#5-99

Simulated Likelihood Estimation of Diffusion with an Application to Exchange Rate Dynamics in Incomplete Markets

First Draft: October 1995
This Draft: January 2001

(Formerly titled "Simulated Likelihood Estimation of Multivariate Diffusions with an Application to Interest Rates and Exchange Rates with Stochastic Volatility")

Michael W. Brandt
The Wharton School
University of Pennsylvania and NBER

and

Pedro Santa-Clara
Anderson Graduate School of Management
University of California, Los Angeles

Finance Working Paper
Sponsored by:

Simulated Likelihood Estimation of Diffusions with an Application to Exchange Rate Dynamics in Incomplete Markets*

Michael W. Brandt
The Wharton School
University of Pennsylvania†
and NBER

Pedro Santa-Clara
The Anderson School
UCLA‡

First Draft: October 1995
This Draft: January 2001 §

Abstract

We present an econometric method for estimating the parameters of a diffusion model from discretely sampled data. The estimator is transparent, adaptive, and inherits the asymptotic properties of the generally unattainable maximum likelihood estimator. We use this method to estimate a new continuous-time model of the joint dynamics of interest rates in two countries and the exchange rate between the two currencies. The model allows financial markets to be incomplete and specifies the degree of incompleteness as a stochastic process. Our empirical results offer several new insights into the dynamics of exchange rates.

*This paper and earlier versions circulated under different titles benefitted from the comments of Yacine Aït-Sahalia, João Amaro de Matos, Domenico Cuoco, Lars Peter Hansen, Lane Hughston, Chris Jones, Frank de Jong, Bob Kimmel, Lars Nielsen, Monika Piazzesi, Matt Richardson, Jesús Saá-Requejo, José Scheinkman, Bill Schwert, Avanidhar Subrahmanyam, Ross Valkanov and Ivo Welch. We are grateful to an anonymous referee whose detailed and knowledgeable comments went far beyond the usual call of duty. Finally, we also thank the seminar participants at the 2000 AFA meetings, Carnegie-Mellon University, London Business School, NBER asset pricing group, NBER forecasting and empirical methods group, Northwestern University, Ohio State University, UCLA, University of Michigan, University of Pennsylvania, University of Rochester, and University of Texas for their comments and suggestions. Bing Han, Canlin Li, Jiang Luo, and Dmitry Livdan provided excellent research assistance. We are responsible for any remaining errors.

†Philadelphia, PA 19104-6367, Phone: (215) 898-3609, E-mail: brandtm@wharton.upenn.edu.

‡Los Angeles, CA 90095-1481, Phone: (310) 206-6077, E-mail: pedro.santa-clara@anderson.ucla.edu.

§The latest revision of this paper is available at: http://brandt.wharton.upenn.edu

1 Introduction

Many theoretical models in economics and finance are formulated in continuous time as a diffusion or a system of diffusions, although the data these models describe can only be sampled at discrete points in time.[1] This popularity of diffusions creates a need for effective econometric methods for estimating continuous-time models. In this paper, we present a simulation-based estimator of the parameters of a diffusion or a system of diffusions from discretely sampled data. The estimator is transparent, adaptive, and inherits the asymptotic properties of the generally unattainable maximum likelihood estimator.

Theorists in various areas like the continuous-time diffusion setting because of the tractability offered by Itô calculus. Furthermore, in financial models the continuous-time setting also plays a conceptual role. Since Black and Scholes (1973) and Merton (1971), many asset pricing models have assumed dynamic trading in continuous-time to allow markets to be complete and hence derivative payoffs or consumption trajectories to be spanned, even when there exists a continuum of states and only a few traded securities. Finally, diffusions are attractive from a statistical perspective because they are fully characterized by their instantaneous mean and variance. The continuous-time setting breaks the link between the model and the sampling frequency of the data, which is particularly important for nonlinear models that have different distributional characteristics at different sampling frequencies.[2]

As for any parametric model, maximum likelihood is the preferred method for estimating the parameters of a diffusion. Unfortunately, exact maximum likelihood estimation is only possible in a few special cases when the distribution of the discretely sampled data is known.[3] In most cases, exact maximum likelihood estimation is impossible because the likelihood function of the model cannot be evaluated explicitly, and the alternative of approximating the likelihood function has until recently proven difficult.

We show how to estimate the parameters of virtually any diffusion model by simulated maximum likelihood (SML). The SML method works as follows. We construct consistent approximations to the transition densities of the diffusion and use these approximations to evaluate the likelihood function. We then maximize this approximated likelihood function. Since the approximations to the transition densities are consistent, so is the approximation

[1] We loosely refer to Itô stochastic differential equations as diffusions.

[2] As an example, consider a GARCH model [Bollerslev (1987)], which is a nonlinear process with Gaussian transitions, specified at a daily frequency. With daily data, maximum likelihood estimation of the model is straightforward. With weekly data, in contrast, the transitions between observations are not Gaussian anymore and maximum likelihood estimation is more complicated.

[3] The distribution of the discretely sampled data is known explicitly only for diffusions with linear mean and constant or proportional variance [see, for example, Chen and Scott (1993) and Pearson and Sun (1994)]. It is known up to an inversion of the characteristic function for all affine jump-diffusions [Singleton (1999)].

to the likelihood function. This implies that asymptotically the SML estimator behaves just like the unattainable exact maximum likelihood estimator.

To approximate the transition densities, we apply an Euler discretization to the diffusion. We split the time interval between any two consecutive observations into smaller intervals and construct a high-frequency discrete time process with Gaussian transitions that converges to the diffusion as the discretization becomes finer. The transition densities of the discretization between observations are convolutions of Gaussian densities and are still unknown in closed form. Therefore, we use an intuitive and computationally efficient simulation scheme to numerically evaluate the transition densities of the Euler discretization. Since both the Euler discretization and our simulation scheme are consistent, so are also the resulting approximations to the transition densities of the diffusion.

The SML method was originally developed by Santa-Clara (1995) in an earlier version of this paper and independently by Pedersen (1995a,1995b). It has since been successfully implemented by Honoré (1997,1998), Piazzesi (2000), and Durham (2000) to estimate a variety of continuous-time term structure models, including models with jumps and with stochastic volatility. Elerian (1998) and Durham and Gallant (2000) develop extensions of the basic SML method described in this paper (based on refined numerical techniques), to further improve the performance of the estimator.

Several other approaches to approximating the likelihood function have been suggested in the literature. Lo (1988) proposes numerically solving the forward Kolmogorov partial differential equation (PDE), subject to the appropriate boundary conditions, to obtain the unknown transition densities of the diffusion. To approximate the likelihood function, this PDE must be solved for every data point, which is not only computationally demanding, especially for multivariate diffusions, but also requires a certain level of expertise in numerically solving PDEs.[4] Aït-Sahalia (2000) suggests using analytical expansions, rather than simulations, to approximate the transition densities. The advantage of analytical expansions is that for the same level of accuracy, it is computationally less demanding than simulations. The disadvantage is that, for the expansions to converge, the diffusion must first be transformed to be sufficiently Gaussian. The need for this transformation limits the transparency and adaptability of the method. Finally, Ogawa (1995), Hurn and Lindsay (1999), and Nicolau (2000) apply nonparametric density estimation to simulated data from the Euler discretizations to approximate the transition densities of the diffusion. This approach suffers from the usual problems with nonparametric density estimation: a slower convergence rate (in the number of simulations) and the curse of dimensionality.[5]

[4] Lo's approach has been implemented by Mella-Barral and Perraudin (1993).

[5] The curse of dimensionality refers to the fact that as the number of variables increases, the convergence

The SML approach relates also to the Markov chain Monte Carlo (MCMC) approach described by Eraker (1998), Jones (2000a), and Elerian, Chib, and Shephard (2000). Aside from the Bayesian interpretation, the MCMC method uses a different simulation scheme to evaluate the transition densities of the Euler discretization. Nevertheless, with sufficiently flat priors, the two methods lead to similar inferences.

The method of moments is the most popular alternative to maximum likelihood. This approach includes estimators based on unconditional moments from an Euler discretization [Chan, Karolyi, Longstaff, and Sanders (1992)], conditional moments from a binomial discretization [He (1990)], simulated unconditional moments [Duffie and Singleton (1993), Gallant and Tauchen (1997)], intentionally misspecified moments [Gouriéroux, Monfort, and Renault (1993)], moments generated from the infinitesimal generator of the diffusion [Duffie and Glynn (1997), Hansen and Scheinkman (1995), Stanton (1997)], and moments that match the unconditional density to an empirical density [Aït-Sahalia (1996a,1996b)]. Most of these estimators are consistent and asymptotically normal. However, except for Gallant and Tauchen's "efficient method of moments," which asymptotically emulates maximum likelihood estimation, they are less efficient than maximum likelihood estimation.

We use the SML method to estimate a new continuous-time model of the joint dynamics of interest rates in two countries and of the exchange rate between the two currencies. The innovation of our model is that it allows for financial markets to be incomplete and specifies the degree of incompleteness as a stochastic process.

The setup of the model is as follows. Each country has its own instantaneous interest rate process. The absence of arbitrage in currency markets determines the drift of the exchange rate between the two currencies. Specifically, the drift consists of the usual interest rate differential and a currency risk premium. We decompose this currency risk premium into a premium that compensates investors for interest rate risk, which arises from the correlation between the exchange rate and the domestic interest rate, and a premium that compensates investors for currency risk orthogonal to interest rate risk. The magnitudes of both of these premia depend on the volatility of the exchange rate. The volatility of the exchange rate, however, can only be identified through the no-arbitrage condition if the financial markets are complete. If markets are incomplete, the volatility of the exchange rate may contain an element that is orthogonal to the priced sources of risk in both countries. This element contributes to what we term the "excess volatility" of exchange rates. To capture this excess volatility, we specify a stochastic process for the degree of market incompleteness.

We estimate a parsimonious parameterization of the model. Like Cox, Ingersoll, and

rate of most nonparametric estimators to their asymptotic distribution deteriorates exponentially.

Ross (CIR, 1985), we assume that the interest rates follow square-root processes and that the market prices of interest rate risk are proportional to the square-root of the interest rates, which allows us to estimates the market prices of interest rate risk from bond prices. The market price of currency risk orthogonal to interest rate risk (or pure currency risk) is assumed either to be constant, or to depend on the exchange rate, the interest rate differential, and the volatility of the exchange rate. A novel aspect of our empirical approach is that, in order to identify the degree of market incompleteness, we make the instantaneous volatility of the exchange rate observable, rather than treat it as a latent state variable. In particular, we use the implied volatility of an at-the-money option with only one-week to maturity as a proxy for the instantaneous volatility of the exchange rate.

We present empirical results for the US dollar per British pound and US dollar per Deutsche mark exchange rates. For both currencies we find that the interest rate risk premium is negligible relative to the risk premium for currency risk orthogonal to interest rate risk. We present evidence that the market prices of pure currency risk are time-varying as a function of the exchange rate and, more importantly, the volatility of the exchange rate. However, even with time-varying market prices of currency risk, a large fraction of the exchange rate volatility is attributed to market incompleteness.

The structure of this paper is as follows. We describe the SML method in Section 2. In Section 3, we present our theoretical model of exchanges rates in incomplete markets. The empirical results are in Section 4 and we conclude in Section 5.

2 Simulated Likelihood Estimation of Diffusions

In this section, we develop the SML method for estimating the parameters of a multivariate diffusion model. We formulate the generic inference problem, describe the estimator, provide its asymptotic properties, and then discuss some practical considerations.[6] The focus is on the underlying intuition. Mathematic details and proofs are in Appendix A.

2.1 The Generic Inference Problem

Consider a K-dimensional continuous-time process Y_t characterized by the following system of stochastic differential equations:

$$dY_t = \mu(Y_t, t; \theta)dt + \Sigma(Y_t, t; \theta)dW_t, \tag{1}$$

[6]The SML method was first described by Santa-Clara (1995) in an earlier version of this paper and, independently, by Pedersen (1995a,1995b).

with initial value $Y_0 \in \mathbb{R}^K$. W_t denotes a vector of independent Brownian motions, defined in a complete probability space $\{\Omega, \mathcal{F}, P\}$. The drift vector $\mu : \mathbb{R}^K \times \mathbb{R}^+ \mapsto \mathbb{R}^K$ and the diffusion matrix $\Sigma : \mathbb{R}^K \times \mathbb{R}^+ \mapsto \mathbb{R}^K \times \mathbb{R}^K$ are functions of the vector Y_t, time t, and an L-dimensional parameter vector θ, whose true value is θ_0.[7]

The objective is to estimate the parameters θ_0, to ultimately draw inferences about the dynamics of Y_t. The following three assumptions make this a well-specified problem:

Assumption 1: The drift μ and diffusion Σ functions are infinitely differentiable with continuous and bounded derivatives of all orders.

This assumption is far stronger than the usual linear growth and uniform Lipschitz continuity conditions that are sufficient to guarantee the existence of a unique strong solution to the system of stochastic differential equations (1) [Karatzas and Shreve (1988)].[8] The extreme degree of smoothness is sufficient, but most likely not necessary, to bound the asymptotic error of the density approximations introduced in Section 2.2.

Assumption 2: The covariance matrix $\Sigma\Sigma'$ is positive definite.

This assumption guarantees that both the diffusion model and its approximation below have well-defined and smooth densities.

Assumption 3: Let $\theta \in \Theta \subset \mathbb{R}^L$, where Θ is a compact set that contains the true θ_0.

For practical reasons, the continuous-time process is sampled only at $N+1$ equally spaced points in time, denoted $t_0, t_1, \ldots t_N$.[9] Notice that although we assume equal spacing of the data for simplicity and ease of notation, our estimator extends trivially to unequally or even randomly spaced observations. Indeed, this is one of the strengths of our method.

Let $p(Y_{t_0}, Y_{t_1}, \ldots Y_{t_N}; \theta)$ denote the joint density of the discrete-time data, generated by the continuous-time diffusion model. As a function of the parameters θ, this density

[7] Although, the setup we use to present the method does not allow for discrete jumps, the estimator does generalize to jump-diffusion processes. See Piazzesi (2000) for details.

[8] This assumption generates an unpleasant gap between the diffusion models for which we can generically establish consistency and the ones commonly used in economics and finance. However, such gap between abstract asymptotic theory and practice is not uncommon. For example, the popular class of affine diffusions [Duffie and Kan (1996)] does not generally satisfy the regularity conditions required for consistency of Euler discretizations. Nevertheless, variants of the simulated method of moments [Duffie and Singleton (1993)] based on Euler discretizations are, according to simulation studies, quite effective for estimating these models.

[9] The sampling frequency often depends on the availability of data. However, sometimes it is better to sample the data less frequently than possible to reduce the market microstructure contamination of the data. See Alizadeh, Brandt, and Diebold (2000) for a discussion of this issue in the context of volatility estimation.

represents the likelihood function:

$$\mathcal{L}_N(\theta) \equiv p(Y_{t_0}, Y_{t_1}, \ldots Y_{t_N}; \theta) = p(Y_{t_0}, t_0; \theta) \prod_{n=0}^{N-1} p(Y_{t_{n+1}}, t_{n+1} | Y_{t_n}, t_n; \theta). \tag{2}$$

The equality follows from the fact that Y_t is Markovian. It shows that, in order to evaluate the likelihood function, we require the initial unconditional density $p(Y_{t_0}, t_0; \theta)$ and the N transition densities $p(Y_{t_{n+1}}, t_{n+1} | Y_{t_n}, t_n; \theta)$, for $n = 0, 1, \ldots N-1$.

The parameter vector that maximizes the log likelihood function $\ln \mathcal{L}_N(\theta)$ is the maximum likelihood estimator $\widehat{\theta}_N$ of θ_0. In our setting, the following two assumptions guarantee that the maximum likelihood estimator has the usual desirable asymptotic properties:[10]

Assumption 4: The likelihood function $\mathcal{L}_N(\theta)$ is twice continuously differentiable in θ in a neighborhood of the true parameter vector θ_0. Furthermore, $\mathrm{E}\big[[\partial \mathcal{L}_N(\theta)/\partial\theta][\partial \mathcal{L}_N(\theta)/\partial\theta']\big]$ has full rank and is bounded for all parameters $\theta \in \Theta$.

With this assumption the true parameters θ_0 are identified through the likelihood function.

Assumption 5: For every vector $\lambda \in I\!R^K$, $\lambda' I_N(\theta)\lambda \to \infty$, where:

$$I_N(\theta) = \mathrm{E}\left[\sum_{n=0}^{N-1} \frac{\partial}{\partial\theta} \ln p(Y_{t_{n+1}}, t_{n+1} | Y_{t_n}, t_n; \theta) \, \frac{\partial}{\partial\theta'} \ln p(Y_{t_{n+1}}, t_{n+1} | Y_{t_n}, t_n; \theta)\right]. \tag{3}$$

This assumption is required to establish that the maximum likelihood estimator $\widehat{\theta}_N$ is consistent. For it to hold, it is sufficient that the gradients of the transition densities are bounded. The matrix I_N is the so-called Fisher information matrix. Its inverse gives the Cramér-Rao lower bound on the covariance matrix of any consistent estimator of the parameter vector. The maximum likelihood estimator typically attains this lower bound.

The classic problem with maximum likelihood estimation of diffusion models is that in most cases a closed form expression for the likelihood function does not exist. This is because the functional forms of the unconditional density and of the transition densities implied by the model are typically unknown. To overcome this problem, we construct an estimator based on a sequence of consistent approximations to the likelihood function.

2.2 The SML Estimator

We use an efficient simulation scheme to construct consistent approximations of the transition densities $p(Y_{t_{n+1}}, t_{n+1} | Y_{t_n}, t_n; \theta)$ and, if the diffusion is stationary, of the initial unconditional

[10] The maximum likelihood estimator is consistent, asymptotically efficient, and asymptotically normal.

density $p(Y_{t_0}, t_0; \theta)$. We then use these approximations to evaluate the likelihood function. The parameter vector that maximizes the resulting approximate log likelihood function is the SML estimator. Although the asymptotic properties of the estimator depend on the accuracy of the initial approximations, in the limit, as the approximations become exact, they are identical to the properties of the exact maximum likelihood estimator.

