEM 01/24/84–12/31/92	JPY 03/14/86–12/31/92
Implied skew		
Mean absolute value	.1065	.1180
Q_3–Q_1	.144	.183
Outlier cutoff	±.72	±.92
No. trimmed	14	7
No. observed	811	477
Model skew		
Mean absolute value	.0891	.1025
Slope	.5773	.6944
Std. err.	(.1047)	(.0991)
Intercept	−.0277	−.0177
Std. err.	(.0228)	(.0149)
R^2 (%)	20.0	39.2
Momentum		
Slope	1.900	7.909
Std. err.	(.5776)	(1.267)
Intercept	−.0326	−.0671
Std. err.	(.0258)	(.0223)
R^2 (%)	19.0	31.3

NOTE: Implicit skewnesses (standardized third central moment) of exchange rate distributions extracted from front-month CME options prices are taken from Bates (1996). The table reports the results of a trimmed regression of these on skewnesses derived from the time series model of Section 2, using the parameter values fit in Section 3. Observations were made twice weekly (Tuesday and Friday), except when there were insufficient options trades to compute the implicit measure. The trimmed observations correspond to values of the dependent variable exceeding five times the interquartile range (Q_3–Q_1) in absolute value. Standard errors (in parentheses) have been adjusted for autocorrelation and heteroscedasticity. The bottom section repeats the regression, using the momentum indicator of Equation (17) as the independent variable.

ACKNOWLEDGEMENTS

My thanks to George Constantinides, John Cochrane, Nick Polson, Jeff Russell, Angel Serrat, and Pietro Veronesi for encouragement and advice. I am grateful to David Bates for graciously sharing the data used in Section 4. This work has also benefited greatly from the thoughtful comments of the Associate Editor and two anonymous referees.

APPENDIX A: THE APPROXIMATING MODEL

As described in Section 2, the evolution of agents' beliefs in the unobserved persistence model is given by an infinite-dimensional system of stochastic differential equations. For empirical work, this system is approximated by a five-dimensional discretized subsystem. This appendix gives the details.

A.1. A Finite System

The exact filter for the system (3) in Section 2 is given by Proposition 2.2 of Johnson (in press), which characterizes the evolution of the conditional moments $\widetilde{\mu_t^k} \equiv E[\mu_t^k \mid \mathcal{F}_t^{\Delta}]$ and $\widetilde{\mu^k S_t} \equiv E[\mu_t^k S_t \mid \mathcal{F}_t^{\Delta}]$:

$$d\widetilde{\mu_t^k} = \left(k(k-1)\widetilde{\mu^{k-2}S_t}\frac{\sigma_0^2}{2}\right)dt + \left(\widetilde{\mu_t^{k+1}} - \widetilde{\mu_t^k}\tilde{\mu}_t + k\widetilde{\mu^{k-1}S_t}\sigma_0^2\right)d\widetilde{W}_t/\sigma_0 \quad \text{(A.1)}$$

$$d\widetilde{\mu^k S_t} = \left(\widetilde{\mu_t^k}\lambda_0 - \widetilde{\mu^k S_t}(\lambda_0+\lambda_1) + k(k-1)\widetilde{\mu^{k-2}S_t}\frac{\sigma_0^2}{2}\right)dt + \left(\widetilde{\mu^{k+1}S_t} - \widetilde{\mu^k S_t}\tilde{\mu}_t + k\widetilde{\mu^{k-1}S_t}\sigma_0^2\right)d\widetilde{W}_t/\sigma_0. \quad \text{(A.2)}$$

The assumption used in this article to close this recursive system is that beliefs about the growth rate μ_t conditional on the current value of the regime indicator S_t are symmetric. To be explicit, put $R \equiv 1 - S$ and define the processes

$$\begin{aligned}
m_t^{(1)} &\equiv \widetilde{E}_t(\mu_t \mid S_t = 1) = \widetilde{\mu S}_t \widetilde{S}_t^{-1} 1_{\{\widetilde{S}_t > 0\}} \\
m_t^{(0)} &\equiv \widetilde{E}_t(\mu_t \mid S_t = 0) = \widetilde{\mu R}_t \widetilde{R}_t^{-1} 1_{\{\widetilde{R}_t > 0\}} \\
v_t^{(1)} &\equiv \widetilde{\mathrm{var}}_t(\mu_t \mid S_t = 1) = \widetilde{\mu^2 S}_t \widetilde{S}_t^{-1} 1_{\{\widetilde{S}_t > 0\}} - \left(m_t^{(1)}\right)^2 \\
v_t^{(0)} &\equiv \widetilde{\mathrm{var}}_t(\mu_t \mid S_t = 0) = \widetilde{\mu^2 R}_t \widetilde{R}_t^{-1} 1_{\{\widetilde{R}_t > 0\}} - \left(m_t^{(0)}\right)^2 \\
s_t^{(1)} &\equiv \widetilde{\mathrm{skew}}_t(\mu_t \mid S_t = 1) \\
&= \widetilde{\mu^3 S}_t \widetilde{S}_t^{-1} 1_{\{\widetilde{S}_t > 0\}} - m_t^{(1)}\left(3v_t^{(1)} + \widetilde{\mu^2 S}_t \widetilde{S}_t^{-1}\right) \\
s_t^{(0)} &\equiv \widetilde{\mathrm{skew}}_t(\mu_t \mid S_t = 0) \\
&= \widetilde{\mu^3 R}_t \widetilde{R}_t^{-1} 1_{\{\widetilde{R}_t > 0\}} - m_t^{(0)}\left(3v_t^{(0)} + \widetilde{\mu^2 R}_t \widetilde{R}_t^{-1}\right),
\end{aligned}$$