2.2.1 Approximating the Transition Densities

To construct a consistent approximation of the transition density $p(Y_{t_{n+1}}, t_{n+1}|Y_{t_n}, t_n; \theta)$ for two adjacent discrete time observations Y_{t_n} and $Y_{t_{n+1}}$, we first discretize the process Y_t between times t_n and t_{n+1}. There exists an infinite number of discrete-time processes that approximate the diffusion process in this interval. We choose the popular Euler discretization scheme because it is both notationally and computationally convenient.[11]

Without loss of generality, we normalize the length of the interval $[t_n, t_{n+1}]$ to one and divide this interval into M subintervals of length $h = 1/M$. The Euler discretization of Y_t in $[t_n, t_{n+1}]$, denoted $\widehat{Y}_{t_n+mh}$, for $m = 0, 1, \ldots M-1$, is a Gaussian process:

$$\widehat{Y}_{t_n+(m+1)h} = \widehat{Y}_{t_n+mh} + \mu\big(\widehat{Y}_{t_n+mh}, t_n + mh; \theta\big)h + \Sigma\big(\widehat{Y}_{t_n+mh}, t_n + mh; \theta\big)\sqrt{h}\,\epsilon_{t_n+(m+1)h}, \quad (4)$$

where ϵ_{t_n} is a vector of independent standard normal random variates. The recursion starts at the initial condition $\widehat{Y}_{t_n} \equiv Y_{t_n}$. Kloeden and Platen (1992) show that, under our assumptions, the Euler approximation converges weakly to the stochastic process Y_t as $M \to \infty$.

By assumption, the one-step ahead transition densities of the Euler discretization are Gaussian. This means that the probability of $\widehat{Y}_{t_n+(m+1)h} = y$, conditional on $\widehat{Y}_{t_n+mh} = x$, is:

$$q_M\big(y, t_n + (m+1)h\big|x, t_n + mh; \theta\big) = \phi\Big(y; x + \mu\big(x, t_n + mh; \theta\big)h, V\big(x, t_n + mh; \theta\big)h\Big), \quad (5)$$

where $\phi(y; \text{mean}, \text{variance})$ denotes a multivariate normal density and $V = \Sigma\Sigma'$. The density q_M is an approximation of $p\big(y, t_n + (m+1)h\big|x, t_n + mh; \theta\big)$. The accuracy of this approximation depends on how much time h elapses between the points x and y. In the limit, as $h \to 0$, the approximation is exact.

The multi-step ahead transition densities of the Euler discretization are unknown in closed form. However, they can be evaluated through recursive integration. In particular,

[11] Kloeden and Platen (1992) describe a number of higher-order discretization schemes that are more efficient than the Euler scheme, such as the Milstein and Platen-Wagner schemes. The SML estimator extends immediately to any discretization with closed-form transitions, including the Milstein scheme. See Elerian (1998) and Durham and Gallant (2000) for details.

the probability that $\widehat{Y}_{t_n+(m+j)h} = y$, conditional on $\widehat{Y}_{t_n+mh} = x$, for $j = 2, 3, \ldots M - m$, is:

$$q_M\big(y, t_n + (m+j)h \big| x, t_n + mh; \theta\big) = \int_{I\!R} q_M\big(y, t_n + (m+j)h \big| z, t_n + (m+j-1)h; \theta\big) \times q_M\big(z, t_n + (m+j-1)h \big| x, t_n + mh; \theta\big) dz. \tag{6}$$

From equation (5), the first term in the integrand is a Gaussian density and is therefore known in closed form. The second term is itself a multi-step ahead transition density that can be computed from the recursion for $j - 1$.

With $y = Y_{t_{n+1}}$, $x = Y_{t_n}$, and $j = M - m$, equations (5) and (6) then yield an intuitive approximation of the continuous-time transition density $p\big(Y_{t_{n+1}}, t_{n+1} \big| Y_{t_n}, t_n; \theta\big)$. For the Euler discretization, the probability that $\widehat{Y}_{t_{n+1}} = Y_{t_{n+1}}$, conditional on $\widehat{Y}_{t_n} = Y_{t_n}$, is:

$$q_M\big(Y_{t_{n+1}}, t_{n+1} \big| Y_{t_n}, t_n; \theta\big) = \int_{I\!R} \phi\Big(Y_{n+1}; z + \mu\big(z, t_n + (M-1)h; \theta\big)h, V\big(z, t_n + (M-1)h; \theta\big)h\Big) \times q_M\big(z, t_n + (M-1)h \big| Y_n, t_n\big) dz. \tag{7}$$

Lemma 1 in Appendix A shows that as the accuracy of the Euler discretization increases, or formally as $M \to \infty$ and thereby $h \to 0$, the transition density of the Euler discretization converges to the corresponding transition density of the continuous-time process.

The approximate transition density $q_M\big(Y_{t_{n+1}}, t_{n+1} \big| Y_{t_n}, t_n; \theta\big)$ is still a convolution of M Gaussian densities that involves solving $M - 1$ integrals. In general, these integrals cannot be computed analytically and quadrature-based numerical integration techniques quickly become computationally infeasible as M increases. This means that the Euler discretization by itself is not sufficient to facilitate maximum likelihood estimation.

The innovation of the SML method is to interpret the integral in equation (7) as an expectation of the function ϕ of a random variable z. The distribution of this variable z is $f(z) \equiv q_M(z, t_n + (M-1)h | Y_{t_n}, t_n)$. Although we cannot easily evaluate the expectation, we can use the Euler discretization to generate a large number of independent random draws z_s, for $s = 1, 2, \ldots S$, from the distribution $f(z)$. Then, we approximate the expectation, and ultimately the corresponding continuous-time transition density p, with a sample average of the function ϕ evaluated at these random draws of z.[12]

[12]This approximation can be viewed as Rao-Blackwellization. Suppose that we are interested in the

In more detail, the method works as follows. Starting at time t_n with $\widehat{Y}_{t_n} = Y_{t_n}$, we iterate on the Euler recursion (4) exactly $M-1$ times. This results in a single draw $z_s = \widehat{Y}_{t_n+(M-1)h}$ of the discrete time process at time $t_n + (M-1)h$ from the distribution $f(z)$. We repeat this procedure S times, which yields the random sample $\{z_1, z_2, \ldots z_S\}$. Finally, we average the function ϕ over this random sample of z to approximate the expectation in equation (7).[13]

Figure 1 further illustrates the mechanics of the approximation. The solid line that connects the two adjacent discrete-time observations Y_0 and Y_1 represents the unobserved continuous-time sample path of a univariate diffusion. The four dashed lines represent incomplete ten-step discretizations of this diffusion. Each discretization is generated by starting the Euler recursion at $\widehat{Y}_0 = Y_0 = 4.00$ and iterating on it nine times. The end points $\widehat{Y}_{9/10}$ of these discretizations represent the random sample z_s, for $s = \{1, 2, 3, 4\}$. The approximation amounts to averaging the function ϕ over the random draws of z_s from $f(z)$. Graphically, we average the probabilities that the final step of the Euler discretization connects the points z_s and $Y_1 = 4.03$ along the four dotted lines.

Formally, our approximation to the transition density q_M of the Euler discretization is:

$$\begin{aligned} &\widehat{q}_{M,S}\big(Y_{t_{n+1}}, t_{n+1} \big| Y_{t_n}, t_n; \theta\big) = \\ &\qquad \frac{1}{S}\sum_{s=1}^{S} \phi\Big(Y_{t_{n+1}}; z_s + \mu\big(z_s, t_n + (M-1)h; \theta\big)h, V\big(z_s, t_n + (M-1)h; \theta\big)h\Big), \end{aligned} \tag{8}$$

where the z_s, for $s = 1, 2, \ldots S$, represent independent realizations of an M-step Euler discretization after $M-1$ iterations, $\widehat{Y}_{t_n+(M-1)h}$. Each discretization starts at $\widehat{Y}_{t_n} = Y_{t_n}$.

The Strong Law of Large Numbers guarantees that the approximation $\widehat{q}_{M,S}$ converges to the transition density q_M of the Euler discretization as $S \to \infty$. Since the transition density of the Euler discretization converges to the transition density p of the continuous-time process as $M \to \infty$, the approximation $\widehat{q}_{M,S}$ also converges to the transition density of the continuous-time process as $S \to \infty$ and $M \to \infty$. Lemmas 2 and 3 in Appendix A formally establish the consistency and asymptotic distribution of the approximation, respectively.

marginal distribution of a scalar summary statistic x (in our case the discrete-time transition probability) of a multivariate random variable y (in our case the continuous sample path) with density $p(y)$ and that the conditional density $p(x|y_{-x})$ is known in closed form (in our case Brownian increments), where y_{-x} denotes all elements of y except for x. The marginal distribution of x can then be obtained by averaging $p(x|y_{-x})$ over simulated y_{-x}, or $\widehat{p}(x) = 1/S \sum_{s=1}^{S} p(x|y_{-x}^s)$. We thank the referee for this observation.

[13]Footnote 11 states that the estimator extends to any discretization scheme with closed-form transitions. The simulation approach described here shows that we actually only need the *final* step of the scheme to have closed-form transitions. One could therefore attempt to capture the benefits of a higher-order discretization scheme with unknown transitions, such as the Platen-Wagner scheme, by iterating on this scheme $M-1$ times and using for the final transition the Euler or Milstein scheme. We thank the referee for this suggestion.

2.2.2 Approximating the Initial Unconditional Density

If the diffusions are stationary and ergodic, the unconditional density can also be evaluated with simulations. Under the assumption of stationarity and ergodicity, the unconditional density does not depend on time, or $p(y, t_0; \theta) = p(y; \theta)$ with $p(y; \theta) = \lim_{t\to\infty} p(y, t|x, 0; \theta)$. This implies that we can start with any initial x and use the Euler discretization to simulate a long continuous sample path of the diffusion. Then, we can approximate the unconditional probability of $y = Y_0$ from the simulated data using standard density estimation tools.

If the diffusions are non-stationary, we need to assume a deterministic Y_0. Fortunately, this assumption has a negligible effect on the likelihood function for sufficiently large samples. In fact, even if the diffusions are stationary and ergodic, estimates that assume Y_0 to be deterministic are often almost identical to estimates that allow Y_0 to be stochastic.

2.2.3 Maximum Likelihood Estimation

Given the above approximations of the transition densities and of the initial unconditional density, we construct a consistent approximation of the likelihood function $\mathcal{L}_N(\theta)$. We define the simulated maximum likelihood estimator $\widehat{\theta}_{N,M,S}$ as the parameters that maximize:[14]

$$\ln \widehat{\mathcal{L}}_{N,M,S}(\theta) = \ln \widehat{q}_{M,S}\big(Y_0, t_0; \theta\big) + \sum_{n=0}^{N-1} \ln \widehat{q}_{M,S}\big(Y_{n+1}, t_{n+1} \big| Y_n, t_n; \theta\big). \tag{9}$$

Since the approximations of the unconditional density and of the transition densities converge to their true counterparts, it follows that this approximate log likelihood function converges to the true log likelihood function [Lemma 4]. Furthermore, convergence occurs for all parameter values $\theta \in \Theta$, which means that the parameters that maximize the approximate log likelihood function converge to the parameters that maximize the true log likelihood function [Lemma 5]. Therefore, as long as the maximum likelihood estimator converges to the true parameter vector θ_0, so does our simulated likelihood estimator [Lemma 6].

To numerically maximize the log likelihood function, we must evaluate it repeatedly for different parameter values. As we vary the parameters, we use the same random numbers ϵ_t in the Euler discretization to generate the draws z_s from $f(z)$. This yields approximate transition densities that are smooth functions of the parameters. As a result, the objective function is continuous and twice differentiable in θ. Not only does this help in the numerical

[14]Since the log likelihood function is constructed from approximations of the transition densities, rather than from approximations of the log transition densities, the non-linearity of the log transformation induces a bias of order $1/S$ in both the log likelihood function and the resulting SML estimator. Gouriéroux and Monfort (1993) suggest a first-order correction for this bias (in a slightly different context).

optimization, but it is also required for our proofs of the asymptotics.

2.3 Asymptotics

The asymptotics of the SML method are summarized in two theorems. The first theorem establishes the consistency of the estimator. The second theorem presents its asymptotic distribution. Proofs of the theorems and supporting lemmas are in Appendix A.

Theorem 1: *Given Assumptions 1 through 5, as $M \to \infty$ and $S \to \infty$, with $S^{1/2}/M \to 0$, the estimator $\widehat{\theta}_{N,M,S}$ converges to the maximum likelihood estimator $\widehat{\theta}_N$, which in turn converges to the true parameter vector θ_0 as $N \to \infty$.*

This theorem summarizes Lemmas 1 through 6 in Appendix A. Lemmas 1 through 3 show that the approximate transition densities converge to their true counterparts. Lemma 4 does the same for the approximate log likelihood function. Lemma 5 then shows that the estimator converges to the maximum likelihood estimator, which according to Lemma 6, is a consistent estimator of the true parameter vector.

To establish the asymptotic distribution, we require one more assumption:

Assumption 6: The gradient $\partial p(Y_{t_{n+1}}, t_{n+1}|Y_{t_n}, t_n; \theta)/\partial\theta$ converges as $N \to \infty$, or, if it diverges, it does so at a rate slower than the rate of convergence of $I_N(\theta_0)^{-1/2}$ to zero.

A sufficient condition for this so-called asymptotic negligibility assumption is that the conditional densities p are strictly positive and that the derivatives $\partial p/\partial\theta$ are bounded.

Lemma 7 in Appendix A and the consistency of our estimator from Theorem 1 imply:

Theorem 2: *Given Assumptions 1 through 6, as $N \to \infty$, $M \to \infty$, and $S \to \infty$, with $S^{1/2}/M \to 0$ and $N/S^{1/4} \to 0$, the asymptotic distribution of the estimator $\widehat{\theta}_{N,M,S}$ is:*

$$I_N(\theta_0)^{1/2}\left[\widehat{\theta}_{N,M,S} - \theta_0\right] \sim \mathrm{N}\big(0, 1\big), \tag{10}$$

where the information matrix $I_N(\theta)$ is defined in Assumption 5.

2.4 Some Practical Considerations

The quality of the estimator depends on three quantities: the sample size N, the number of discretization steps M, and the simulation size S. While the data determines the sample size, the econometrician controls the other two parameters. Increasing M, S, or both, improves the approximation of the transition densities and thus results in an estimator that behaves

more like the exact maximum likelihood estimator. However, at the same time, it increases the computing time required to evaluate the approximate likelihood function.

Simulation studies by Honoré (1997) and Durham and Gallant (2000) suggest that for fairly persistent daily or weekly data an M of five to ten is sufficient to capture the shape of the transition densities for reasonably well-behaved univariate and multivariate diffusions. Regarding the choice of S, our experience with the estimator suggests that even for a four dimensional diffusion, 2500-5000 simulations are sufficient. However, in practice, we always perform a final round of optimizations, in which we double both S and M. The resulting changes in the estimates and the objective function are typically negligible.

Variance reduction techniques are an effective way to enhance the quality of the estimator for a given number of simulations. In practice, we always use the method of antithetic variates, where for every z_s generated by the sequence of Gaussian innovations $\{\epsilon_{t_n+(m+1)h}\}_{m=0}^{M-1}$, we also include the z_s^- generated by $\{-\epsilon_{t_n+(m+1)h}\}_{m=0}^{M-1}$.[15]

As with any simulation-based econometric method, computational feasibility is an important issue. Given modern computing power, SML estimation is practical. Efficiently programmed (in Fortran or C), a single evaluation of the likelihood function and the gradient (see below) for a $K = 4$ dimensional diffusion with $L = 15$ parameters, $N = 500$ observations, $M = 10$ discretization steps, and $S = 5000$ simulations, takes less than tree minutes on a Pentium III 800 MHz personal computer. The whole simulated likelihood estimation takes less than a day (with approximately 25 iterations per parameter).[16] Furthermore, the required computing time grows only linearly as we increase any of the quantities $\{K, N, M, S\}$ to estimate a more elaborate model or to improve the precision of the estimator.

Optimizing the simulated log likelihood function is computationally feasible, even with 15 parameters, because both the gradient and the Hessian can be computed explicitly. In particular, for the gradient we need the derivatives:

$$\frac{\partial \widehat{q}_S}{\partial \theta} = \frac{1}{S}\sum_{s=1}^{S} \frac{\partial \phi}{\partial \theta} + \frac{\partial \phi}{\partial z_s}\frac{\partial z_s}{\partial \theta}. \tag{11}$$

Given $\phi = \phi\big(Y_{n+1}; z_s + \mu(z_s, t_n + (M-1)h; \theta)h, V(z_s, t_n + (M-1)h; \theta)h\big)$, we can analytically compute the derivatives $\partial\phi/\partial\theta$ and $\partial\phi/\partial z_s$. The term $\partial z_s/\partial\theta$, which stands for a realization

[15] Kloeden and Platen (1992) discuss the use of antithetic variates and other variance reduction techniques.

[16] For the maximization we use NPSOL 5.0, which is a software package for efficiently solving constrained high-dimensional optimization problems. It employs a dense successive quadratic programming (SQP) algorithm and is designed particularly for nonlinear problems whose functions and gradients are expensive to evaluate. The algorithm is described in detail by Gill, Murray, Saunders, and Wright (1998) and the code is distributed in Fortran with a Matlab interface by Stanford Business Software Inc.

of $\partial \widehat{Y}_{t_n+(M-1)h}/\partial\theta$, can be obtained recursively by differentiating the Euler discretization:

$$\frac{\partial \widehat{Y}_{t_n+(m+1)h}}{\partial\theta} = \frac{\partial \widehat{Y}_{t_n+mh}}{\partial\theta} + \left[\mu_\theta\big(\widehat{Y}_{t_n+mh}, t_n+mh;\theta\big) + \mu_Y\big(\widehat{Y}_{t_n+mh}, t_n+mh;\theta\big)\frac{\partial \widehat{Y}_{t_n+mh}}{\partial\theta}\right] h + \left[\Sigma_\theta\big(\widehat{Y}_{t_n+mh}, t_n+mh;\theta\big) + \Sigma_Y\big(\widehat{Y}_{t_n+mh}, t_n+mh;\theta\big)\frac{\partial \widehat{Y}_{t_n+mh}}{\partial\theta}\right] \sqrt{h}\,\epsilon_{t_n+(m+1)h}, \tag{12}$$

where the subscripts denote the corresponding partial derivations. The recursion starts with the initial condition $\partial \widehat{Y}_{t_n}/\partial\theta = 0$. We can use the same reasoning to explicitly compute the second derivatives for the Hessian.

3 Exchange Rate Dynamics in Incomplete Markets

In this section, we present a new model of the joint dynamics of interest rates in two countries and of the exchange rate between the two currencies.[17] Our model is both arbitrage-free and consistent with equilibrium pricing in incomplete markets. The key insight of our model is that when markets are incomplete the volatility of the exchange rate is not uniquely determined by the domestic and foreign stochastic discount factors. Market incompleteness causes the exchange rate to exhibit "excess volatility." We capture this excess volatility in our model through a stochastic process for the degree of market incompleteness.

3.1 General Setup

Consider a world with two countries, a home country and a foreign country, each with its own currency. Quantities denominated in the foreign currency are superscripted by $\star$.

We postulate the existence of a stochastic discount factor (SDF) in the home country that prices all domestic assets. We denote this SDF by M_t and refer to it alternatively as the pricing kernel or the state price density.[18] The absence of arbitrage in financial markets is equivalent to the existence of a strictly positive SDF. It also requires that the product M_tV_t is a martingale, where V_t is the value process of any admissible self-financing trading

[17] The model is similar in spirit to the specifications of Nielsen and Saá-Requejo (1993), Saá-Requejo (1993), Ahn (1995), Ahn and Gao (1999), Bansal (1997), and Backus, Foresi, and Telmer (1998).

[18] In an exchange economy, the SDF can be interpreted as the representative agent's nominal, intertemporal, marginal rate of substitution of consumption. See Duffie (1996) or Cochrane (2001) for details on SDFs.

strategy. This implies the fundamental pricing equation:

$$V_t = \mathrm{E}_t\left[\frac{M_s}{M_t}V_s\right] = \frac{1}{M_t}\mathrm{E}_t\left[M_s\right]\mathrm{E}_t\left[V_s\right] + \frac{1}{M_t}\mathrm{Cov}_t\left[M_s, V_s\right], \tag{13}$$

where $s > t$ is some future date. Intuitively, the ratio M_s/M_t discounts the future payoff V_s at a rate that adjusts for the risk associated with the trading strategy.

When the financial markets are complete, meaning that the space of all payoffs is spanned by trading strategies in the domestic assets, the SDF is unique. If market are incomplete, however, there exists an infinite number of SDFs that price all domestic assets.[19]

We assume the following joint dynamics of the domestic SDF:[20]

$$\frac{dM_t}{M_t} = -r_t dt - \phi_t dW_t - \psi_t dZ_t \tag{14}$$

and of the domestic instantaneous interest rate:

$$dr_t = \mu_t dt + \sigma_t dW_t, \tag{15}$$

where ϕ_t, ψ_t, μ_t, and σ_t are stochastic processes that satisfy the usual conditions for the diffusions to be well defined [Karatzas and Shreve (1988)]. When markets are complete, the SDF in equation (14) is unique. Otherwise, we assume it is the minimum-variance SDF.[21]

There are two sources of risk that are priced in the domestic economy, corresponding to the Brownian motions W_t and Z_t. Without loss of generality, we assume that these risks are orthogonal. The instantaneous expected return of any trading strategy V_t is then:

$$\begin{aligned}\mathrm{E}_t\left[\frac{dV_t}{V_t}\right] &= r_t dt - \mathrm{Cov}_t\left[\frac{dV_t}{V_t}, \frac{dM_t}{M_t}\right] \\ &= r_t dt + \phi_t \mathrm{Cov}_t\left[\frac{dV_t}{V_t}, dW_t\right] + \psi_t \mathrm{Cov}_t\left[\frac{dV_t}{V_t}, dZ_t\right].\end{aligned} \tag{16}$$

From this expression, we interpret ϕ_t and ψ_t as the market prices of risk (or instantaneous

[19] If M_t prices V_t such that $V_t = \mathrm{E}_t[(M_s/M_t)V_s]$, then $\hat{M}_t = M_t U_t$, where U_t is a martingale orthogonal to M_t and V_t, also prices V_t because $\mathrm{E}_t[(\hat{M}_s/\hat{M}_t)V_s] = \mathrm{E}_t[(M_s/M_t)V_s]\mathrm{E}_t[U_s/U_t] = \mathrm{E}_t[(M_s/M_t)V_s] = V_t$.

[20] The drift of the SDF guarantees that M_t prices a riskless bank account that grows at the instantaneous interest rate r_t. With continuous compounding, the value at time t of an initial deposit of B_0 currency units in an interest bearing account is $B_t = B_0 \exp\{\int_0^t r_s ds\}$. The absence of arbitrage requires that $M_t B_t$ is a martingale, or $\mathrm{E}_t\left[d(M_t B_t)/(M_t B_t)\right] = 0$, which requires that the drift of the SDF is $\mathrm{E}_t\left[dM_t/M_t\right] = -r_t dt$.

[21] The minimum-variance SDF plays a special role because in that case $\phi_t^2 + \psi_t^2$ corresponds to the maximum Sharpe ratio obtainable by trading in the assets priced by the SDFs [Hansen and Jagannathan (1991)].

Sharpe ratio) associated with the Brownian motions W_t and Z_t, respectively. They are the instantaneous excess returns to the strategy V_t for covarying with the systematic risks in the domestic economy. Since W_t governs the dynamics of the domestic interest rate, the process ϕ_t represents the market price of interest rate risk. Analogously, we refer to the process ψ_t as the market price of risk that is orthogonal to interest rate risk.

Symmetric to the domestic economy, we postulate the existence of a foreign SDF that prices all foreign assets and model either the unique or minimum-variance SDF (depending on whether markets are complete) jointly with the foreign instantaneous interest rate as:

$$\frac{dM_t^\star}{M_t^\star} = -r_t^\star dt - \phi_t^\star dW_t^\star - \psi_t^\star dZ_t^\star \tag{17}$$

and

$$dr_t^\star = \mu_t^\star dt + \sigma_t^\star dW_t^\star, \tag{18}$$

where ${\phi_t}^\star$, ${\psi_t}^\star$, ${\mu_t}^\star$, and ${\sigma_t}^\star$ are again stochastic processes such that the diffusions are well defined. We assume that the foreign Brownian motions ${W_t}^\star$ and ${Z_t}^\star$ are correlated with the domestic Brownian motions W_t and Z_t, respectively, but are uncorrelated with each other. We can then interpret ${\phi_t}^\star$ as the market prices of foreign interest rate risk and ${\psi_t}^\star$ as the market price of all foreign risks that are orthogonal to interest rate risk.

We define the exchange rate E_t as the number of domestic currency units required to purchase one unit of foreign currency. The dynamics of this exchange rate are:

$$\frac{dE_t}{E_t} = \kappa_t dt + v_t dX_t, \tag{19}$$

where the drift κ_t, volatility v_t, and Brownian motion X_t are to be determined endogenously. For symmetry of the exchange rate from the domestic and foreign perspectives, we prefer working with the log exchange rate. By Itô's lemma, the dynamics of $e_t = \ln E_t$ are:

$$de_t = \left(\kappa_t - \frac{1}{2}{v_t}^2\right)dt + v_t dX_t, \tag{20}$$

which differs from the dynamics of E_t only by a Jensen's inequality drift correction.

The absence of arbitrage uniquely determines the drift κ_t of the exchange rate. Consider the following trading strategy of a domestic investor: buy one unit of foreign currency, deposit it with the foreign bank for one instant, and then convert the proceeds back into the home currency. The profit from this strategy consists of "capital gains" from the appreciation

or depreciation of the currency and "dividends" from the foreign interest. In units of the domestic currency, the value of this strategy follows the process:

$$\frac{dE_tB_t^\star}{E_tB_t^\star} = \left(\kappa_t + r_t^\star\right)dt + v_t dX_t. \tag{21}$$

The absence of arbitrage across the two countries requires that $M_tE_tB_t^\star$ is a martingale, or that $\mathrm{E}_t\left[d(M_tE_tB_t^\star)/(M_tE_tB_t^\star)\right] = 0$, which, in turn, implies that:

$$\kappa_t = \left(r_t - r_t^\star\right) + \left(\rho_{wx}\phi_t + \rho_{xz}\psi_t\right)v_t, \tag{22}$$

where ρ_{wx} and ρ_{xz} denote the instantaneous correlations (covariances) between the Brownian motions W_t and X_t and the Brownian motions X_t and Z_t, respectively.

The first term in the drift of the exchange rate is the usual interest rate differential. The second term is the currency risk premium. It is the excess return that domestic investors require to deposit money in the foreign bank account. This currency risk premium is made up of two components. The first component is $\sigma_{we}\phi_t \equiv \phi_t \operatorname{Cov}_t[dE_t/E_t, dW_t]$ and compensates domestic investors for the covariance between the exchange rate and domestic interest rate risk. The second component is $\sigma_{we}\psi_t \equiv \psi_t \operatorname{Cov}_t[dE_t/E_t, dZ_t]$ and compensates for the covariance between the exchange rate and domestic risk orthogonal to interest rate risk. We refer to this component as the risk premium for pure currency risk.

To determine the volatility v_t of the exchange rate, we note that any foreign security with price process $V_t^\star$ must be valued correctly by both the foreign and the domestic SDFs:

$$\mathrm{E}_t\left[\frac{M_s^\star}{M_t^\star}V_s^\star\right] = \mathrm{E}_t\left[\frac{M_s}{M_t}\frac{E_s}{E_t}V_s^\star\right]. \tag{23}$$

This equation is trivially satisfied if the foreign SDF is defined as $M_t^\star = M_tE_t$. Furthermore, if markets are complete, this definition of the foreign SDF is unique, which implies that one of the three quantities M_t, $M_t^\star$, and E_t is redundant and can be inferred from the other two.[22] Therefore, we can obtain an explicit expression for the exchange rate innovations $v_t\,dX_t$ by applying Itô's lemma to $e_t = \ln M_t^\star - \ln M_t$.

The key insight of our model is that if markets are incomplete the exchange rate volatility is not fully determined by the two minimum-variance SDFs. Depending on which set of markets is incomplete, either or both SDFs that satisfy $\hat{M}_t^\star = \hat{M}_tE_t$ can contain additional sources of unpriced uncertainty (recall footnote 19), such that $\hat{M}_t = M_tU_t$ and $\hat{M}_t^\star = M_t^\star U_t^\star$, where U_t and $U_t^\star$ are martingales that are orthogonal to the minimum-variance SDFs, all

[22] This redundancy is explained carefully by Backus, Foresi, and Telmer (1998).

domestic and foreign security price processes, and each other. To maintain the symmetry of our model, we therefore assume, without further loss of generality, the following relationship between the exchange rate and the two minimum-variance SDFs:

$$E_t = \frac{M_t^\star}{M_t} O_t, \tag{24}$$

where O_t is a martingale that is orthogonal to M_t, $M_t^\star$, and all domestic and foreign assets. It is easy to verify that with this exchange rate both SDFs assign the same value to any foreign security price process $V_t^\star$. This means that although markets are incomplete, the law of one prices holds across the two countries.

In our diffusion framework, we let:

$$\frac{dO_t}{O_t} = \epsilon_t dU_t, \tag{25}$$

where the Brownian motion U_t is uncorrelated with the Brownian motions W_t, $W_t^\star$, Z_t and $Z_t^\star$. By Itô's lemma, the dynamics of the log exchange rate are then:[23]

$$\begin{aligned} de_t = \Big[\big(r_t - r_t^\star\big) + \frac{1}{2}\big({\phi_t}^2 - \phi_t^{\star 2}\big) + \frac{1}{2}\big({\psi_t}^2 - \psi_t^{\star 2}\big) - \frac{1}{2}{\epsilon_t}^2\Big] dt \\ + \phi_t dW_t - \phi_t^\star dW_t^\star + \psi_t dZ_t - \psi_t^\star dZ_t^\star + \epsilon_t dU_t, \end{aligned} \tag{26}$$

which we can rewrite in the form of equation (20) as:

$$de_t = \Big[\big(r_t - r_t^\star\big) + \big(\rho_{wx}\phi_t + \rho_{zx}\psi_t\big)v_t - \frac{1}{2}v_t^2\Big] dt + v_t dX_t. \tag{27}$$

with

$$v_t dX_t = \phi_t dW_t - \phi_t^\star dW_t^\star + \psi_t dZ_t - \psi_t^\star dZ_t^\star + \epsilon_t dU_t \tag{28}$$

and

$$v_t^2 = \big({\phi_t}^2 + \phi_t^{\star 2} - 2\rho_{ww^\star}\phi_t\phi_t^\star\big) + \big({\psi_t}^2 + \psi_t^{\star 2} - 2\rho_{zz^\star}\psi_t\psi_t^\star\big) + {\epsilon_t}^2, \tag{29}$$

where from equation (28) we have $\rho_{wx} = (\phi_t - \rho_{ww^\star}\phi_t^\star)/v_t$ and $\rho_{xz} = (\psi_t - \rho_{zz^\star}\psi_t^\star)/v_t$.

Equation (29) makes more precise the effect of market incompleteness. With incomplete markets, the variance of the exchange rate is inflated by an amount ${\epsilon_t}^2$ relative to the variance

[23] Equation (26) illustrates nicely the symmetry of our model. This symmetry is not as transparent in the mathematically equivalent formulation (20) because of the convoluted functional forms of ρ_{wx}, ρ_{zx}, and v_t.

with complete markets.[24] The degree of "excess volatility" of the exchange rate depends on the amount of uncertainty due to market incompleteness. We therefore interpret ϵ_t, which determines the volatility of O_t, as a measure of both the degree of market incompleteness and the degree of excess volatility of the exchange rate.

The final building block of our model is the following stochastic process for the degree of market incompleteness:

$$d{\epsilon_t}^2 = \alpha_t dt + \beta_t dY_t \tag{30}$$

where α_t and β_t are stochastic processes such that the diffusion is well defined, and the Brownian motion Y_t may be correlated with the other Brownian motions in the model.

Assuming that the market prices of risk ϕ_t, $\phi_t^\star$, ψ_t, and $\psi_t^\star$ depend only on the variables r_t, $r_t^\star$, e_t, and ϵ_t, equation (29) defines the exchange rate volatility as a function of the same variables. We can then use Itô's lemma to obtain the dynamics of v_t in the form:

$$\begin{aligned} d{v_t}^2 = \alpha(r_t, r_t^\star, e_t, v_t)dt + \beta_w(r_t, r_t^\star, e_t, v_t)dW_t + \beta_{w^\star}(r_t, r_t^\star, e_t, v_t)dW_t^\star \\ + \beta_x(r_t, r_t^\star, e_t, v_t)dX_t + \beta_y(r_t, r_t^\star, e_t, v_t)dY_t, \end{aligned} \tag{31}$$

where we use equation (29) again to replace ϵ_t with r_t, $r_t^\star$, e_t, and v_t.

In summary, our model specifies the dynamics of four quantities: the domestic and foreign interest rates [equations (15) and (18)], the log exchange rate [equations (26) and (29)], and the degree of market incompleteness [equation (30)] *or equivalently* the volatility of the log exchange rate [equation (31)]. These four quantities are governed by four Brownian motions with the following correlation structure:

$$\mathrm{Corr}_t \begin{bmatrix} dW_t \\ dW_t^\star \\ dX_t \\ dY_t \end{bmatrix} = \begin{bmatrix} 1 & & & \\ \rho_{ww^\star} & 1 & & \\ \frac{\phi_t - \rho_{ww^\star}\phi_t^\star}{v_t} & \frac{\rho_{ww^\star}\phi_t - \phi_t^\star}{v_t} & 1 & \\ \rho_{wy} & \rho_{w^\star y} & \rho_{xy} & 1 \end{bmatrix}. \tag{32}$$

[24]Notice that market incompleteness does not effect the drift of the exchange rate and only enters the drift of the log exchange rate through the Jensen's inequality correction. Therefore, whether markets are complete is irrelevant for studying issues that involve only the drift, such as the "forward premium puzzle". We thank the referee for this observation.

3.2 Parsimonious Parameterization

We need to parameterize the model to make it suitable for estimation. As usual, our objective in choosing a parameterization is two-fold. We want a model that is parsimonious and, at the same time, flexible enough to capture the stylized features of the data.

Motivated by an extensive literature on the term structure of interest rates, we assume that the domestic and foreign interest rates follow correlated square-root processes:

$$\begin{aligned} dr_t &= \lambda(\theta - r_t)dt + \sigma\sqrt{r_t}\, dW_t \\ dr_t^\star &= \lambda^\star(\theta^\star - r_t^\star)dt + \sigma^\star\sqrt{r_t^\star}\, dW_t^\star \end{aligned} \tag{33}$$

The two interest rates are mean reverting to their unconditional means θ and $\theta^\star$ with mean reversion speeds λ and $\lambda^\star$ and conditional volatilities $\sigma\sqrt{r_t}$ and $\sigma^\star\sqrt{r_t^\star}$, respectively.

Like Cox, Ingersoll, and Ross (CIR, 1995), we assume that the market prices of interest rate risk are proportional to the square-root of the respective interest rate:

$$\phi_t = \phi\sqrt{r_t} \quad \text{and} \quad \phi_t^\star = \phi^\star\sqrt{r_t^\star}. \tag{34}$$

With this assumption, we can price domestic and foreign zero-coupon bonds in closed form, which we use to identify the parameters ϕ and $\phi^\star$ (see Section 4.2).

The theoretical literature offers little guidance on parameterizing the market prices of currency risk orthogonal to interest rate risk. We explore specifications with constant market prices $\psi_t = \psi$ and $\psi_t^\star = \psi^\star$, as well as with time-varying market prices:

$$\psi_t = \psi_0 + \psi_1(r_t - r_t^\star) + \psi_2 e_t + \psi_3 v_t \quad \text{and} \quad \psi_t^\star = \psi_0^\star + \psi_1^\star(r_t - r_t^\star) + \psi_2^\star e_t + \psi_3^\star v_t. \tag{35}$$

We specify the follow square-root process for the market incompleteness measure ϵ_t^2:

$$d\epsilon_t^2 = \left(-2\alpha\epsilon_t^2 + \beta^2\right)dt + 2\beta\left|\epsilon_t\right|dY_t, \tag{36}$$

where the drift and volatility specifications are motivated (via Itô's lemma) by an Orstein-Uhlenbeck process for ϵ_t that mean-reverts to zero (which represents market completeness) with a mean-reversion speed of α and a conditional volatility of β.