where $1_{\{\bullet\}}$ denotes the indicator function, which is 1 when its argument is true and zero otherwise. Now the closure condition is simply $s_t^{(1)} = s_t^{(0)} = 0$, $\forall\ t$. (Notice that this does not restrict the fourth or higher moments of the conditional distributions.)

This assumption gives expressions for two moments in terms of lower ones. When Itô's Lemma and the exact filter above are used to derive evolution equations for the central conditional moments $m_t^{(1)}$, $m_t^{(0)}$, $v_t^{(1)}$, and $v_t^{(0)}$, these two conditions suffice to eliminate all reference to higher moments. After some algebra, the resulting system is

$$d\widetilde{S}_t = (\lambda_0 - (\lambda_0+\lambda_1)\widetilde{S}_t)\,dt + \widetilde{\mathrm{cov}}_t(\mu, S)\frac{d\widetilde{W}_t}{\sigma_0} \quad \text{(A.3)}$$

$$dm_t^{(1)} = -\frac{\widetilde{\mathrm{cov}}_t(\mu, S)}{\widetilde{S}_t}\left[\frac{\lambda_0}{\widetilde{S}_t} + \frac{(\sigma_0^2 + v_t^{(1)})}{\sigma_0^2}\right] \times 1_{\{\widetilde{S}_t > 0\}}dt + \left(\sigma_0^2 + v_t^{(1)}\right)\frac{d\widetilde{W}_t}{\sigma_0} \quad \text{(A.4)}$$

$$dm_t^{(0)} = -\frac{\widetilde{\mathrm{cov}}_t(\mu, R)}{\widetilde{R}_t}\left[\frac{\lambda_1}{\widetilde{R}_t} + \frac{v_t^{(0)}}{\sigma_0^2}\right] \times 1_{\{\widetilde{R}_t > 0\}}dt + v_t^{(0)}\frac{d\widetilde{W}_t}{\sigma_0} \quad \text{(A.5)}$$

$$dv_t^{(1)} = \left\{\sigma_0^2 - \frac{(\sigma_0^2 + v_t^{(1)})^2}{\sigma_0^2} + \lambda_0\frac{\widetilde{R}_t}{\widetilde{S}_t} \times \left[(v_t^{(0)} - v_t^{(1)}) + (m_t^{(1)} - m_t^{(0)})^2\right]1_{\{\widetilde{S}_t > 0\}}\right\}dt \quad \text{(A.6)}$$

$$dv_t^{(0)} = \left\{-\frac{(v_t^{(0)})^2}{\sigma_0^2} + \lambda_1\frac{\widetilde{S}_t}{\widetilde{R}_t} \times \left[(v_t^{(1)} - v_t^{(0)}) + (m_t^{(1)} - m_t^{(0)})^2\right]1_{\{\widetilde{R}_t > 0\}}\right\}dt. \quad \text{(A.7)}$$

Now a technical issue arises: the existence of a solution. The original system that this approximates is known (from filtering theory) to have a unique strong solution. But this theory no longer applies because the quantities above are not a true filter for any similar inference problem. Because the system is an approximation anyway, we may simply deal with the issue by replacing it with a further approximation. To this end, use the substitution $\widetilde{\mathrm{cov}}_t(\mu, S) \equiv \widetilde{S}_t\widetilde{R}_t(m_t^{(1)} - m_t^{(0)})$ and define the quantities $\tilde{m}_t^{(1)}$, $\tilde{m}_t^{(0)}$, $\tilde{v}_t^{(1)}$, $\tilde{v}_t^{(0)}$ by the differentials

$$d\tilde{m}_t^{(1)} = -\kappa(\Phi_t)\widetilde{R}_t\left[\frac{\lambda_0}{\widetilde{S}_t} + \frac{(\sigma_0^2 + (\tilde{v}_t^{(1)} \wedge K))}{\sigma_0^2}\right] \times 1_{\{\widetilde{S}_t > 1/K\}}\, dt + \left(\sigma_0^2 + \tilde{v}_t^{(1)}\right)\frac{d\widetilde{W}_t}{\sigma_0} \quad \text{(A.8)}$$