Even with this parsimonious parameterization, the implied volatility dynamics (31) are quite involved and the details are therefore left to Appendix B.

4 Empirical Results

We estimate the above parameterization of our model using the SML method and data for the US as the home country and the UK or Germany as the foreign country.

4.1 Data

4.1.1 Interest and Exchange Rates

We collect weekly observations of one-week Euro-currency interest rates as proxies for the instantaneous interest rates r in the US and $r^{\star}_{UK}$ in the UK or $r^{\star}_{DM}$ in Germany. We use one-week interest rates, instead of overnight rates, because of the market micro-structure anomalies documented in the overnight market [Hamilton (1996)]. In addition to the one-week interest rates, we also collect weekly observations of one-year Euro-currency yields y, $y^{\star}_{UK}$, and $y^{\star}_{DM}$, as well as the spot dollar-per-pound and dollar-per-mark exchange rates E_{UK} and E_{DM}, respectively. The interest and exchange rates are sampled every Tuesday from January 1990 through May 2000 (544 observations). The Euro-currency rates are taken from the Financial Times and the exchange rates are obtained from Morgan Stanley Capital International (MSCI), both provided by the Datastream database.

4.1.2 Volatility of the Exchange Rate

Traditionally, the empirical literature on stochastic volatility treats the volatility as a latent state variable. We explore an alternative approach. We use the implied volatility of an at-the-money option on the spot exchange rates with one week to maturity as a proxy for the instantaneous volatility of the exchange rates. This is analogous to using the one-week interest rate as proxy for the instantaneous interest rate.

The results of Ledoit and Santa-Clara (1998) formally justify this approach. They show that in a stochastic volatility model, the Black-Scholes implied volatility of an at-the-money option converges to the instantaneous volatility of the underlying asset as the option approaches expiration. Intuitively, one instant before the option expires, the effect of stochastic volatility on the option price is negligible. In that case, the Black-Scholes formula prices the option correctly and, as a result, its implied volatility corresponds to the instantaneous volatility of the underlying asset.[25] Since we cannot observe the implied

[25] Ledoit and Santa-Clara's result is not as trivial as this intuitive argument suggests. They prove that *only* the implied volatility of an at-the-money option converges to the instantaneous volatility. The limit of the implied volatilities of out-of- or in-the-money options is indeterminate.

volatility of an option one instant before it expires, we use the implied volatility of an at-the-money option with one week to expiration as a proxy for the instantaneous volatility.

We merge the weekly interest and exchange rates with over-the-counter (OTC) quotes of one-week at-the-money dollar-per-pound and dollar-per-mark exchange rate options. The data is provided by a major money-center banks that actively makes a market in OTC currency options. It comes in the form of Garman-Kohlhagen implied volatilities.[26] Relative to their exchange traded counterparts, OTC currency options are very liquid, especially when they are short-dated and at-the-money. Furthermore, OTC options are well-suited for our purpose because they are quoted for a fixed set of times-to-maturities and moneyness, as opposed to a fixed set of maturity dates and strike prices. This means that we always observe an at-the-money option with exactly one week to maturity. Finally, since OTC options are quoted as implied volatilities, rather than prices, the effect of non-synchronicities between the exchange rates, interest rates, and option quotes is minimal.[27]

4.1.3 Descriptive Statistics

Tables 1 and 2 and Figure 2 describe the data. Table 1 presents in Panel A summary statistics for the interest rates r, $r^{\star}_{UK}$, and $r^{\star}_{DM}$, for the yields y, $y^{\star}_{UK}$, and $y^{\star}_{DM}$, for the interest rate differential $r-r^{\star}_{UK}$ and $r-r^{\star}_{DM}$, for the log exchange rates e_{UK} and e_{DM}, and for the implied volatilities v_{UK} and v_{DM} of the exchange rates. It describes in Panel B the weekly differences Δr, $\Delta r^{\star}_{UK}$, $\Delta r^{\star}_{DM}$, Δe_{UK}, Δe_{DM}, Δv_{UK}, and Δv_{DM}. Table 2 shows selected cross-correlations of the data. Panel A of Figure 2 plots as solid lines the log exchange rates and as dashed lines the corresponding volatilities of the exchange rates. Panel B plots as solid and dashed lines the domestic and foreign interest rates, respectively.

Toward the end of 1992 the dollar-per-pound exchange rate, in Panel A of Figure 2, experienced a dramatic drop. At the time, the British pound faced a severe currency crisis that resulted in it leaving the Exchange Rate Mechanism (ERM) of the European Monetary System (EMS). On Tuesday, September 15, 1992, the Bank of England intervened by buying more than 15 billion pounds and raising interest rates twice, from about ten percent to 12 percent, and ultimately to 15 percent. However, the intervention was in vain and on the next day, "Black Wednesday," the exchange rate was allowed to move freely.

Since the Financial Times samples the exchange and Euro-currency interest rates in the morning, we do not observe the extreme swings in rates that occurred on September 15, 1992. Nevertheless, there are still unusually large swings in the exchange and interest rate series

[26] Garman and Kohlhagen (1983) modify the Black and Scholes (1973) formula to price currency options.
[27] For further details on the OTC currency options market, see Campa, Chang, and Reider (1998).

surrounding this period. From September 9 through October 27, the price of the British pound fell from two dollars to 1.5 dollars, the volatility of the exchange rate increased from 15 percent to 22 percent, while the UK interest rate was approximately ten percent. By December 15, the exchange rate was still around 1.5 dollars per pound, but its volatility had returned to 15 percent and the interest rate had declined to 7.1 percent.

It is difficult to judge the impact of the ERM crisis on the estimation results for the British pound. However, it is reassuring that this time period was not that unusual. Panel A of Figure 2 shows other periods during which the data series experienced sizable shocks. Unfortunately, the limited availability of OTC options data does not allow us to further explore whether September 1992 represents a structural break in the data.

Finally, notice that the annualized standard deviation of Δe of 10.1 percent for the UK or 10.6 percent for Germany virtually matches the average implied volatility of 10.3 or 10.7 percent, respectively. Therefore, the one-week implied volatility is (at least on average) a good proxy for the instantaneous volatility of the exchange rate.

4.2 Identifying the Market Prices of Risk

Market prices of risk are notoriously difficult to estimate and, to make matters worse, they are not well identified in our model. In particular, the market prices of interest rate risk ϕ_t and $\phi_t^\star$ appear in the drift of the log exchange rate as $\sigma_{we}\phi_t = \phi^2\, r_t - \rho_{ww^\star}\, \phi\phi^\star \sqrt{r_t r_t^\star}$ and in the variance as $\sigma_{we}\phi_t + \sigma_{w^\star e}\phi_t^\star = \phi^2\, r_t + \phi^{\star 2}\, r_t^\star - 2\, \rho_{ww^\star}\, \phi\phi^\star \sqrt{r_t r_t^\star}$. Since both interest rates are highly persistent (see Table 1), the two terms $\sigma_{we}\phi_t$ and $\sigma_{we}\phi_t + \sigma_{w^\star e}\, \phi_t^\star$ are virtually constant and estimating ϕ and $\phi^\star$ from the exchange rate dynamics alone is difficult.

This problem is even worse for the market prices of currency risk orthogonal to interest rate risk, ψ_t and $\psi_t^\star$, which appear in the drift of the exchange rate as $\sigma_{ze}\psi_t = {\psi_t}^2 - \rho_{zz^\star}\psi_t\psi_t^\star$ and in the variance as $\sigma_{ze}\psi_t + \sigma_{z^\star e}\psi_t^\star = {\psi_t}^2 + \psi_t^{\star 2} - 2\, \rho_{zz^\star}\psi_t\psi_t^\star$. Since $\rho_{zz^\star}$ is a coefficient that, in contrast to $\rho_{ww^\star}$, cannot be directly estimated from observables, it follows that when ψ_t and $\psi_t^\star$ are constant, we have two equation in three unknowns. In that case, we can only identify $\sigma_{ze}\psi$ and $\sigma_{z^\star e}\psi^\star$ from the exchange rate dynamics. When ψ_t and $\psi_t^\star$ are time-varying as a function of observables, we can separately identify the coefficients of the two functions, but, depending on the persistence of ψ_t and $\psi_t^\star$, the estimation may then be subject to the same problems as that of the market prices of interest rate risk.[28]

[28] Suppose $\psi_t = \psi_0 + \psi_1 x_t$ and $\psi_t^\star = \psi_0^\star + \psi_1^\star x_t$, for some regressor x_t. Then, $\sigma_{ze}\psi_t = a_0 + a_1 x_t + a_2 {x_t}^2$ and $\sigma_{ze}\psi_t + \sigma_{z^\star e}\psi_t^\star = b_0 + b_1 x_t + b_2 {x_t}^2$, where we can use the six reduced-form coefficients: $a_0 = \psi_0 - \rho_{zz^\star}\psi_0\psi_0^\star$, $a_1 = \psi_1 - \rho_{zz^\star}\psi_0\psi_1^\star - \rho_{zz^\star}\psi_1\psi_0^\star$, $a_2 = -\rho_{zz^\star}\psi_1\psi_1^\star$, $b_0 = a_0 + \psi_0^\star - \rho_{zz^\star}\psi_0\psi_0^\star$, $b_1 = a_1 + \psi_1^\star - \rho_{zz^\star}\psi_0\psi_1^\star - \rho_{zz^\star}\psi_1\psi_0^\star$, and $b_2 = a_2 - \rho_{zz^\star}\psi_1\psi_1^\star$ to solve for the five structural coefficients ψ_0, ψ_1, $\psi_0^\star$, $\psi_1^\star$, and $\rho_{zz^\star}$.

Finally, market incompleteness makes the estimation of the market prices of risk even more difficult because it introduces the excess variance term ${\epsilon_t}^2$ into the variance of the log exchange rate in equation (29). The problem is that subtracting a constant from either $\sigma_{we}\phi_t + \sigma_{w^\star e}\phi_t^\star$ or $\sigma_{ze}\psi_t + \sigma_{z^\star e}\psi_t^\star$ and adding this constant to ${\epsilon_t}^2$ results in the same variance of the exchange rate. In principle, the dynamics of ${\epsilon_t}^2$ in equation (36) resolve this problem. Although the variance of the exchange rate is the same, the log likelihood function of the whole model is not because ϵ_t is assumed to mean-revert to zero (with constant volatility). In practice, however, market incompleteness clearly adds to the identification problem.

We therefore take another route to estimate the market prices of risk. In particular, we estimate the market price of interest rate risk from the domestic and foreign term structures of default-free zero-coupon bonds. We use CIR's bond pricing formula, the parameters of the interest rate processes, and the interest rates to solve in every iteration of the SML estimation for the coefficient ϕ and $\phi^\star$ that best price, in a least-squares sense, the domestic and foreign one-year bonds (with yields y_{US} and $y_{UK}^\star$ or $y_{DM}^\star$).[29] We then evaluate the likelihood function conditional on these values of ϕ and $\phi^\star$.

When the market prices of currency risk orthogonal to interest rate risk are constant, in which case we can only identify the products $\sigma_{ze}\,\psi$ and $\sigma_{z^\star e}\psi^\star$, we estimate $\sigma_{ze}\psi$ by SML but choose $\sigma_{z^\star e}\psi^\star$ in every iteration of the SML estimation to minimize, in a least-squares sense, the degree of market incompleteness.[30] The intuition of this estimation scheme is that $\sigma_{ze}\psi$ is identified by the drift of the exchange rate while $\sigma_{z^\star e}\psi^\star$ is difficult to tell apart from ${\epsilon_t}^2$ in equation (29). By choosing $\sigma_{z^\star e}\psi^\star$ to minimize the sum of ${\epsilon_t}^2$, we impose a prior of sorts that markets are complete. When the market prices of risk are time-varying as a function of two or more observables, we we can identify the parameters of ψ_t and of the product $\rho_{zz^\star}\psi_t^\star$ from the drift of the exchange rate (by expanding $\sigma_{ze}\psi_t$ as a polynomial of the observables as in footnote 28). In that case, we estimate the correlation $\rho_{zz^\star}$ by minimizing the degree of market incompleteness or excess volatility.

To compute joint standard errors of the parameters estimated by SML and the market prices of risk estimated by least-squares, we stack the first-order conditions of the SML estimation (the scores) and the first-order conditions of the least-squares estimation into a single vector of moments. We then compute standard GMM [Hansen (1982) and Hansen and Singleton (1982)] standard errors for the parameter estimates, using the autocorrelation and heteroskedasticity adjusted covariance matrix estimator of Andrews (1991).

[29] For each country, we solve $\min_\phi \sum_{t=1}^T [y_t - y(r_t, \phi; \lambda, \theta, \sigma)]^2$, where $y(r_t, \phi; \lambda, \theta, \sigma)$ denotes the theoretical one-year yield as a function of r_t and ϕ, conditional on the parameters λ, θ, and σ of the interest rate process.

[30] We choose $\sigma_{z^\star e}\psi^\star$ to minimize $\sum_{t=1}^T {\epsilon_t}^2$, where ${\epsilon_t}^2 = v_t^2 - ({\phi_t}^2 + {\phi_t^\star}^2 - 2\rho_{ww^\star}\phi_t\phi_t^\star) - (\sigma_{ze}\psi + \sigma_{z^\star e}\psi^\star)$.

4.3 Estimation Results

Table 3 presents SML estimates of our model with constant market prices of currency risk orthogonal to interest rate risk ψ and $\psi^\star$. We use $M = 10$ Euler discretization steps and $S = 5000$ simulations (plus 5000 antithetic variates) to approximate the likelihood function. The likelihood maximization is performed using the NPSOL optimizer (see footnote 16).[31]

4.3.1 Interest Rate Dynamics and Market Prices of Interest Rate Risk

Our estimates of the square-root interest rate dynamics are in line with the results of Chen and Scott (1993), Gibbons and Ramaswamy (1993), Pearson and Sun (1994), and Geyer and Pichler (1995). The long-run means of the US, UK, and German rates are 5.3 (or 5.8), 7.4, and 6.4 percent, respectively. The US interest rates appear to mean-revert faster than the German rates but slower than the UK rates. The half-life of a shock to $r^\star_{US}$ is about 2.5 years, compared to more than 7.5 years for $r^\star_{DM}$ and less than 1.5 years for $r^\star_{UK}$. The UK rates are significantly more volatile than the US and German rates, with an annualized volatility of 1.2 to 2.2 percent, compared to 0.5 to 0.9 percent and 0.7 to 1.3 percent, respectively. Finally, the US and UK interest rates are fairly independent, with a correlation of only 0.06, while the US and German rates covary more, with a correlation of 0.21.

The linear drift and square-root volatility functions are admittedly simplistic.[32] However, estimating more elaborate interest rate dynamics is difficult because short-term interest rates are highly persistent (see Table 1). It requires samples much larger than ours to precisely estimate nonlinearities in the drift and volatility of interest rates. As a result, we only explore the square-root dynamics for the interest rates.

The advantage of the square-root dynamics is that we can use CIR's bond pricing formula to estimate the market prices of interest rate risk $\phi_t = \phi\sqrt{r_t}$ and $\phi^\star_t = \phi^\star\sqrt{r_t{}^\star}$ from the cross-sections of domestic and foreign bond yields (see Section 4.2). Our estimates imply an average market price of interest rate risk of -3.1 (or -2.8) percent in the US, -0.9 percent in Germany, and $+0.8$ percent in the UK. The relatively small magnitude of these market prices of risk reflects the fact that in our sample the average yield curve was virtually flat in

[31] We treat the initial observation as deterministic. However, we checked in lower-dimensional applications that the conditional (on the first observation) estimates are not very different from the unconditional ones.

[32] Aït-Sahalia (1996b), Conley, Hansen, Luttmer, and Scheinkman (1997), and Stanton (1997) document non-linearities in the drift of US interest rates. However, Pritsker (1998), Chapman and Pearson (2000) and Jones (2000b) argue that these results are due to the poor small sample properties of the estimators. There is also disagreement about the correct diffusion function. Nowman (1997) cannot reject that the volatility of UK interbank rates is independent of the level, while Ball and Torous (1999) present estimates that favor a square-root specification for four Euro-currency rates, UK interbank rates, and US Treasury bill rates.

the US and in Germany and was even slightly inverted in the UK (see Table 1).[33]

4.3.2 Exchange Rate Dynamics and Currency Risk Premium

Constant Market Prices of Currency Risk

The dynamics of the exchange rate are fully characterized by the instantaneous drift and volatility. The volatility v_t is observed while the drift κ_t consists of three terms: the interest rate differential $r_t - r_t^\star$, the interest rate risk premium $\sigma_{we}\phi_t = \phi_t{}^2 - \rho_{ww^\star}\phi_t\phi_t^\star$, and the risk premium for currency risk orthogonal to interest rate risk $\sigma_{ze}\psi_t = \psi_t{}^2 - \rho_{zz^\star}\psi_t\psi_t^\star$. Since the interest rate differential is observed and the interest rate risk premium is determined by the domestic and foreign market prices of interest rate risk, which in turn depend on the interest rate dynamics and the observed cross-sections of bond yields, the only free parameter in the exchange rate dynamics is the pure currency risk premium.[34] With constant market prices of pure currency risk, our estimates of the currency risk premium are an annualized 2.4 percent for the British pound and minus one percent for the Deutsche mark.

To make sense of these risk premia estimates, we can approximate (in discrete-time) the unconditional pure currency risk premium as $\mathrm{E}[\Delta e_t] - \mathrm{E}[r_t - r_t{}^\star] - \mathrm{E}[\phi_t{}^2 - \rho_{ww^\star}\phi_t\phi_t^\star] + \mathrm{E}[0.5\, v_t^2]$. Plugging in the sample moments from Table 1 and our estimates of the market prices of interest rate risk and of the correlation between the domestic and foreign interest rates, this approximation implies a pure currency risk premium of $-0.007+0.025-0.001+0.006=0.023$ for the British pound and of $-0.019+0.003-0.001+0.006=-0.011$ for the Deutsche mark. These approximations are within a few basis points of the corresponding SML estimates.

The implications of the estimates for currency investments depend on the nationality of the investor.[35] From the perspective of a US investor, the British pound offers a positive risk premium (interest rate risk and pure currency risk) and the Deutsche mark demands a negative premium throughout the whole the sample. For both exchange rates, the magnitude of the risk premium is the same as that of the pure currency risk premium $\sigma_{ze}\psi$ because the interest rate risk premium $\sigma_{we}\phi_t$ is economically negligible (less than 0.1 percent).

There are three hypotheses commonly considered for the currency risk premium. The

[33] To make sense of the magnitude of the market price of interest rate risk, we can in a one-factor model interpret $-\phi_t = -\phi\sqrt{r_t}$ as the instantaneous Sharpe ratio of default-free bonds [see equation (16)].

[34] The currency risk premium is not literally "free" because there must exist parameters $\{\psi, \psi^\star, \rho_{zz^\star}\}$ that generate $\psi^2 - \rho_{zz^\star}\psi\psi^\star$ and satisfy $v_t{}^2 \geq (\phi_t{}^2 + \phi_t^{\star 2} - 2\,\rho_{ww^\star}\phi_t\phi_t^\star) + (\psi^2 + \psi^{\star 2} - 2\rho_{zz^\star}\psi\psi^\star)$ for all t. This constraint can be binding, since the correlation $\rho_{zz^\star}$ must take on a value in $[-1, 1]$, but empirically it is slack.

[35] Currency investments are typically carried out with currency forward contracts, by buying the foreign currency forward and selling it in the spot market at maturity of the forward contract. In our model, the instantaneous log (dollars per foreign currency) forward price is $f_t = e_t + (r_t - r_t^\star)\,dt$, and the log return on this (zero investment) trading strategy is $e_{t+dt} - f_t = \left(\sigma_{we}\phi_t + \sigma_{ze}\psi_t - \frac{1}{2}v_t{}^2\right)dt + v_t dX_t$.

first hypothesis is that the risk premium is zero, which is typically phrased as uncovered interest rate parity or unbiasedness of the forward exchange rate as a predictor of the future spot exchange rate. The second hypothesis is that the premium cancels out the interest rate differential, causing the spot exchange rate to be a martingale. The third hypothesis is that the risk premium is time-varying in a way that does not cancel out the interest rate differential.[36] In our model, uncovered interest rate parity requires that $\sigma_{we}\phi_t = \sigma_{ze}\psi_t$. The estimates suggest otherwise, but the standard errors are so large, especially for the pure currency risk premium, that we cannot reject this hypothesis with a likelihood ratio test. For the British pound, the estimates are unconditionally consistent with the martingale hypothesis, since $\sigma_{ze}\psi \simeq \mathrm{E}[r_r{}^\star - r_t]$ and $\sigma_{we}\phi_t$ is negligible, but for the Deutsche mark the risk premium and average interest rate differential share the same sign. To address the third hypothesis, we need to allow the market prices of currency risk to be time-varying.

Time-Varying Market Prices of Currency Risk

Table 4 presents estimates of three different specifications (denoted models A, B, and C) of the market prices of pure currency risk ψ_t and $\psi_t^\star$. Model A is the constant market prices of risk case from Table 3, where we can only identify the terms $\psi^2 - \rho_{zz^\star}\psi\psi^\star$ and $\psi^{\star 2} - \rho_{zz^\star}\psi\psi^\star$. In model B, the market prices depend linearly on the interest rate differential $r_t - r_t{}^\star$ and the log exchange rate e_t.[37] In model C, the market prices depend also on the volatility of the exchange rate v_t.[38] For each model, the table shows only estimates of the coefficients entering the currency risk premium. The corresponding estimates of the other 14 parameters of the model are not shown to save space, but they are very similar to the estimates in Table 3. Finally, the table also describes the implied currency risk premium $\psi_t{}^2 - \rho_{zz^\star}\psi_t\psi_t^\star$.

From model B, there is some evidence (at the ten-percent level) that both market prices of currency risk depend on the log exchange rate. The point estimates are such that for both currencies the risk premium $\psi_t{}^2 - \rho_{zz^\star}\psi_t\psi_t^\star$ is positive throughout the whole sample and increases monotonically in the log exchange rate. For the British pound, the risk premium ranges from zero to 1.4 percent per year, with an average of 36 basis points. For the Deutsche

[36] Baillie and Bollerslev (1989,1990), Diebold (1988), Diebold and Nason (1990), and Westerfield (1977) find that exchange rate changes are virtually unpredictable, which supports the second hypothesis. Domowitz and Hakkio (1985), Fama (1984), Hansen and Hodrick (1980), Hodrick and Srivastava (1984,1986), and Hsieh (1984) find that exchange rates tend to vary in the opposite direction of the interest rate differential, which is evidence in favor of the third hypothesis and is referred to as the "forward premium puzzle". In the recent literature, only Roll and Yan (1998) find support for the uncovered interest rate parity hypothesis.

[37] We only consider cases in which the market prices depend on two or more variables because otherwise we cannot identify all parameters of ψ_t and $\rho_{zz^\star}\psi_t^\star$ from the drift of the log exchange rate (see Section 4.2).

[38] We treat models B and C separately in order to measure (in Section 4.3.3) the role the volatility plays in reducing the degree of market incompleteness. From equation (29) it is clear that if the market prices of currency risk depend on the exchange rate volatility, the ϵ_t^2 can in principle be reduced significantly.

mark, it ranges from zero to 65 basis points per year, with an average of 17 basis points.

Model C enhances the time-variation of the market prices of currency risk (and indirectly of the currency risk premium), both in terms of economic and statistical significance. The coefficients ψ_3 and $\psi_3^\star$ on the exchange rate volatility v_t are significant at the five-percent level for the British pound and at the ten-percent level for the Deutsche mark.

In the case of the British pound, the estimates imply an annualized currency risk premium that ranges from -4.1 to 2.9 percent, with an average of 32 basis points. Holding the volatility constant, the risk premium increases monotonically in the log exchange rate, and holding the log exchange rate constant, the risk premium decreases monotonically in the volatility of the exchange rate. Since for the British pound e_t and v_t are positively correlated in our sample (see Table 2), these marginal effects tend to partially offset each other. In the case of the Deutsche mark, the implied currency risk premium ranges from -56 basis point to 1.3 percent, with an average of 22 basis points. Most of this time-variation is still due to the positive relationship between the risk premium and the log exchange rate, and not due to the exchange rate volatility (notice the different magnitude of ψ_3 and $\psi_3^\star$).

To get a better sense for the time-variation in the currency risk premium and for the role this risk premium plays in the drift of the exchange rate, Figure 3 shows a decomposition of the drift into its three components. We plot the interest rate differential $r_t - r_t^\star$ as a dashed line, the interest rate risk premium $\sigma_{we}\phi_t = \phi_t^2 - \rho_{ww^\star}\phi_t\phi_t^\star$ as a dotted line, and the pure currency risk premium $\sigma_{ze}\psi_t = \psi_t^2 - \rho_{zz^\star}\psi_t\psi_t^\star$ as a solid line. Panels A and B correspond to the models B and C in Table 4, respectively.

Comparing the two panels, it is clear that model C produces substantially more time-variation in the currency risk premium. Consider first the British pound. The risk premium is large and positive during the first two years of the sample, turns large and negative after the ERM crisis, and then remains fairly steady around 50 basis points throughout the second half of the sample. Except for the first year after the ERM crisis, the premium tends to be positive and partially offsets the mostly negative interest rate differential. In terms of economic significance, the interest rate differential and the currency risk premium appear on equal footing, whereas, as we noted before, the interest rate risk premium is negligible.

In contrast, for the Deutsche mark the interest rate differential clearly dominates the drift of the exchange rate in both panels. Nevertheless, in Panel B the currency risk premium exhibits broadly the same pattern as the premium for the British pound. When the interest rate differential is large and negative, throughout the first half of the sample, the currency risk premium is positive. As the interest rate differential turns positive in the second half of the sample, the risk premium decreases and ultimately becomes negative.

To further measure the significance of the time-variation in the currency risk premium (beyond the statistical significance of the coefficients of ψ_t and $\psi_t^\star$), we compute the following incremental R^2 for the exchange rate equation of the diffusion model:

$$1 - \frac{\mathrm{Var}\Big[\Delta e_t - \Big((r_t - r_t^\star) + ({\phi_t}^2 - \rho_{ww^\star}\phi_t\phi_t^\star) + ({\psi_t}^2 - \rho_{zz^\star}\psi_t\psi_t^\star) - \frac{1}{2}\,{v_t}^2\Big)\Delta t\Big]}{\mathrm{Var}\Big[\Delta e_t - \Big((r_t - r_t^\star) + ({\phi_t}^2 - \rho_{ww^\star}\phi_t\phi_t^\star) + \ (\psi^2 - \rho_{zz^\star}\psi\psi^\star) \ - \frac{1}{2}\,{v_t}^2\Big)\Delta t\Big]}, \tag{37}$$

where the numerator is evaluated using the estimates of either model B or model C and the denominator corresponds to model A. For the British pound, the R^2 is 0.011 for model B and 0.029 for model C. For the Deutsche mark, the R^2 are 0.015 and 0.021. In the context of the predictability literature, where an R^2 of five percent is hailed as a great success, the time-variation in the currency risk premium is therefore not negligible, especially considering that the horizon in our study is only one week.

Returning to the three hypotheses about the currency risk premium, when the market prices of currency risk depend on the volatility of the exchange rate, we can decisively (at the five-percent level for the British pound and at the ten-percent level for the Deutsche mark) reject both the uncovered interest rate parity and martingale exchange rate hypotheses with likelihood ratio tests (not shown to save space). The reason for the rejections is two-fold. First, the volatility improves the fit of the exchange rate drift (which is obvious from the significance of ψ_3 and $\psi_3^\star$ and from the incremental R^2). Second, when the market prices of currency risk depend on the volatility of the exchange rate, the fluctuations of the ϵ_t [defined implicitly by equation (29)] are dampened and the Ornstein-Uhlenbeck process for the degree of excess volatility or market incompleteness fits the data better.

In contrast to our results, Baillie and Bollerslev (1989,1990), Bekaert and Hodrick (1993), and Domowitz and Hakkio (1985) find only weak or no support for the inclusion of the conditional exchange rate volatility in the exchange rate drift. The evidence presented in Table 4 is stronger for two reasons. First, we *observe* the volatility of the exchange rate, rather than filter it with potentially large error from past changes in the exchange rate using a GARCH model. Second, we impose an economic model that implies a specific functional form for the drift. In particular, our drift includes both v_t and ${v_t}^2$ through the polynomial form of $\sigma_{ze}\psi_t = {\psi_t}^2 - \rho_{zz^\star}\psi_t\psi_t^\star$. To illustrate the importance of including both terms, consider a GLS regression of $\Delta e_t - [(r_t - r_t^\star) - 0.5\,v_t^2]\Delta t$ on v_t, on v_t^2, and on both terms.[39] For the British pound, the adjusted R^2 of the regression is 0.009 with v_t, 0.013 with v_t^2, and 0.023 with both terms. For the Deutsche mark, the R^2 are 0.002, 0.001, and 0.016.

[39]We use a GLS not OLS regression to correct for the heteroskedasticity induced by the time-varying volatility v_t. We also checked that the regressions are not sensitive to the Jensen's inequality adjustment.

The estimates of ψ_t and $\psi_t^\star$ in Table 4 are subject to the constraint that ${\epsilon_t}^2 \geq 0$ for all t, or equivalently from equation (29) that:

$$v_t^2 \geq \left({\phi_t}^2 + \phi_t^{\star 2} - 2\rho_{ww^\star}\phi_t\phi_t^\star\right) + \left({\psi_t}^2 + \psi_t^{\star 2} - 2\rho_{zz^\star}\psi_t\psi_t^\star\right) \quad \text{for all } t. \tag{38}$$

This constraint is binding for models B and C, which can be detected from the fact that the average conditional currency risk premium deviates substantially from the unconditional risk premium (from model A). For the British pound, models B and C imply average currency risk premia of only 36 and 32 basis points, compared to an unconditional premium of 2.4 percent. Similarly for the Deutsche mark, models B and C generate average risk premia of 17 and 21 basis points, instead of an unconditional premium of minus one percent.

To further measure the impact of this constraint on our results, we plot in Figure 4 as solid line an unconstrained version of the currency risk premium of model C. Since we need equation (29) for the SML estimation, to infer ϵ_t^2 from the data, we cannot actually lift the constraint in the context of our diffusion model. We therefore compute the unconstrained risk premium through a GLS regression of $\Delta e_t - [(r_t - r_t^\star) + \sigma_{we}\phi_t - 0.5\, {v_t}^2]\Delta t$ on the variables $(r_t - r_t^\star)$, e_t, and v_t, their squares, and their cross-products (the terms in ${\psi_t}^2 - \rho_{zz^\star}\psi_t\psi_t^\star$). For comparison, we also plot in Figure 4 as a dashed line the corresponding constrained currency risk premium (from Figure 3) and as a dotted line the interest rate differential.

There is no question that lifting the constraint has a substantial effect on our estimates of the time-varying currency risk premia. Not only do the unconstrained risk premia match on average the unconditional premia from model A, but the time-variation in the risk premia also increases dramatically when we lift the constraint. For the British pound, for example, the unconstrained premium exceeds 20 percent in 1991, drops to nearly -20 percent after the ERM crisis, and remains fairly steady around five percent throughout the second half of the sample. Even more strikingly, when we lift the constraint the currency risk premium clearly dominates the interest rate differential in the drift of the exchange rate.

It is not obvious, however, that lifting the constraint leads to a statistically significant improvement of the estimated currency risk premium. For the British pound, the R^2 of the unconstrained regression is 0.031, compared to 0.029 for the constrained model C. For the Deutsche mark, the constrained and unconstrained R^2 are 0.019 and 0.021, respectively. At least in the context of the GLS regression, we cannot reject (at the five-percent level) the constrained risk premium implied by model C in favor of the unconstrained risk premium. Unfortunately, we cannot formally test the constraint in our diffusion model because we need to impose it to construct the likelihood function.

4.3.3 Market Incompleteness or Excess Volatility

The degree of market incompleteness or excess volatility implied by the observed volatility of the exchange rate v_t, the estimated market prices of interest rate risk ϕ_t and $\phi_t^\star$, and the estimated market prices of pure currency risk ψ_t and $\psi_t^\star$ is defined by equation (29) as:

$$|\epsilon_t| = \sqrt{{v_t}^2 - \left({\phi_t}^2 + {\phi_t^\star}^2 - 2\,\rho_{ww^\star}\phi_t\phi_t^\star\right) - \left({\psi_t}^2 + {\psi_t^\star}^2 - 2\rho_{zz^\star}\psi_t\psi_t^\star\right)}. \tag{39}$$

Table 4 describes the implied ϵ_t for models A, B, and C.

With constant market prices of currency risk (model A), the average level of excess volatility is surprisingly large. For the British pound, the average ϵ_t is an annualized 8.2 percent, which means that approximately 80 percent of the exchange rate volatility is not explained by the domestic and foreign pricing kernels (from Table 1, the average v_t is 10.3 percent). For the Deutsche mark, the average ϵ_t is 8.8 percent, which leaves 82 percent of the exchange rate volatility unexplained (the average v_t is 10.7 percent).

To understand this result recall that interest rates are very persistent. As a result, the term ${\phi_t}^2 + {\phi_t^\star}^2 - 2\rho_{ww^\star}\phi_t\phi_t^\star$ in equation (39) is nearly constant. When the market prices of currency risk are also constant, the excess volatility is then essentially $\epsilon_t = \sqrt{{v_t}^2 - \min\{v_t\}_{t=1}^T}$ because the constraint (38) must hold for all dates including the date on which the volatility is the lowest. Since v_t is quite volatile (see Figure 2), the average v_t is very different from the lowest v_t, which means that our model leaves a large fraction of the exchange rate volatility unexplained. It is important to realize that it is not the level but the *volatility* of exchange rate volatility that generates the excess volatility in model A.[40] If the exchange rate volatility was high but steady (so that the difference between the average and lowest v_t is small), the constant market prices of currency risk could explain a much larger portion of it.

Given that the exchange rate volatility is volatile and that the interest rate risk premium term ${\phi_t}^2 + {\phi_t^\star}^2 - 2\rho_{ww^\star}\phi_t\phi_t^\star$ is nearly constant, the only way to significantly reduce the level of excess volatility is through time-varying market prices of pure currency risk ψ_t and $\psi_t^\star$. However, time-variation in the market prices of currency risk alone is not enough. The term ${\psi_t}^2 + {\psi_t^\star}^2 - 2\rho_{zz^\star}\psi_t\psi_t^\star$ must be positively correlated with the variance ${v_t}^2$, or more precisely it must be small on the same dates that v_t^2 is small to prevent the constraint (38) from binding. Of course, ${\psi_t}^2 - \rho_{zz^\star}\psi_t\psi_t^\star$ must also help explain the drift of the exchange rate.

Model B, in which ψ_t and $\psi_t^\star$ depend on the interest rate differential and log exchange

[40] This suggests the literature on the excess volatility of exchange rates [Huang (1981), Wadhwani (1987), Bartolini and Bodnar (1996), and Bartolini and Giorgianni (1999)] is somewhat misguided in its focus on the level, instead of the volatility of volatility.

rate, has only mixed success in reducing the level of excess volatility. For the British pound the average ϵ_t drops substantially from 8.2 to 7.1 percent but for the Deutsche mark it drops only less than 10 basis points. The difference between the results for the two currencies is explained by the fact that the correlation between the log exchange rate (which in model B drives the time-variation in the market prices of currency risk) and the volatility is 0.33 for the British pound and only 0.04 for the Deutsche mark (see Table 2).

When ψ_t and $\psi_t^\star$ also depend on the exchange rate volatility, in model C, the reduction in excess volatility is more impressive. The average ϵ_t drops to 6.6 percent (64 percent of the average volatility) for the British pound and 6.2 percent (58 percent of the average volatility) for the Deutsche mark. Furthermore, especially for the British pound (for which the market prices of currency risk load more on the volatility) the extreme realizations of ϵ_t are reduced substantially relative to the constant market prices of risk model A (from 22 to 12 percent for the British pound and from 18 to 15 percent for the Deutsche mark).

To better visualize the magnitude and time-series properties of the degree of market incompleteness, we plot in Figure 5 as a solid line the excess volatility ϵ_t and as a dashed line the exchange rate volatility v_t. Panels A and B correspond to models B and C, respectively. Comparing the two panels, the importance of allowing the market prices of pure currency risk to depend on the volatility emerges clearly. However, there is still a substantial amount of excess volatility, which raises the challenge of finding variables that help predict changes in the exchange rate and at the same time help reduce the excess volatility.