$$d\tilde{m}_t^{(0)} = -\kappa(\Phi_t)\widetilde{S}_t\left[\frac{\lambda_1}{\widetilde{R}_t} + \frac{(\tilde{v}_t^{(0)} \wedge K)}{\sigma_0^2}\right] \times 1_{\{\widetilde{R}_t > 1/K\}}\, dt + \tilde{v}_t^{(0)}\frac{d\widetilde{W}_t}{\sigma_0} \quad \text{(A.9)}$$

$$d\tilde{v}_t^{(1)} = \left\{\sigma_0^2 - \frac{(\sigma_0^2 + \tilde{v}_t^{(1)})^2}{\sigma_0^2} + \lambda_0\frac{\widetilde{R}_t}{\widetilde{S}_t} \times \left[\kappa(\tilde{v}_t^{(0)} - \tilde{v}_t^{(1)}) + (\kappa(\Phi_t))^2\right]1_{\{\widetilde{S}_t > 1/K\}}\right\} dt \quad \text{(A.10)}$$

$$d\tilde{v}_t^{(0)} = \left\{-\frac{(\tilde{v}_t^{(0)})^2}{\sigma_0^2} + \lambda_1\frac{\widetilde{S}_t}{\widetilde{R}_t} \times \left[\kappa(\tilde{v}_t^{(1)} - \tilde{v}_t^{(0)}) + (\kappa(\Phi_t))^2\right]1_{\{\widetilde{R}_t > 1/K\}}\right\} dt, \quad \text{(A.11)}$$

where $\Phi \equiv (\tilde{m}^{(1)} - \tilde{m}^{(0)})$, $\kappa(a) \equiv ((a \wedge K) \vee -K)$, and K is some large constant. Clearly this system can be made arbitrarily close to the previous one. Equations (6)–(9) in the text correspond to (A.8)–(A.11), with the truncation suppressed for notational simplicity.

It is now straightforward to apply standard existence results to deduce the existence of a unique strong solution for $\widetilde{S}$, $\tilde{m}_t^{(1)}$, $\tilde{m}_t^{(0)}$, $\tilde{v}_t^{(1)}$, and $\tilde{v}_t^{(0)}$. These may then be interpreted as estimators (though not the exact conditional expectations) for the unknown conditional moments of μ and S. Using these, we may then define estimators of the unconditional moments according to $\tilde{\mu} = \widetilde{S}\tilde{m}^{(1)} + \widetilde{R}\tilde{m}^{(0)}$, $\widetilde{\mathrm{var}}_t(\mu) = \widetilde{S}\tilde{v}^{(1)} + \widetilde{R}\tilde{v}^{(0)} + \widetilde{S}\widetilde{R}\Phi^2$, and so on.

A.2. Comparison of Filters

One might reasonably wonder whether modeling investors as using the modified estimators (given by the system above) entails an assumption of unreasonably suboptimal behavior, or, more delicately, whether it entails behavior that differs in any economically meaningful way from that of investors using the exact posterior means. The answer to both is no.

Under the model of Section 2, the only estimated quantity that enters into the determination of exchange rates is $\tilde{\mu}$, the growth rate estimator. So, in comparing the economic implications of the two filtering systems, it suffices to examine the differences in this quantity. To do this, realizations of the underlying fundamental variables in the economy (S, μ, and Δ) are simulated for a variety of parameter configurations, and the two estimators are computed. [The true posterior mean can be calculated using the method of Proposition 2.1 in Johnson (in press) when the observation series Δ is available.] Each configuration is simulated at daily frequency for 25 years.

Let $\tilde{\mu}$ be the estimator derived from the system (A.3), (A.8)–(A.11) and let $\tilde{\mu}^*$ be the exact conditional expectation. The overall performance of $\tilde{\mu}$ can be assessed by the root mean the square of $(\tilde{\mu} - \mu)$. Theoretically, $\tilde{\mu}^*$ minimizes the mean square error. So the relative and absolute efficiency of $\tilde{\mu}$ is measured by the ratio of mean square errors. To detect any imputed behavioral bias by users of $\tilde{\mu}$, the mean difference $\tilde{\mu} - \tilde{\mu}^*$ is also reported. The mean absolute difference gives a measure of the overall ability of $\tilde{\mu}$ to track $\tilde{\mu}^*$. The truncation constant K was set to 10^6. The results are shown in Table 3.

The first four rows of the table are parameter choices analyzed by Johnson (in press). The last two are the values fitted in Section 3 for the DEM and the JPY (cf. Table A.1). The root mean square errors show that the modified estimator suffers no efficiency loss relative to the true conditional expectation (and even outperforms it in some realizations). The second column from the right shows that there is no systematic deviation of one from the other. (The differences are scaled by σ_0.) The last column shows that, in fact, they are tracking one another extremely precisely. Indeed, the figures may not quite convey how close the two are. On graphs, they literally plot on top of each other. The differences are of the order of the margins of error of the numerical integrations used to evaluate each.

In practical terms, the last column can be converted to percentage differences in exchange rates by multiplying by the structural parameters $\theta_2\sigma_0$ for the simulated economies. Doing this for the JPY and DEM models gives values of .0015 and .0022, about the size of the bid–offer spread in the data series used in Section 3.