4.3.4 Specification Tests

We conduct a number of specification tests on the model. The results of these tests, which are not tabulated to save space, can be summarized as follows. First, to verify that the one-week implied volatility is a reasonable proxy for the instantaneous volatility of the exchange rate, we estimate specifications of the model in which the true volatility, denoted h_t, is constant (A: $h_t = h_0$), proportional to the implied volatility (B: $h_t = h_1 v_t$), or linear in the implied volatility (C: $h_t = h_0 + h_1 v_t$). For both currencies, specification A is easily rejected. The estimates of h_1 in specifications B and C range from 0.9 to 0.98 (depending on the currency and on whether the market prices of currency risk are time-varying), which is consistent with an upward-sloping term structure of implied volatilities, but the standard errors are so large that for both currencies we cannot reject (at the ten-percent level) the hypothesis $h_1 = 1$ for B and the joint hypothesis $\{h_0 = 0, h_1 = 1\}$ for C.

Second, we allow the correlations ρ_{wy}, $\rho_{w^\star y}$, and ρ_{xy} to be time-varying as functions of the interest rate differential, log exchange rate, and exchange rate volatility.[41] The results show clearly that ρ_{wy} and $\rho_{w^\star y}$ are not time-varying as a function of these variables. For the third correlation, the results are less clear-cut. When the market prices of currency risk are constant, as in Table 3, there is some evidence (at the ten-percent level) in the case of the US dollar per British pound exchange rate that ρ_{xy} depends on the volatility of the exchange rate. However, when the market prices are time-varying, as in Table 4, this relationship between the correlation and volatility disappears completely.

Finally, we test for non-linearities in the market prices of pure currency risk by including second- and third-order polynomial terms of the interest rate differential, log exchange rate, and exchange rate volatility (with and without cross-terms) in the linear specification (35). For both currencies, these higher-order terms are insignificant.

5 Conclusion

Empiricists now have a transparent, adaptive, and (as our application illustrates) practical econometric method for estimating the parameters of a continuous time diffusion model that inherits the desirable asymptotic properties of the typically unattainable maximum likelihood estimator. The list of potential applications of the SML method in economics and finance is virtually endless. We use it to estimate a new continuous-time model of the joint dynamics of interest rates in two countries and of the exchange rate between the two currencies. The innovation of our model is that it allows for financial markets to be incomplete and specifies the degree of incompleteness as a stochastic process.

Our empirical results for the US dollar per British pound and per Deutsche mark exchange rates offer some new insights into the dynamics of exchange rates. For both currencies we find that the interest rate risk premium is negligible relative to the premium for currency risk orthogonal to interest rate risk. We present evidence that the market prices of pure currency risk are time-varying as a function of the exchange rate and, more importantly, the volatility of the exchange rate. However, even with time-varying market prices of currency risk, a large fraction of the exchange rate volatility is attributed to market incompleteness. We identify the volatility of the exchange rate volatility, rather than the level, as the quantity that is difficult to explain with our parsimoniously parameterized model.

[41] We ensure that the correlation can only taken on values in $[-1, 1]$ by parameterizing it as:

$$\rho_t = 2\frac{\exp\{\rho_0 + \rho_1(r_t - r_t^\star) + \rho_2 e_t + \rho_3 v_t\}}{1 + \exp\{\rho_0 + \rho_1(r_t - r_t^\star) + \rho_2 e_t + \rho_3 v_t\}} - 1.$$

References

Ahn, Dong-Hyun, 1995, Common factors and local factors: Implications for term structures and exchange rates, Working Paper, New York University.

Ahn, Dong-Hyun, and Bin Gao, 1999, A parametric nonlinear model of term structure dynamics, *Review of Financial Studies* 12, 721–762.

Aït-Sahalia, Yacine, 1996a, Nonparametric pricing of interest rate derivatives, *Econometrica* 64, 527–560.

Aït-Sahalia, Yacine, 1996b, Testing continuous-time models of the spot interest rate, *Review of Financial Studies* 9, 385–426.

Aït-Sahalia, Yacine, 2000, Maximum likelihood estimation of discretely-sampled diffusions: A closed form approach, *Econometrica*, forthcoming.

Alizadeh, Sassan, Michael W. Brandt, and Francis X. Diebold, 2000, Range based estimation of stochastic volatility models, Working Paper, University of Pennsylvania.

Andrews, Donald W. K., 1991, Heteroskedasticity and autocorrelation consistent covariance matrix estimation, *Econometrica* 59, 817–858.

Backus, David, Silverio Foresi, and Chris Telmer, 1998, Affine models of currency pricing: Accounting for the forward premium anomaly, Working Paper, New York University.

Baillie, Richard T., and Tim Bollerslev, 1989, The message in daily exchange rates: A conditional variance tale, *Journal of Business and Economic Statistics* 7, 297–305.

Baillie, Richard T., and Tim Bollerslev, 1990, A multivariate generalized GARCH approach to modeling risk premia in foreign exchange markets, *Journal of International Money and Finance* 9, 309–324.

Ball, Clifford A., and Walter N. Torous, 1999, The stochastic volatility of short-term interest rates, *Journal of Finance* 54, 2339-2359.

Bally, Vlad, and Denis Talay, 1995a, Probabilistic numerical methods for partial differential equations: Elements of analysis, in Talay, D., and Tubaro, L. (ed.) *Probabilistic Models for Nonlinear Partial Differential Equations*, Springer Verlag: New York, NY, 148–196.

Bally, Vlad, and Denis Talay, 1995b, The Euler scheme for stochastic differential equations: Error analysis with Malliavin calculus, *Mathematics and Computers in Simulation* 38, 35–41.

Bally, Vlad, and Denis Talay, 1996a, The law of the Euler scheme for stochastic differential equations: I. Convergence rate of the distribution function, *Probability Theory and Related Fields* 104, 43–60.

Bally, Vlad, and Denis Talay, 1996b, The law of the Euler scheme for stochastic differential equations: II. Convergence rate of the density, *Monte Carlo Methods and Applications* 2, 93–128.

Bansal, Ravi, 1997, An exploration of the forward premium puzzle in currency markets, *Review of Financial Studies* 10, 369–403.

Bartolini, Leonardo, and Gordon M. Bodnar, 1996, Are exchange rates excessively volatile? And what does "excessively volatile" mean?, *IMF Staff Papers* 43, 72–96.

Bartolini, Leonardo, and Lorenzo Giogianni, 1999, Excess volatility of exchange rates with unobservable fundamentals, Working Paper, Federal Reserve Bank of New York.

Bekaert, Geert, and Robert J. Hodrick, 1993, On biases in the measurement of foreign exchange risk premiums, *Journal of International Money and Finance* 12, 115–138.

Black, Fischer, and Myron S. Scholes, 1973, The pricing of options and corporate liabilities, *Journal of Political Economy* 81, 637–654.

Bollerslev, Tim, 1987, A conditional heteroskedastic time series model for speculative prices and rates of return, *Review of Economics and Statistics* 69, 542–547.

Campa, Jose M., P.H. Kevin Chang, and Robert L. Reider, 1998, Implied exchange rate distributions: Evidence from OTC options markets, *Journal of International Money and Finance* 17, 117–160.

Chan, K.C., G. Andrew Karolyi, Francis A. Longstaff, and Anthony B. Sanders, 1992, An empirical comparison of alternative models of the short-term interest rate, *Journal of Finance* 47, 1209–1228.

Chapman, David A., and Neil D. Pearson, 1998, Is the short rate drift actually non-linear?, *Journal of Finance* 55, 355–388.

Chen, Ren-Raw, and Louis Scott, 1993, Maximum likelihood estimation for a multifactor equilibrium model of the term structure of interest rates, *Journal of Fixed Income* 4, 14–31.

Cochrane, John H., 2001, *Asset Pricing.* Princeton University Press: Princeton, NJ.

Conley, Timothy G., Lars Peter Hansen, Erzo G.J. Luttmer, and José A. Scheinkman, 1997, Short-term interest rates as subordinated diffusions, *Review of Financial Studies* 10, 525–577.

Cox, John C., Jonathan E. Ingersoll, Jr., and Stephen A. Ross, 1985, A theory of the term structure of interest rates, *Econometrica* 53, 385–407.

Diebold, Francis X., 1988, *Empirical Modeling of Exchange Rate Dynamics*, Springer Verlag: New York, NY.

Diebold, Francis X., and James N. Nason, 1990, Nonparametric exchange rate prediction, *Journal of International Economics* 28, 315–332.

Domowitz, Ian, and Craig S. Hakkio, 1985, Conditional variance and the risk premium of the foreign exchange market, *Journal of International Economics* 19, 47–66.

Duffie, Darrell, 1996, *Dynamic Asset Pricing Theory, Second Edition*, Princeton University Press: Princeton, NJ.

Duffie, Darrell, and Kenneth J. Singleton, 1993, Stimulated moments estimation of Markov models of asset prices, *Econometrica* 50, 987–1007.

Duffie, Darrell, and Kenneth J. Singleton, 1997, Estimation of continuous-time Markov processes sampled at random time intervals, Working Paper, Stanford University.

Duffie, Darrell, and Rui Kan, 1996, A yield-factor model of interest rates, *Mathematical Finance* 6, 379–406.

Duffie, Darrell, and Peter Glynn, 1997, Estimation of Continuous-Time Markov Processes Sampled at Random Time Intervals, Working Paper, Stanford University.

Durham, Garland B., 2000, Likelihood-based specification analysis of models of the short-term interest rate, Working Paper, University of North Carolina.

Durham, Garland B., and A. Ronald Gallant, 2000, Numerical Techniques for Maximum Likelihood estimation of continuous-time diffusion processes, Working Paper, University of North Carolina.

Elerian, Ola, 1998, A note on the existence of a closed form conditional transition density for the Milstein scheme, Working Paper, Nuffield College, Oxford University.

Elerian, Ola, Siddhartha Chib, and Neil Shephard, 2000, Likelihood inference for discretely observed non-linear diffusions, *Econometrica*, forthcoming.

Eraker, B., 1998, *Markov Chain Monte Carlo Analysis of Diffusion Models with Application to Finance*, HAE thesis, Norwegian School of Economics and Business Administration.

Fama, Eugene F., 1984, Forward and spot exchange rates, *Journal of Monetary Economics* 14, 319–338.

Gallant, A. Ronald, and George Tauchen, 1997, Estimation of continuous-time models for stock returns and interest rates, *Macroeconomic Dynamics* 1, 135–168.

Garman, Mark B., and Steven W. Kohlhagen, 1983, Foreign currency option values, *Journal of International Money and Finance* 2, 231–237.

Geyer, Alois L.J., and Stefan Pichler, 1999, A state-space approach to estimate and test multifactor Cox-Ingersoll-Ross models, *Journal of Financial Research* 22, 107–130.

Gibbons, Michael R., and Krishna Ramaswamy, 1993, A test of the Cox, Ingersoll, and Ross model of the term structure, *Review of Financial Studies* 6, 519–568.

Gill, Phillip E., Walter Murray, Michael A. Suanders, and Margaret H. Wright, 1998, User's Guide for NPSOL 5.0: A Fortran package for nonlinear programming, Working Paper, University of California at San Diego.

Gourieroux, Christian, and Alain Monfort, 1993, *Simulation-Based Econometric Methods*, Oxford University Press: Oxford, UK.

Gourieroux, Christian, Alain Monfort, and Eric Renault, 1993, Indirect Inference, *Journal of Applied Econometrics* 8, 85–118.

Hall, Peter, and Christopher Charles Heyde, 1980, *Martingale Limit Theory and Its Applications*, Academic Pres: San Diego, CA.

Hamilton, James D., 1996, The daily market for federal funds, *Journal of Political Economy* 104, 26–56.

Hansen, Lars Peter, 1982, Large sample properties of generalized method of moments estimators, *Econometrica* 50, 1029–1053.

Hansen, Lars Peter, and Kenneth J. Singleton, 1982, Generalized instrumental variables estimation of nonlinear rational expectations models, *Econometrica* 50, 1269–1285.

Hansen, Lars Peter, and Ravi Jagannathan, 1991, Implications of security market data for models of dynamic economies, *Journal of Political Economy* 99, 225–262.

Hansen, Lars Peter, and José A. Scheinkman, 1995, Back to the future: Generating moment implications for continuous-time Markov processes, *Econometrica* 63, 767–804.

Hansen, Lars Peter, and Robert J. Hodrick, 1980, Forward exchange rates as optimal predictors of future exchange rates: An econometric analysis, *Journal of Political Economy* 88, 829–853.

He, Hua, 1990, Moment approximation and estimation of diffusion models of asset prices, Working Paper, University of California at Berkeley.

Hodrick, Robert J., and Sanjay Srivastava, 1984, An investigation of risk and return in forward foreign exchange, *Journal of International Money and Finance* 3, 5–29.

Hodrick, Robert J., and Sanjay Srivastava, 1986, The covariation of risk premiums and expected future spot exchange rates, *Journal of International Money and Finance* 5, 5–22.

Honoré, Peter, 1997, Maximum likelihood estimation of non-linear continuous-time term-structure models, Working Paper, Aarhus School of Business.

Honoré, Peter, 1998, Pitfalls in estimating jump-diffusion models, Working Paper, Aarhus School of Business.

Hsieh, David, 1984, Tests of rational expectations and no risk premium in forward exchange markets, *Journal of International Economics* 17, 173–184.

Huang, Roger D., 1981, The monetary approach to exchange rates in an efficient foreign market: Tests based on volatility, *Journal of Finance* 36, 31–41.

Hurn, A. Stan, and K.A. Lindsay, 1999, Estimating the parameters of stochastic differential equations, *Mathematics and Computers in Simulation* 48, 373–384.

Jones, Christopher S., 2000a, Bayesian estimation of continuous-time finance models, Working Paper, Rochester University.

Jones, Christopher S., 2000b, Nonlinear mean reversion in the short-term interest rate, Working Paper, Rochester University.

Karatzas, Ioannis, and Steven E. Shreve, 1988, *Brownian Motion and Stochastic Calculus*, Springer-Verlag: New York, NY.

Kloeden, Peter, and Eckhard Platen, 1992, *Numeric Solutions of Stochastic Differential Equations*, Springer-Verlag: New York, NY.

Ledoit, Olivier, and Pedro Santa-Clara, 1998, Relative pricing of options with stochastic volatility, Working Paper, University of California at Los Angeles.

Lo, Andrew W., 1988, Maximum likelihood estimation of generalized Itô processes with discretely sampled data, *Econometric Theory* 4, 231–247.

Mella-Barral, Pierre, and William R. Perraudin, 1993, New methods for estimating nonlinear continuous time interest rate processes, Working Paper, London School of Economics.

Merton, Robert C., 1971, Optimum consumption and portfolio rules in a continuous-time model, *Journal of Econometric Theory* 3, 373–413.

Nielsen, Lars Tyge, and Jesús Saá-Requejo, 1993, Exchange Rate and Term Structure Dynamics and the Pricing of Derivative Securities, Working Paper, INSEAD.

Nicolau, João, 2000, New techniques for simulating the likelihood of stochastic differential equations, Working Paper, Instituto Superior de Economia e Gestao/UTL.

Nowman, Khalid B., 1997, Gaussian estimation of single-factor continuous time models of the term structure of interest rates, *Journal of Finance* 52, 1695–1706.

Ogawa, Shigeyoshi,, 1995, Some problems in the simulation of nonlinear diffusion processes, *Mathematics and Computers in Simulations* 28, 217–223.

Pearson, Neil D., and Tong-Sheng Sun, 1994, Exploiting the conditional density in estimating the term structure: An application to the Cox, Ingersoll, and Ross model, *Journal of Finance* 49, 1279–1304.

Pedersen, Asger Roer, 1995a, A new approach to maximum likelihood estimation for stochastic differential equations based on discrete observations, *Scandinavian Journal of Statistics* 22, 55–71.

Pedersen, Asger Roer, 1995b, Consistency and asymptotic normality of an approximate maximum likelihood estimator for discretely observed diffusion processes, *Bernoulli* 1, 257–279.

Piazzesi, Monika, 1999, An econometric model of the yield curve with macroeconomic jump effects, Working Paper, University of California at Los Angeles.

Pritsker, Matt, 1998, Nonparametric density estimation and tests of continuous time interest rate models, *The Review of Financial Studies* 11, 449–487.

Roll, Richard, and Shu Yan, 1998, An explanation of the forward premium puzzle, Working Paper, University of California at Los Angeles.

Saá-Requejo, Jesús, 1993, The dynamics and the term structure of risk premia in the foreign exchange markets, Working Paper, University of Chicago.

Santa-Clara, Pedro, 1995, Simulated likelihood estimation of diffusions with an application to the short term interest rate, in *Essays on the Theory and Estimation of Term Structure Models*. Ph.D. Dissertation, INSEAD.

Singleton, Kenneth J., 1999, Estimation of affine asset pricing models using the empirical characteristic function, Working Paper, Stanford University.

Stanton, Richard, 1997, A nonparametric model of term structure dynamics and the market price of interest rate risk, *Journal of Finance* 52, 1973–2002.

Wadhwani, Sushil B., 1987, Are exchange rates "excessively" volatile?, *Journal of International Economics* 22, 339–348.

Westerfield, Janice Moulton, 1977, An examination of foreign exchange risk under fixed and floating rate regimes, *Journal of International Economics* 7, 181–200.

A Asymptotics

This appendix presents a collection of results for the asymptotics of the SML estimator. Lemmas 1 through 6 combine into a proof of Theorem 1 and Lemma 7 establishes Theorem 2.

Lemma 1: *Given Assumptions 1 and 2, as $M \to \infty$:*[42]

$$q_M\big(Y_{t_{n+1}}, t_{n+1}\big|Y_{t_n}, t_n; \theta\big) - p\big(Y_{t_{n+1}}, t_{n+1}\big|Y_{t_n}, t_n; \theta\big) = O\big(1/M\big). \tag{A1}$$

Proof of Lemma 1: Assumptions 1 and 2 are sufficient for the expansion of $q_M - p$ in powers of $1/M$ by Bally and Talay (1995a,1995b,1996a,1996b). The leading terms of the expansion are proportional to $1/M$ and $1/M^2$ with bounded coefficients.