Although simulations are not proof, this is certainly strong evidence that two economies—one consisting of agents using $\tilde{\mu}^*$ and one of agents using $\tilde{\mu}$—would be observationally indistinguishable. It should be noted, though, that this may be partly due to the simplicity of the pricing model used, in which higher moments (and, in particular, risk) play no role.

Table A.1. Comparison of Estimators

Parameters			Root mean square		Relative	Mean	
σ_0	λ_0	λ_1	$(\tilde{\mu} - \mu)$	$(\tilde{\mu}^* - \mu)$	efficiency	$(\tilde{\mu} - \tilde{\mu}^*)$	$\lvert\tilde{\mu} - \tilde{\mu}^*\rvert$
.030	2.00	6.00	.4466	.4507	1.009	−.0008	.0058
.030	0.06	6.00	.3512	.3538	1.007	.0007	.0056
.030	17.3	52.0	.4975	.4910	.987	−.0111	.0133
. 030	0.53	52.0	.1336	.1338	1.002	−.0010	.0013
.028	0.23	8.13	.2043	.2050	1.003	.0037	.0039
.068	0.45	2.65	.6908	.6811	.986	−.0016	.0223

NOTE: The table compares the growth rate estimator $\tilde{\mu}$ of the system (A.3), (A.8)–(A.11) with the posterior mean $\tilde{\mu}^*$ in simulated 25-year histories for six parameter configurations. Relative efficiency is the ratio of the fifth column to the fourth. All averages have been divided by the relevant scale factor σ_0.

A.3. Discretization and Likelihood

To use the model empirically, Euler approximations are used to discretize equations (A.3), (A,8)–(A.11):

$$\Delta\widetilde{S}_t = (\lambda_0 - (\lambda_0+\lambda_1)\widetilde{S}_t)\Delta t + \widetilde{S}_t\widetilde{R}_t\Phi_t\frac{\Delta\widetilde{W}_t}{\sigma_0} \tag{A.12}$$

$$\Delta\tilde{m}_t^{(1)} = -\kappa(\Phi_t)\widetilde{R}_t\left[\frac{\lambda_0}{\widetilde{S}_t} + \frac{(\sigma_0^2 + (\tilde{v}_t^{(1)} \wedge K))}{\sigma_0^2}\right] \times 1_{\{\widetilde{S}_t > 1/K\}}\Delta t + \left(\sigma_0^2 + \tilde{v}_t^{(1)}\right)\frac{\Delta\widetilde{W}_t}{\sigma_0} \tag{A.13}$$

$$\Delta\tilde{m}_t^{(0)} = -\kappa(\Phi_t)\widetilde{S}_t\left[\frac{\lambda_1}{\widetilde{R}_t} + \frac{(\tilde{v}_t^{(0)} \wedge K)}{\sigma_0^2}\right] * \times 1_{\{\widetilde{R}_t > 1/K\}}\Delta t + \tilde{v}_t^{(0)}\frac{\Delta\widetilde{W}_t}{\sigma_0} \tag{A.14}$$

$$\Delta\tilde{v}_t^{(1)} = \left\{\sigma_0^{(2)} - \frac{(\sigma_0^2 + \tilde{v}_t^{(1)})^2}{\sigma_0^2} + \lambda_0\frac{\widetilde{R}_t}{\widetilde{S}_t} \times \left[\kappa(\tilde{v}_t^{(0)} - \tilde{v}_t^{(1)}) + (\kappa(\Phi_t))^2\right]1_{\{\widetilde{S}_t > 1/K\}}\right\}\Delta t \tag{A.15}$$

$$\Delta\tilde{v}_t^{(0)} = \left\{-\frac{(\tilde{v}_t^{(0)})^2}{\sigma_0^2} + \lambda_1\frac{\widetilde{S}_t}{\widetilde{R}_t} \times \left[\kappa(\tilde{v}_t^{(1)} - \tilde{v}_t^{(0)}) + (\kappa(\Phi_t))^2\right]1_{\{\widetilde{R}_t > 1/K\}}\right\}\Delta t, \tag{A.16}$$

where $\Delta\widetilde{W}_t \sim \mathcal{N}(0, \Delta t)$.

If $\{e_t\}_{t=1}^T$ are the observations of the log exchange rate, then the log likelihood function for the parameter vector $\{\lambda_0, \lambda_1, \sigma_0, \theta_2, \widetilde{S}_0, \tilde{m}_0^{(0)}, \tilde{m}_0^{(1)}, \tilde{v}_0^{(1)}, \tilde{v}_0^{(0)}\}$ is given (up to an additive constant) by

$$\sum_{t=1}^{T}\left\{-\log\tilde{h}_t - \frac{1}{2}\left(\frac{\Delta e_t - \tilde{\mu}_t\Delta t}{\tilde{h}_t}\right)^2\right\}, \tag{A.17}$$

where $\tilde{h}$ and $\tilde{\mu}$ are given in terms of the preceding moments by Equations (11)–(13).