Lemma 2: *Given Assumptions 1 and 2, as $M \to \infty$ and $S \to \infty$:*

$$\widehat{q}_{M,S}\big(Y_{t_{n+1}}, t_{n+1}\big|Y_{t_n}, t_n; \theta\big) \to p\big(Y_{t_{n+1}}, t_{n+1}\big|Y_{t_n}, t_n; \theta\big) \quad \textit{almost surely.} \tag{A2}$$

Proof of Lemma 2: Recall that from equation (8):

$$\widehat{q}_{M,S}\big(Y_{t_{n+1}}, t_{n+1}\big|Y_{t_n}, t_n; \theta\big) = \frac{1}{S}\sum_{s=1}^{S} \phi\big(Y_{t_{n+1}}; z_s + \mu(z_s)h, V(z_s)h\big), \tag{A3}$$

where we abbreviate $\mu(z_s) \equiv \mu\big(z_s, t_n + (M-1)h; \theta\big)$ and $V(z_s) \equiv V\big(z_s, t_n + (M-1)h; \theta\big)$. The elements of the sum are i.i.d. with finite expectation (by Assumption 2):

$$\mathrm{E}\Big[\phi\big(Y_{t_{n+1}}; z_s + \mu(z_s)h, V(z_s)h\big)\Big] = q_M\big(Y_{t_{n+1}}, t_{n+1}\big|Y_{t_n}, t_n; \theta\big). \tag{A4}$$

Hence, the Strong Law of Large Numbers applies, and as $S \to \infty$:

$$\widehat{q}_{M,S}\big(Y_{t_{n+1}}, t_{n+1}\big|Y_{t_n}, t_n; \theta\big) \to q_M\big(Y_{t_{n+1}}, t_{n+1}\big|Y_{t_n}, t_n; \theta\big) \quad \text{almost surely.} \tag{A5}$$

Use of Lemma 1 completes the proof.

Lemma 3: *Given Assumptions 1 and 2, as $M \to \infty$ and $S \to \infty$, with $S^{1/2}/M \to 0$:*

$$\begin{aligned} S^{1/2}\Big[\widehat{q}_{M,S}\big(Y_{t_{n+1}}, t_{n+1}\big|Y_{t_n}, t_n; \theta\big) - p\big(Y_{t_{n+1}}, t_{n+1}\big|Y_{t_n}, t_n; \theta\big)\Big] \sim \\ \mathrm{N}\Big(0, \mathrm{var}\big[\phi\big(Y_{t_{n+1}}; z_s + \mu(z_s)h, V(z_s)h\big)\big]\Big). \end{aligned} \tag{A6}$$

[42] We use the standard notation $c_n = O(1/n^\delta)$ to denote a sequence c_n for which $\operatorname{plim} n^\delta c_n$ is a finite non-zero constant and $c_n = o(1/n^\delta)$ to denote a sequence c_n for which $\operatorname{plim} n^\delta c_n$ is zero.

Proof of Lemma 3: Write:

$$S^{1/2}\Big[\widehat{q}_{M,S}\big(Y_{t_{n+1}}, t_{n+1}\big|Y_{t_n}, t_n;\theta\big) - p\big(Y_{t_{n+1}}, t_{n+1}\big|Y_{t_n}, t_n;\theta\big)\Big] =$$
$$\frac{1}{S^{1/2}}\sum_{s=1}^{S}\phi\big(Y_{t_{n+1}}; z_s + \mu(z_s)h, V(z_s)h\big) - q_M\big(Y_{t_{n+1}}, t_{n+1}\big|Y_{t_n}, t_n;\theta\big) + \tag{A7}$$
$$S^{1/2}\Big[q_M\big(Y_{t_{n+1}}, t_{n+1}\big|Y_{t_n}, t_n;\theta\big) - p\big(Y_{t_{n+1}}, t_{n+1}\big|Y_{t_n}, t_n;\theta\big)\Big].$$

Lemma 1 and the condition $S^{1/2}/M \to 0$ ensure that as $M \to \infty$ the second term in the sum converges to zero. Applying the Central Limit Theorem to the first term, as in Duffie and Glynn (1996), completes the proof.

Lemma 4: *Given Assumptions 1–4, as $N \to \infty$, $M \to \infty$, and $S \to \infty$, with $S^{1/2}/M \to 0$:*

$$\ln \widehat{\mathcal{L}}_{N,M,S}(\theta) - \ln \mathcal{L}_N(\theta) = o\big(N/S^{1/2}\big). \tag{A8}$$

Proof of Lemma 4: Let x_n denote the errors of the simulated transition densities:

$$x_n \equiv \widehat{q}_{M,S}\big(Y_{t_{n+1}}, t_{n+1}\big|Y_{t_n}, t_n;\theta\big) - p\big(Y_{t_{n+1}}, t_{n+1}\big|Y_{t_n}, t_n;\theta\big). \tag{A9}$$

Abbreviate $p_n \equiv p\big(Y_{t_{n+1}}, t_{n+1}\big|Y_{t_n}, t_n;\theta\big)$ and write:

$$\ln \widehat{q}_{M,S}\big(Y_{t_{n+1}}, t_{n+1}\big|Y_{t_n}, t_n;\theta\big) - \ln p_n = \ln\big(x_n + p_n\big) - \ln p_n = \ln\left(1 + \frac{x_n}{p_n}\right). \tag{A10}$$

Expanding the last term around $x_n = 0$ for a fixed p_n implies that for a sufficiently small x_n:

$$\ln\left(1 + \frac{x_n}{p_n}\right) \approx \frac{x_n}{p_n} + o\big(x_n\big) = o\big(1/S^{1/2}\big). \tag{A11}$$

The last equality follows from Lemma 3. Substituting the expansion (A11) into equation (A10) and summing over the N sample points completes the proof.

Lemma 5: *Given Assumptions 1–4, as $N \to \infty$, $M \to \infty$, and $S \to \infty$, with $S^{1/2}/M \to 0$:*

$$\widehat{\theta}_{N,M,S} - \widehat{\theta}_N = o\big(N^{1/2}/S^{1/4}\big), \tag{A12}$$

where $\widehat{\theta}_N$ is the parameter vector that maximizes $\ln \mathcal{L}_N(\theta)$.

Proof of Lemma 5: A second-order expansion of $\ln \mathcal{L}_N\big(\widehat{\theta}_{N,M,S}\big)$ around $\widehat{\theta}_N$ yields:

$$\ln \mathcal{L}_N\big(\widehat{\theta}_{N,M,S}\big) = \ln \mathcal{L}_N\big(\widehat{\theta}_N\big) +$$
$$\frac{\partial \ln \mathcal{L}_N\big(\widehat{\theta}_N\big)}{\partial\theta}\big(\widehat{\theta}_{N,M,S} - \widehat{\theta}_N\big) + \frac{1}{2}\frac{\partial^2 \ln \mathcal{L}_N\big(\bar{\theta}\big)}{\partial\theta^2}\big(\widehat{\theta}_{N,M,S} - \widehat{\theta}_N\big)^2, \tag{A13}$$

where $\bar{\theta}$ is a convex combination of $\widehat{\theta}_{N,M,S}$ and $\widehat{\theta}_N$. The first order term of the expansion is

zero, since $\widehat{\theta}_N$ maximizes $\ln \mathcal{L}_N(\theta)$.

An analogous second-order expansion of $\ln \widehat{\mathcal{L}}_{N,M,S}(\widehat{\theta}_N)$ around $\widehat{\theta}_{N,M,S}$ yields:

$$\begin{aligned}\ln \widehat{\mathcal{L}}_{N,M,S}(\widehat{\theta}_N) &= \ln \widehat{\mathcal{L}}_{N,M,S}(\widehat{\theta}_{N,M,S}) + \\ &\frac{\partial \ln \widehat{\mathcal{L}}_{N,M,S}(\widehat{\theta}_{N,M,S})}{\partial \theta}(\widehat{\theta}_N - \widehat{\theta}_{N,M,S}) + \frac{1}{2}\frac{\partial^2 \ln \widehat{\mathcal{L}}_{N,M,S}(\tilde{\theta})}{\partial \theta^2}(\widehat{\theta}_N - \widehat{\theta}_{N,M,S})^2,\end{aligned} \tag{A14}$$

where $\tilde{\theta}$ is another convex combination of $\widehat{\theta}_{N,M,S}$ and $\widehat{\theta}_N$. The first order term of the expansion is again zero, since $\widehat{\theta}_{N,M,S}$ maximizes $\ln \widehat{\mathcal{L}}_{N,M,S}(\theta)$.

Summing the two expansions (A13)–(A14) and rearranging terms:

$$\begin{aligned}&\left[\ln \widehat{\mathcal{L}}_N(\widehat{\theta}_{N,M,S}) - \ln \widehat{\mathcal{L}}_{N,M,S}(\widehat{\theta}_{N,M,S})\right] + \left[\ln \widehat{\mathcal{L}}_N(\widehat{\theta}_N) - \ln \widehat{\mathcal{L}}_{N,M,S}(\widehat{\theta}_N)\right] = \\ &\qquad \left[\frac{\partial \ln \widehat{\mathcal{L}}_{N,M,S}(\widehat{\theta}_{N,M,S})}{\partial \theta} - \frac{\partial \ln \widehat{\mathcal{L}}_N(\widehat{\theta}_N)}{\partial \theta}\right](\widehat{\theta}_N - \widehat{\theta}_{N,M,S}) + \\ &\qquad\qquad \frac{1}{2}\left[\frac{\partial^2 \ln \widehat{\mathcal{L}}_N(\bar{\theta})}{\partial \theta^2} + \frac{\partial^2 \ln \widehat{\mathcal{L}}_{N,M,S}(\tilde{\theta})}{\partial \theta^2}\right](\widehat{\theta}_N - \widehat{\theta}_{N,M,S})^2.\end{aligned} \tag{A15}$$

From Lemma 4, the two differences between the actual and simulated log likelihood functions are $o(N/S^{1/2})$. The first-order derivatives are both zero by construction and the second-order derivatives are bounded by Assumption 4. The result (A12) follows.

Lemma 6 *Given Assumptions 1–5, as $N \to \infty$:*

$$\widehat{\theta}_N \to \theta_0. \tag{A16}$$

Proof of Lemma 6: Define:

$$u_{n+1}(\theta) = \frac{\partial \ln p(Y_{t_{n+1}}, t_{n+1} | Y_{t_n}, t_n; \theta)}{\partial \theta} \quad \text{and} \quad v_{n+1}(\theta) = \frac{\partial^2 \ln p(Y_{t_{n+1}}, t_{n+1} | Y_{t_n}, t_n; \theta)}{\partial \theta^2}. \tag{A17}$$

It is well know that $\mathrm{E}[v_{n+1}(\theta_0)|\mathcal{F}_n] = -\mathrm{E}[{u_{n+1}}^2(\theta_0)|\mathcal{F}_n]$, which implies that:

$$I_N(\theta_0) = -\mathrm{E}\left[\sum_{n=0}^{N-1} v_{n+1}(\theta_0) \middle| \mathcal{F}_n\right] \equiv -\mathrm{E}[J_N(\theta_0)|\mathcal{F}_n]. \tag{A18}$$

Therefore, $J_N(\theta) + I_N(\theta)$ is a martingale.

An expansion of the first-order conditions $\partial \ln \mathcal{L}_N(\widehat{\theta}_N)/\partial \theta = 0$ around θ_0 yields:

$$\frac{\partial \ln \mathcal{L}_N(\widehat{\theta}_N)}{\partial \theta} = \sum_{n=0}^{N-1} u_{n+1}(\theta_0) - I_N(\theta_0)(\widehat{\theta}_N - \theta_0) + \left(J_N(\bar{\theta}) + I_N(\theta_0)\right)(\widehat{\theta}_N - \theta_0) = 0, \tag{A19}$$

where $\bar{\theta}$ is a convex combination of $\widehat{\theta}_N$ and θ_0. Rearrange this expansion to isolate $\widehat{\theta}_N - \theta_0$:

$$(\widehat{\theta}_N - \theta_0)\Big[1 - I_N(\theta_0)^{-1}\Big(J_N(\bar{\theta}) + I_N(\theta_0)\Big)\Big] = I_N(\theta_0)^{-1}\sum_{n=0}^{N-1} u_{n+1}(\theta_0). \tag{A20}$$

As in Theorem 2.18 of Hall and Heyde (1980), if $I_N(\theta_0) \to \infty$ as $N \to \infty$, the right-hand side of equation (A20) converges to zero, or:

$$(\widehat{\theta}_N - \theta_0)\Big[1 - I_N(\theta_0)^{-1}\Big(J_N(\bar{\theta}) + I_N(\theta_0)\Big)\Big] \to 0 \quad \text{almost surely.} \tag{A21}$$

Since there always exists a $\tilde{\theta}$ between $\bar{\theta}$ and θ_0 such that:

$$J_N(\bar{\theta}) = J_N(\theta_0) + \frac{\partial J_N(\tilde{\theta})}{\partial \theta}(\tilde{\theta} - \theta_0), \tag{A22}$$

the bracketed term in equation (A21) can be rewritten as:

$$\Big[1 - I_N(\theta_0)^{-1}\Big(J_N(\theta_0) + I_N(\theta_0)\Big)\Big] + I_N(\theta_0)^{-1}\frac{\partial J_N(\tilde{\theta})}{\partial \theta}(\tilde{\theta} - \theta_0). \tag{A23}$$

If $I_N(\theta_0) \to \infty$ as $N \to \infty$, the second term converges to zero, since J_N has a bounded gradient, from Assumption 4. The first term converges to one almost surely, since $J_N + I_N$ is a martingale.

Lemma 7: *Given Assumptions 1–6, as $N \to \infty$:*

$$I_N(\theta_0)^{1/2}\big[\widehat{\theta}_N - \theta_0\big] \sim \mathrm{N}(0,1). \tag{A24}$$

Proof of Lemma 7: From Hall and Hyde (1980), sufficient conditions for the Central Limit Theorem are that as $N \to \infty$:

$$\sum_{n=0}^{N-1} \mathrm{E}\Big[I_N(\theta_0)^{-1} u_{n+1}{}^2(\theta_0)\Big|\mathcal{F}_n\Big] \to 1 \quad \text{in probability,} \tag{A25}$$

and

$$\max_{0\le n\le N-1}\Big|I_N(\theta_0)^{-1/2} u_{n+1}(\theta_0)\Big| \to 0 \quad \text{in distribution.} \tag{A26}$$

The first condition is trivial, since $I_N = \sum u_n^2$. The second condition is satisfied if $I_N(\theta_0) \to \infty$ and if u_{n+1} behaves according to Assumption 6.

B Exchange Rate Volatility Dynamics

In this appendix, we derive the dynamics of the volatility of the exchange rate for the model of section 3.2. This derivation is basically of a tedious but straightforward application of Itô's Lemma in matrix form.

Define $d\overline{W} \equiv \begin{bmatrix} dW \\ dW^\star \\ dX \\ dY \end{bmatrix}$, and its correlation matrix $\Sigma = \text{Corr}(d\overline{W})$. We know that:

$$dv^2 = \mu dt + \sigma_W dW + \sigma_{W^\star} dW^\star + \sigma_X dX + \sigma_Y dY. \tag{B1}$$

Denoting $\sigma_v = \begin{bmatrix} \sigma_W \\ \sigma_{W^\star} \\ \sigma_X \\ \sigma_Y \end{bmatrix}$, we have:

$$dv = \left(\frac{\mu}{2v} - \frac{1}{8v^3}\sigma_v'\Sigma\sigma_v\right)dt + \frac{1}{2v}\left(\sigma_W dW + \sigma_{W^\star} dW^\star + \sigma_X dX + \sigma_Y dY\right). \tag{B2}$$

Let $R_1 = \begin{bmatrix} \sqrt{r} \\ \sqrt{r^\star} \end{bmatrix}$, and $R_2 = \begin{bmatrix} r \\ r^\star \\ e \\ v \end{bmatrix}$. Then R_1 and R_2 are governed by the following processes:

$$\begin{aligned} dR_1 =& f_1 dt + g_1 d\overline{W} \\ dR_2 =& f_2 dt + g_2 d\overline{W}, \end{aligned}$$

with

$$f_1 = \begin{bmatrix} \frac{1}{2}r^{-\frac{1}{2}}\lambda(\theta - r) - \frac{1}{8}r^{-\frac{1}{2}}\sigma^2 \\ \frac{1}{2}r^{\star-\frac{1}{2}}\lambda^\star(\theta^\star - r^\star) - \frac{1}{8}r^{\star-\frac{1}{2}}\sigma^{\star 2} \end{bmatrix}, f_2 = \begin{bmatrix} \lambda(\theta - r) \\ \lambda^\star(\theta^\star - r^\star) \\ (r - r^\star) + (\rho_{wx}\phi + \rho_{zx}\psi)v - \frac{1}{2}v^2 \\ \frac{\mu}{2v} - \frac{1}{8v^3}\sigma_v'\Sigma\sigma_v \end{bmatrix}, \tag{B4}$$

$$g_1 = \frac{1}{2}\begin{bmatrix} \sigma & 0 & 0 & 0 \\ 0 & \sigma^\star & 0 & 0 \end{bmatrix}, \text{ and } g_2 = \begin{bmatrix} \sigma\sqrt{r} & 0 & 0 & 0 \\ 0 & \sigma^\star\sqrt{r^\star} & 0 & 0 \\ 0 & 0 & v & 0 \\ \frac{1}{2v}\sigma_W & \frac{1}{2v}\sigma_{W^\star} & \frac{1}{2v}\sigma_X & \frac{1}{2v}\sigma_Y \end{bmatrix}. \tag{B5}$$

Note that:

$$\begin{bmatrix} \phi \\ \phi^\star \end{bmatrix} = B_1 R_1 \quad \text{and} \quad \begin{bmatrix} \psi \\ \psi^\star \end{bmatrix} = A_2 + B_2 R_2, \tag{B6}$$

where

$$B_1 = \begin{bmatrix} \phi & 0 \\ 0 & \phi^\star \end{bmatrix}, \quad A_2 = \begin{bmatrix} \psi_0 \\ \psi_0^\star \end{bmatrix}, \quad \text{and} \quad B_2 = \begin{bmatrix} \psi_2 & -\psi_2 & \psi_1 & \psi_3 \\ \psi_2^\star & -\psi_2^\star & \psi_1^\star & \psi_3^\star \end{bmatrix}. \tag{B7}$$