Evaluation of the log likelihood for a given parameter vector is then performed by iterating through the data and at each time t performing the following steps:

1. Impute the unexpected innovation $\Delta\widetilde{W}_t$ from the return Δe_t as $(\Delta e_t - \tilde{\mu}_t\Delta t)/\tilde{h}_t$.
2. Compute the addition to the log likelihood from (A.17).
3. Update the state variables above via (A.12)–(A.16) and construct $\tilde{\mu}_{t+1}$ and $\tilde{h}_{t+1}$ from these according to (11)–(13).

Thus each parameter vector, together with the data, determines a complete history of agents' conditional beliefs in-sample. No additional data augmentation is required to evaluate these latent variables.

APPENDIX B: GENERATING THE POSTERIOR SAMPLE

This appendix describes the procedure used to produce the samples used in the article from the joint posterior distribution of the model's parameters.

B.1. Data

The data consist of daily exchange rates for the U.S. dollar against the Japanese yen and the German mark from the beginning of 1980 to 3/31/94. The series are those used in Drost Nijman, and Werker (1998) and were taken from the web site of the *Journal of Business and Economic Statistics (ftp://www.amstat.org/JBES_View/98-2-Apr.)*. All weekday returns are present. However, quotes from Bank of England holidays, which seem to be repetitions of the previous day's quotes, are deleted. This leaves 3,605 returns for each series. For further information on the data, consult the above reference.

B.2. Parameters

As presented in Section 2, there are four unknown parameters in the model: $\sigma_0, \lambda_0, \lambda_1, \theta_2$. Up to five additional parameters may be used to describe the initial conditions. These are the starting values of the state variables $\widetilde{S}, \tilde{m}^{(0)}, \tilde{m}^{(1)}, \tilde{v}^{(0)}$, and $\tilde{v}^{(1)}$ or, equivalently, $\widetilde{S}, \tilde{\mu}, \widetilde{\mu S}, \widetilde{\mu^2}$, and $\widetilde{\mu^2 S}$. In the structural model, these describe the time-zero beliefs about the unknown growth rate and persistence. The simplifying assumption that this initial distribution is binomial is imposed, reducing the number of start-up parameters to three. The full set was found to be weakly identified by the data, leading to poor convergence properties of the Markov chain.

Last, one additional parameter is included for the effective time length of returns over more than 1 day. That is, the exchange rate equation (10) (in discretized form) is modified to

$$\Delta e_t = \tilde{\mu}\Delta t + d^{I_t}\tilde{h}\Delta\widetilde{W}_t, \tag{B.1}$$

where I_t is an indicator variable taking the value 1 when the interval Δt is not one calendar day. This is the only deseasonalization used.

In total, then, the parameter space has eight dimensions. Constructing a convergent sampling routine is greatly facilitated by reducing dependencies between the parameters in the posterior through transforming the original variables. The mappings shown in Table B.1 were found to yield significant improvement in this regard.

B.3. Priors

Convergence conditions for the Metropoli–Hastings algorithm implemented below include the existence of a proper posterior distribution. This necessitates the use of a proper prior. To minimize the impact of the choice of this distribution, an extremely dispersed density is employed. That choice and its effect are now described.

The specific assumption used in the sampler later is that

$$\omega \sim \mathcal{N}(0, c^2 \cdot I),$$

where c is a large constant and I is the 8×8 identity matrix. Both chains reported were run with $c = 64$. To put this in perspective, the largest posterior standard deviation of any parameter (for either currency) turns out to be 2.1. Moreover, the

Table B.1. Parameter Transformations Used in the Markov Chain Monte Carlo Algorithm

Transformed parameter	Function of original parameters	Original parameter	Function of transformed parameters
ω_1	$\log(\theta_2)$	θ_2	$\exp(\omega_1)$
ω_2	$\log(\sigma_0)$	σ_0	$\exp(\omega_2)$
ω_3	$\log(\lambda_1+\lambda_0)$	λ_0	$\dfrac{\exp(\omega_3+\omega_4)}{1+\exp(\omega_4)}$
ω_4	$\operatorname{logit}(\overline{S})$	λ_1	$\dfrac{\exp(\omega_3)}{1+\exp(\omega_4)}$
ω_5	$\operatorname{logit}(\widetilde{S}_0-\overline{S})$	$\widetilde{S}_0$	$\dfrac{\exp(\omega_5)}{1+\exp(\omega_5)}+\dfrac{\exp(\omega_4)}{1+\exp(\omega_4)}$
ω_6	$\tilde{\mu}_0/\sigma_0$	$\tilde{m}_0^{(1)}$	$\exp(\omega_2)\dfrac{\omega_6+\omega_7}{2}$
ω_7	$(\tilde{m}_0^{(1)}-\tilde{m}_0^{(0)})/\sigma_0$	$\tilde{m}_0^{(0)}$	$\exp(\omega_2)\dfrac{\omega_6-\omega_7}{2}$
ω_8	$\log(d)$	d	$\exp(\omega_8)$

NOTE: The first two columns show transformations of the model parameters used in the construction of the posterior sample. The transformation yields a set of unrestricted parameters that exhibit lower covariance under the posterior, as described in Appendix B. ω_1–ω_4 transform the structural parameters. ω_5–ω_7 transform the start-up values the state variables. ω_8 transforms the multiday dummy. The inverse mapping used to recover the original parameters is also shown (last two columns).

largest absolute posterior mean is 6.2. Thus the entire posterior is effectively contained within a fraction of one prior standard deviation from the prior mean. As a result the prior is extremely flat over the range of parameter values sampled for both currencies. Clearly, too, the same would apply if any reasonable nonzero prior mean were used with this variance. Likewise, with this large a dispersion, the choice of normality is irrelevant because only a small neighborhood of the density's domain is ever encountered.

An interesting numerical measure of the effect of a particular prior on the resulting posterior is the fraction of variation in the log posterior density due to variation in the log of the prior. That is,

$$\operatorname{Var}[\log(p(\omega))]/\operatorname{Var}[\log(p(\omega \mid Y))],$$

where $p(\omega)$ and $p(\omega \mid Y)$ are the prior and posterior densities, respectively, and the variance is computed with respect to the posterior. Variance in the denominator can be due to nonuniformities in either the likelihood function or the prior. The fraction thus captures the relative flatness of the two product terms over the region of interest in the parameter space. A uniform prior would produce zero. Combining two independent univariate normals with the same mean (e.g., a normal likelihood and a normal prior) would give just the square of the ratio of the prior precision to that of posterior. For Markov chain Monte Carlo (MCMC) inference, numerator and denominator are both readily computed from the posterior sample, with neither normalization constant required. In the present case, the fraction is .00000052 for the yen and 0.0000031 for the mark.

The use of vague priors in this article does not mean that no information has been imposed. Use of different transformations of the model variables can cause seemingly flat priors to become highly informative. Section 3.1 examined some functions of the structural parameters that related to implied time series properties of returns. It is worth noting what the current prior imposes on them. Their marginal priors are evaluated by simply drawing from the prior and retransforming the draws.

For the half-life of volatility shocks [cf. Equation (15)], the prior places half its mass at the lower bound, 0.347, and distributes the remainder with extremely wide dispersion (median $\approx 10^8$) over the positive half-axis. Similarly, the average volatility of volatility [Equation (16)] places roughly 3/4 of its mass at zero (or below 0.0001, on the scale of Fig. 4), with the remainder being diffuse (median $\approx 10^{18}$). The prior is, then, very informative, due to the point masses at the lower bounds of these parameters. The conclusion in the text that the posterior has no mass at these points is thus even more significant, in that the data had to overcome strong prior inclinations.

Another basic sense in which the article's priors are informative is that, by using logs of the structural variables, negative values have been precluded. As noted in the text, σ_0, λ_0, λ_1, and θ_2 are all restricted to be positive based on their interpretation in the underlying filtering problem described in Section 2. However, no further weight has been placed on that interpretation in the sense of incorporating prior information about money supply volatilities or interest rate elasticities, and so on. This is appropriate here because it is the implied return dynamics that are the object of study, not the model of exchange rate determination. The intention is to allow the data the maximum flexibility to accommodate the restrictions that the form of the model already imposes in describing those dynamics.

B.4. Sampling

Evaluation of the likelihood function for the model is straightforward, as described at the end of Section A.3. This makes the log posterior similarly straightforward. For a given point, ω, in the transformed parameter space simply (a) invert the transformations given in Table B.1; (b) loop through the data to compute the log likelihood; and (c) add the log prior.

However, expressing the posterior density in terms of standard conditional distributions for the unknown parameters is impossible, rendering Gibbs sampling infeasible. In this situation, the Metropolis–Hastings algorithm still offers a way to draw variates from the posterior by constructing a Markov chain kernel from a suitably chosen dominating density (called the proposal density). Because it is impossible to establish the required domination property here, the two-stage acceptance/rejection regenerative kernel of Tierney (1994, 1996) is employed. Tierney has shown that this procedure is convergent under very weak conditions, even when the proposal density does not blanket the posterior everywhere.

To be computationally feasible, the proposal density must still closely resemble the posterior in general shape. A first step is to locate the modes by using an optimizer. This turns up a surprising feature: bimodality for each currency. (This is a consequence of the two possible signs for the initial correlation between returns and volatility, about which the data do not speak strongly. Imposing either sign on the start-up value of $\tilde{m}_0^{(1)}-\tilde{m}_0^{(0)}$, to which this correlation is equivalent,

causes the bimodality to collapse.) The bimodality was handled by designing separate proposal densities for each mode and mixing by an appropriate independent binomial. Multivariate Cauchy densities were used for the individual modes, matching the second-derivative matrices of their logs to the numerically calculated curvature of the posterior at the mode.

Tierney's algorithm then proceeds by the following steps:

1. Draw a candidate parameter $\hat{\omega}$ from the proposal density and compute the ratio

$$\hat{\gamma} = \frac{c \cdot h(\hat{\omega})}{\pi(\hat{\omega})},$$

where $\pi(\,)$ is the unnormalized posterior density (likelihood $\times$ prior), $h(\,)$ is the proposal density, and c is a tuning constant to be determined.

2. Draw an independent uniform variate u. If $u \leq \hat{\gamma}$ discard $\hat{\omega}$ and return to step 1. Else, proceed to step 3.

3. Let $\gamma_{\text{new}} = \hat{\gamma}$ and compute

$$\gamma_{\text{old}} = \frac{ch(\omega^{(i-1)})}{\pi(\omega^{(i-1)})}$$

for the previous draw $\omega^{(i-1)}$ in the chain. Define

$$\alpha \equiv \min[1, \gamma_{\text{old}}] \cdot 1_{\{\gamma_{\text{new}} > 1\}} + \min\left[1, \frac{\gamma_{\text{old}}}{\gamma_{\text{new}}}\right] \cdot 1_{\{\gamma_{\text{new}} \leq 1\}}.$$

4. Draw an independent uniform variate v. If $v \leq \alpha$ set $\omega^{(i)} = \hat{\omega}$. Else set $\omega^{(i)} = \omega^{(i-1)}$.

5. Return to step 1 until N draws have been accepted.

Intuitively, steps 1 and 2 attempt to draw from π by a standard rejection algorithm which one would use if ch were known to dominate it. Steps 3 and 4 "correct" this draw for the fact that domination might fail in certain regions. Since these regions are undersampled, the procedure sets the second-stage acceptance level α to make it easier for the chain to move into them and harder for it to move out.

Choice of the tuning parameter c determines how often domination will fail. A large value makes it extremely hard to get through the first stage of the algorithm but more likely that the chain will move to the new draw at the second stage. In the implementation here, c was chosen so that the first-stage acceptance rate was approximately 2%. This produced chains whose probability of moving at each step was between 20% and 40%.

The chains were cycled for 20,000 iterations. Note that because successive candidate draws are independent, conditional on the chain moving, the new points are i.i.d. Thus the chain forgets its old position each time it moves: the current position of the chain conveys no information about the previous distinct position. Hence the choice of initial value of the chain is irrelevant as soon as the first transition has been made. To effectively ensure this, the first 2,000 values of each sample were discarded.

With any MCMC sampler there are two important issues to address in evaluating the success of the algorithm: whether the chain has been run long enough to forget its initial conditions and whether, given the autocorrelation exhibited by the chain, a large enough sample has been drawn to represent the stationary distribution accurately.

As emphasized in the survey of Cowles and Carlin (1996), convergence diagnosis for Markov chains is best approached from multiple directions simultaneously, since all existing techniques are vulnerable to failure of one sort or another. Three separate checks were applied to each of the generated chains.

First, the stationarity tests of Geweke (1992) were applied to check that subsamples from the chain look similar. Some traces of an initial transient phase were detected for two of the Deutschemark parameters, which was adequately dealt with by dropping an additional 2,000 draws (leading to changes in all reported posterior means of less than .02%). Otherwise the test indicated good convergence characteristics for all quantities.

Next, tests proposed by Raftery and Lewis (1992) were performed to assess the length of chains. Specifically, these tests verified that, given the degree of autocorrelation exhibited by each chain, all of the parameters have been sampled sufficiently to conclude that their 5th and 95th percentiles are estimated within $\pm.01$ with 90% probability.

Last, both stationarity and run length were also assessed via the method of Heidelberger and Welch (1983). Again the verdict is positive: the tests imply that all parameters' posterior means have been estimated within 10% with 95% probability, and no component paths exhibit the unusual tendency to wonder from their full-sample means.

Computation of all of the above test statistics was conducted using CODA (Best Cowles, and Vines 1995). Cowles and Carlin (1996) provide a detailed discussion of these and many other procedures.

In conclusion, the algorithm appears to have succeeded in simulating a convergent Markov chain whose stationary distribution is the posterior of the model's parameters. Samples drawn from the chain were long enough to permit accurate inference for the quantities examined in the text. Further details of the algorithm and files containing the sampled chains are available from the author upon request.

[Received December 1999. Revised March 2001.]

REFERENCES

Barberis, N., Shleifer, A., and Vishny, R. (1998), "A Model of Investor Sentiment," *Journal of Financial Economics*, 49, 307–344.

Barsky, R. B., and DeLong, J. B. (1993), "Why Does the Stock Market Fluctuate?" *Quarterly Journal of Economics*, 108, 291–311.

Bates, D. S. (1996), "Dollar Jump Fears, 1984–1992: Distributional Abnormalities Implicit in Currency Futures Options Prices," *Journal of International Money and Finance*, 15, 65–93.

Best, N., Cowles, M. K., and Vines, K. (1995), *CODA: Convergence Diagnosis and Output Analysis Software for Gibbs Sampling Output* (Version 0.30), MRC Biostatistics Unit, Cambridge, UK.

Campa, J. M., Chang, P. H. K., and Reider, R. L. (1998), "Implied Exchange Rate Distributions: Evidence From OTC Options Markets," *Journal of International Money and Finance*, 17, 117–160.

Chib, S., and Greenberg, E. (1995), "Understanding the Metropolis Hastings Algorithm," *The American Statistician*, 49, 327–336.

Cowles, M. K., and Carlin, B. P. (1996), "Markov Chain Monte Carlo Convergence Diagnostics: A Comparative Review," *Journal of the American Statistical Association*, 91, 883–904.

Dacorogna, M. M., Müller, U. A., Olsen, R. B., and Pictet, O. V. (1998), "Modelling Short-Term Volatility With GARCH and HARCH Models," in *Nonlinear Modelling of High Frequency Financial Time Series*, eds. C. Dunis and B. Zhou, New York: Wiley.

David, A. (1997), "Fluctuating Confidence in Stock Markets: Implications for Returns and Volatility," *Journal of Financial and Quantitative Analysis*, 32, 427–462.

Detemple, J. B. (1986), "Asset Pricing in a Production Economy With Incomplete Information," *Journal of Finance*, 41, 383–391.

Detemple, J. B. (1991), "Further Results on Asset Pricing With Incomplete Information," *Journal of Economic Dynamics and Control*, 15, 425–453.

Drost, F. C., Nijman, T., and Werker, J. M. (1998), "Estimation and Testing in Models Containing Both Jump and Conditional Heteroscedasticity," *Journal of Business & Economic Statistics*, 16, 237–243.

Evans, M. D. D. (1998), "Dividend Variability and Stock Market Swings," *Review of Economic Studies*, 65, 711–740.

Evans, M. D. D., and Lewis, K. K. (1995), "Do Expected Shifts in Inflation Affect Estimates of the Long-Run Fisher Relation?" *Journal of Finance*, 50, 225–253.

Feldman, D. (1989), "The Term Structure of Interest Rates in a Partially Observable Economy," *Journal of Finance*, 44, 789–812.

Gelfand, A. E. (1996), "Model Determination Using Sampling-Based Methods," in *Markov Chain Monte Carlo in Practice*, eds. W. R. Gilks, S. Richardson, and D. J. Spiegelhalter, London: Chapman and Hall.

Gelman, A., Meng, X., and Stern, H. (1996), "Posterior Predictive Assessment of Model Fitness via Realized Discrepancies" (with discussion), *Statistica Sinica*, 6, 733–807.

Geweke, J. (1992), "Evaluating the Accuracy of Sampling-based Approaches to Calculating Posterior Moments," in *Bayesian Statistics 4*, eds. J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith, Oxford: Clarendon Press.

Heidelberger, P., and Welch, P. (1983), "Simulation Run Length Control in the Presence of an Initial Transient," *Operations Research*, 31, 1109–1144.

Jackwerth, J. C., and Rubenstein, M. (2000), "Recovering Probabilities and Risk Aversion From Options Prices and Realized Returns," in *Essays in Honor of Fisher Black*, ed. B. Lehman, Oxford: Oxford University Press.

Johnson, T. C. (1999), "Unobservable Persistence: An Economic Theory of Stochastic Volatility," Ph.D. thesis, University of Chicago.

——— (in press), "Return Dynamics When Persistence Is Unobservable," *Mathematical Finance*.

Meng, X.-L. (1994), "Posterior Predictive p-Values," *Annals of Statistics*, 22, 1142–1160.

Müller, U. A., Dacorogna, M. M., Davé, R. D., Olsen, R. B., Pictet, O. V., and von Weizsäcker, J. E. (1997), "Volatilities of Different Time Resolutions—Analyzing the Dynamics of Market Components," *Journal of Empirical Finance*, 4, 213–240.

Raftery, A. L., and Lewis, S. (1992), "How Many Iterations in the Gibbs Sampler?" in *Bayesian Statistics 4*, ed. J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith. Oxford: Clarendon Press.

Sentana, E. (1995), "Quadratic ARCH Models," *Review of Economic Studies*, 62, 639–661.

Tierney, L. (1994), "Markov Chains for Exploring Posterior Distributions," (with discussion), *Annals of Statistics*, 22, 1701–1762.

Tierney, L. (1996), "Introduction to General State-Space Markov Chain Theory," in *Markov Chain Monte Carlo in Practice*, ed. W. R. Gilks, S. Richardson, and D. J. Spiegelhalter. London: Chapman & Hall.

Veronesi, P. (1999), "Stock Market Overreaction to Bad News in Good Times: A Rational Expectations Equilibrium Model," *Review of Financial Studies*, 12, 975–1007.

——— (2000), "How Does Information Quality Affect Stock Returns?", *Journal of Finance*, 55, 807–837.

Wang, J. (1993), "A Model of Intertemporal Asset Prices Under Asymmetric Information," *Review of Economic Studies*, 60, 249–282.