Denote: $C_1 = \begin{bmatrix} 1 & -\rho_{ww^\star} \\ -\rho_{ww^\star} & 1 \end{bmatrix}$, and $C_2 = \begin{bmatrix} 1 & -\rho_{zz^\star} \\ -\rho_{zz^\star} & 1 \end{bmatrix}$, we can write:

$$v^2 = \frac{1}{2} R_1' \Omega_1 R_1 + \frac{1}{2}\Theta_2 + \Gamma_2 R_2 + \frac{1}{2} R_2' \Omega_2 R_2 + \epsilon^2 \tag{B8}$$

where $\Omega_1 = 2B_1' C_1 B_1$, $\Theta_2 = 2A_2' C_2 A_2$, $\Gamma_2 = 2A_2' C_2 B_2$, and $\Omega_2 = 2B_2' C_2 B_2$. Thus:

$$\begin{aligned}
dv^2 &= R_1'\Omega_1 dR_1 + \frac{1}{2} dR_1'\Omega_1 dR_1 + (\Gamma_2 + R_2'\Omega_2) dR_2 + \frac{1}{2} dR_2'\Omega_2 dR_2 + d\epsilon^2 \\
&= R_1'\Omega_1 dR_1 + \frac{1}{2}\mathrm{Tr}(dR_1'\Omega_1 dR_1) + (\Gamma_2 + R_2'\Omega_2) dR_2 + \frac{1}{2}\mathrm{Tr}(dR_2'\Omega_2 dR_2) + d\epsilon^2 \\
&= R_1'\Omega_1 (f_1 dt + g_1 d\overline{W}) + \frac{1}{2}\mathrm{Tr}(\Omega_1 g_1 \Sigma g_1') dt \\
&\quad + (\Gamma_2 + R_2'\Omega_2)(f_2 dt + g_2 d\overline{W}) + \frac{1}{2}\mathrm{Tr}(\Omega_2 g_2 \Sigma g_2') dt + d\epsilon^2 \qquad \text{(B9)} \\
&= \left[R_1'\Omega_1 f_1 + (\Gamma_2 + R_2'\Omega_2) f_2 + \frac{1}{2}\mathrm{Tr}(\Omega_1 g_1 \Sigma g_1') + \frac{1}{2}\mathrm{Tr}(\Omega_2 g_2 \Sigma g_2') - 2\alpha\epsilon^2 + \beta^2 \right] dt \\
&\quad + \left[R_1'\Omega_1 g_1 + (\Gamma_2 + R_2'\Omega_2) g_2 \right] d\overline{W} + 2\beta|\epsilon| dY \\
&= v_\mu dt + v_W dW + v_{W^\star} dW^\star + v_X dX + v_Y dY
\end{aligned}$$

Therefore, we have a set of five equations, which determine μ, σ_W, $\sigma_{W^\star}$, σ_X and σ_Y:

$$\begin{aligned}
\mu &= v_\mu \\
\sigma_W &= v_W \\
\sigma_{W^\star} &= v_{W^\star} \\
\sigma_X &= v_X \\
\sigma_Y &= v_Y.
\end{aligned} \tag{B10}$$

Table 1
Descriptive Statistics

This table presents descriptive statistics of one-week Euro-currency rates r in the US and $r^\star$ in the UK or Germany, one-year Euro-currency yields y and $y^\star$, interest rate differentials $r-r^\star$, log US dollar per British pound or Deutsche mark exchange rates e, implied volatilities v of one-week at-the-money British pound or Deutsche mark options, and weekly differences Δr, $\Delta r^\star$, Δe, and Δv. The means and standard deviations of the first differences are annualized. The data are weekly observations from January 1990 through May 2000 (544 observations).

Panel A: Levels

	US		UK					Germany				
Statistic	r	y	$r^\star$	$y^\star$	$r-r^\star$	e	v	$r^\star$	$y^\star$	$r-r^\star$	e	v
Mean	0.052	0.056	0.078	0.077	-0.025	0.492	0.103	0.056	0.057	-0.003	-0.499	0.107
Standard Deviation	0.014	0.013	0.031	0.025	0.025	0.071	0.030	0.025	0.024	0.029	0.092	0.023
Skewness	0.291	0.208	1.266	1.334	-0.678	0.879	0.933	0.404	0.519	-0.754	-0.254	0.456
Kurtosis	3.109	2.845	3.295	3.780	1.905	3.468	3.710	1.516	1.676	2.051	2.893	3.308
Auto-Correlation	0.996	0.994	0.996	0.997	0.993	0.981	0.928	0.999	0.999	0.999	0.987	0.824

Panel B: Differences

	US	UK			Germany		
Statistic	Δr	$\Delta r^\star$	Δe	Δv	$\Delta r^\star$	Δe	Δv
Mean	-0.002	-0.009	-0.007	-0.000	-0.004	-0.019	0.002
Standard Deviation	0.009	0.020	0.101	0.082	0.009	0.106	0.098
Skewness	1.674	-0.396	-1.360	0.853	0.734	-0.363	0.644
Kurtosis	24.757	6.433	9.870	9.903	22.995	5.467	5.017
Auto-Correlation	0.028	-0.249	0.066	-0.196	-0.027	0.001	-0.243

Table 2
Correlation Matrix

This table presents pairwise correlations of one-week Euro-currency rates r in the US and $r^\star$ in the UK or Germany, one-year Euro-currency yields y and $y^\star$, interest rate differentials $r-r^\star$, log US dollar per British pound or Deutsche mark exchange rates e, implied volatilities v of one-week at-the-money British pound or Deutsche mark options, and weekly differences Δr, $\Delta r^\star$, Δe, and Δv.

	Levels							Differences			
	r	y	$r^\star$	$y^\star$	$r-r^\star$	e	v	Δr	$\Delta r^\star$	Δe	Δv
Levels											
US											
r	1.000							-0.090			
y	0.942	1.000						0.005			
UK											
$r^\star$	0.590	0.494	1.000					-0.163			
$y^\star$	0.662	0.608	0.964	1.000				-0.134			
$r-r^\star$	-0.178	-0.092	-0.899	-0.816	1.000			0.151			
e	0.420	0.317	0.727	0.717	-0.659	1.000		-0.172			
v	-0.166	-0.215	0.484	0.402	-0.680	0.363	1.000	-0.112			
Germany											
$r^\star$	-0.087	-0.106	1.000					-0.142			
$y^\star$	0.103	0.099	0.971	1.000				-0.136			
$r-r^\star$	0.535	0.525	-0.888	-0.775	1.000			0.079			
e	-0.009	-0.109	0.251	0.211	-0.217	1.000		-0.083			
v	-0.095	-0.117	0.377	0.360	-0.364	0.036	1.000	-0.058			
Differences											
UK											
$\Delta r^\star$	0.036	0.038	-0.100	-0.025	0.141	-0.005	-0.132	0.010	1.000		
Δe	0.077	0.083	0.018	0.039	0.020	-0.090	-0.113	-0.077	-0.047	1.000	
Δv	0.028	0.022	0.011	0.028	0.002	0.082	-0.190	-0.044	0.195	-0.125	1.000
Germany											
$\Delta r^\star$	0.090	0.116	-0.046	0.004	0.081	-0.076	0.002	0.403	1.000		
Δe	0.037	0.049	0.044	0.062	-0.021	-0.043	0.007	-0.097	0.062	1.000	
Δv	0.011	0.008	-0.006	0.002	0.010	-0.002	-0.294	-0.040	-0.010	-0.069	1.000

Table 3
Simulated Maximum Likelihood Estimates of the Model

This table presents simulated maximum likelihood estimates of the model:

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

$$dr_t^\star = \lambda^\star\left(\theta^\star - r_t^\star\right)dt + \sigma^\star\sqrt{r_t^\star}\,dW_t^\star$$

$$de_t = \left[\left(r_t - r_t^\star\right) + \sigma_{wx}\phi_t + \sigma_{ze}\psi - \frac{1}{2}{v_t}^2\right]dt + v_t dX_t,$$

$$v_t^2 = ({\phi_t}^2 + {\phi_t^\star}^2 - 2\,\rho_{ww^\star}\phi_t\phi_t^\star) + ({\psi}^2 + {\psi^\star}^2 - 2\rho_{zz^\star}\psi\psi^\star) + {\epsilon_t}^2,$$

$$d{\epsilon_t}^2 = \left(-2\,\alpha\,{\epsilon_t}^2 + \beta^2\right)dt + 2\,\beta\left|\epsilon_t\right|dY_t,$$

where $\sigma_{we} = \rho_{wz}v_t = \phi_t - \rho_{ww^\star}\phi_t^\star$, $\sigma_{ze} = \rho_{zx}v_t = \psi - \rho_{zz^\star}\psi^\star$, $\phi_t = \phi\sqrt{r_t}$, $\phi_t^\star = \phi^\star\sqrt{r_t^\star}$, and:

$$\mathrm{Corr}\begin{bmatrix} dW_t \\ dW_t^\star \\ dX_t \\ dY_t \end{bmatrix} = \begin{bmatrix} 1.00 & & & \\ \rho_{ww^\star} & 1.00 & & \\ \dfrac{\phi_t - \rho_{ww^\star}\phi_t\phi_t^\star}{v_t} & \dfrac{\phi_t^\star - \rho_{ww^\star}\phi_t\phi_t^\star}{v_t} & 1.00 & \\ \rho_{wy} & \rho_{w^\star y} & \rho_{xy} & 1.00 \end{bmatrix}.$$

In brackets are asymptotic standard errors. All parameters are annualized.

	US vs. UK		US vs. Germany	
	Interest Rates			
λ and $\lambda^\star$	0.284 [0.239]	0.486 [0.255]	0.305 [0.263]	0.088 [0.299]
θ and $\theta^\star$	0.053 [0.017]	0.074 [0.038]	0.058 [0.021]	0.064 [0.038]
σ and $\sigma^\star$	0.028 [0.001]	0.056 [0.001]	0.027 [0.001]	0.042 [0.001]
ρ_{ww}	0.057 [0.012]		0.213 [0.046]	
	Market Prices of Risk			
ϕ and $\phi^\star$	−0.138 [0.048]	0.027 [0.016]	−0.125 [0.043]	−0.036 [0.018]
$\psi^2 - \rho_{zz^\star}\psi\psi^\star$	0.024	[0.022]	-0.010	[0.012]
$\psi^{\star 2} - \rho_{zz^\star}\psi\psi^\star$	-0.021	[0.017]	0.013	[0.010]
	Market Incompleteness / Excess Volatility			
α	0.320 [0.112]		0.338 [0.125]	
β	0.088 [0.001]		0.101 [0.002]	
ρ_{wy}	0.048 [0.005]		0.058 [0.008]	
$\rho_{w^\star y}$	0.034 [0.005]		−0.034 [0.009]	
ρ_{xy}	−0.012 [0.003]		0.006 [0.002]	

Table 4
Time-Varying Market Prices of Pure Currency Risk

This table presents simulated maximum likelihood estimates of the model described in Table 3 with constant (model A) and time-varying (models B and C) market prices of currency risk orthogonal to interest rate risk:

$$\psi_t = \psi_0 + \psi_1(r_t - r_t^\star) + \psi_2\, e_t + \psi_3\, v_t \quad \text{and} \quad \psi_t^\star = \psi_0^\star + \psi_1^\star(r_t - r_t^\star) + \psi_2^\star e_t + \psi_3^\star v_t.$$

The table also describes the implied pure currency risk premium:

$$\sigma_{ze}\psi_t = \psi_t^{\,2} - \rho_{zz^\star}\psi_t\psi_t^\star$$

and excess volatility:

$$|\epsilon_t| = \sqrt{v_t^{\,2} - \left(\phi_t^{\,2} + \phi_t^{\star 2} - 2\,\rho_{ww^\star}\phi_t\phi_t^\star\right) - \left(\psi_t^{\,2} + \psi_t^{\star 2} - 2\rho_{zz^\star}\psi_t\psi_t^\star\right)}.$$

In brackets are asymptotic standard errors. All entries are annualized.

	ψ_t				$\psi_t^\star$					Risk Premium		Excess Volatility	
	ψ_0	ψ_1	ψ_2	ψ_3	$\psi_0^\star$	$\psi_1^\star$	$\psi_2^\star$	$\psi_3^\star$	$\rho_{zz^\star}$	Mean StdDev	Min Max	Mean StdDev	Min Max
						US vs. UK							
A				$\psi^2 - \rho_{zz^\star}\psi\psi^\star =$	0.024 [0.022]					0.0241		0.0820	0.0000
				$\psi^{\star 2} - \rho_{zz^\star}\psi\psi^\star =$	−0.021 [0.017]							0.0364	0.2151
B	0.062	0.117	-0.263		-0.101	0.418	0.135		0.54	0.0036	0.0001	0.0711	0.0000
	[0.043]	[0.237]	[0.144]		[0.061]	[0.451]	[0.078]		[0.13]	0.0029	0.0141	0.0353	0.2099
C	-0.026	0.019	0.134	1.514	0.065	0.002	-0.329	2.581	0.89	0.0032	-0.0414	0.0660	0.0000
	[0.019]	[0.025]	[0.076]	[0.633]	[0.041]	[0.007]	[0.179]	[1.001]	[0.17]	0.0101	0.0286	0.0195	0.1234
						US vs. Germany							
A				$\psi^2 - \rho_{zz^\star}\psi\psi^\star =$	−0.010 [0.012]					-0.0104		0.0875	0.0000
				$\psi^{\star 2} - \rho_{zz^\star}\psi\psi^\star =$	0.013 [0.010]							0.0290	0.1791
B	-0.140	0.066	-0.197		0.089	-0.003	0.251		0.36	0.0017	0.0001	0.0867	0.0000
	[0.121]	[0.049]	[0.104]		[0.131]	[0.004]	[0.129]		[0.09]	0.0016	0.0065	0.0277	0.1782
C	-0.072	-0.073	0.006	-0.142	0.029	0.028	0.238	0.127	0.94	0.0022	-0.0056	0.0624	0.0000
	[0.050]	[0.059]	[0.004]	[0.072]	[0.021]	[0.035]	[0.134]	[0.072]	[0.21]	0.0034	0.0132	0.0245	0.1450

Figure 1
Approximating the Transition Densities

This figure illustrates the approximation of the transition densities of a diffusion. The solid line represents the unobserved continuous-time sample path of a univariate diffusion. The four dashed lines represent incomplete ten-step Euler discretizations.

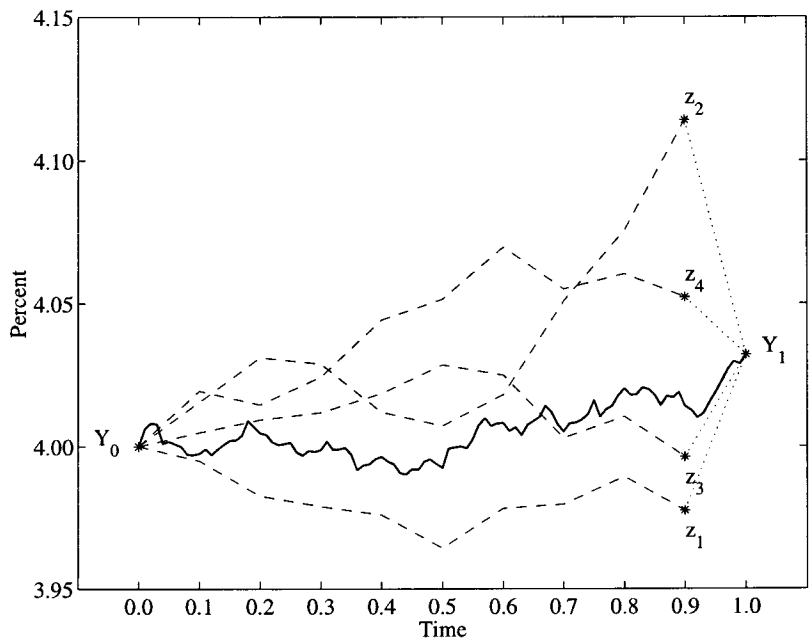
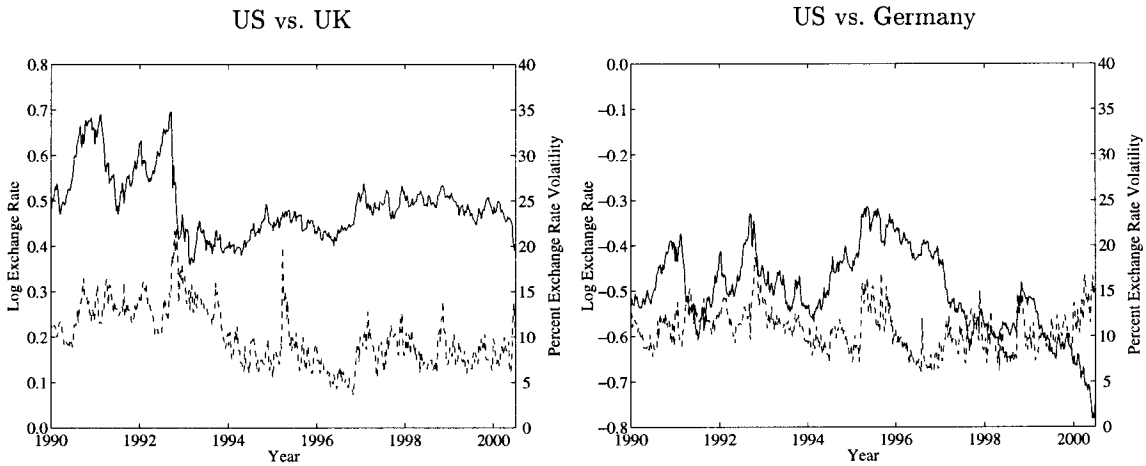


Figure 2
Log Exchange Rate, Exchange Rate Volatility, and Interest Rates

This figure shows in Panel A as solid lines the log US dollar per British pound or log US dollar per Deutsche mark exchange rates and as dashed lines the one-week implied volatilities of at-the-money British pound or Deutsche mark options. It shows in Plot B as solid and dashed lines the one-week Euro-dollar and Euro-sterling or Euro-dollar and Euro-mark interest rates, respectively

Panel A: Log Exchange Rate and Exchange Rate Volatility



Panel B: Interest Rates

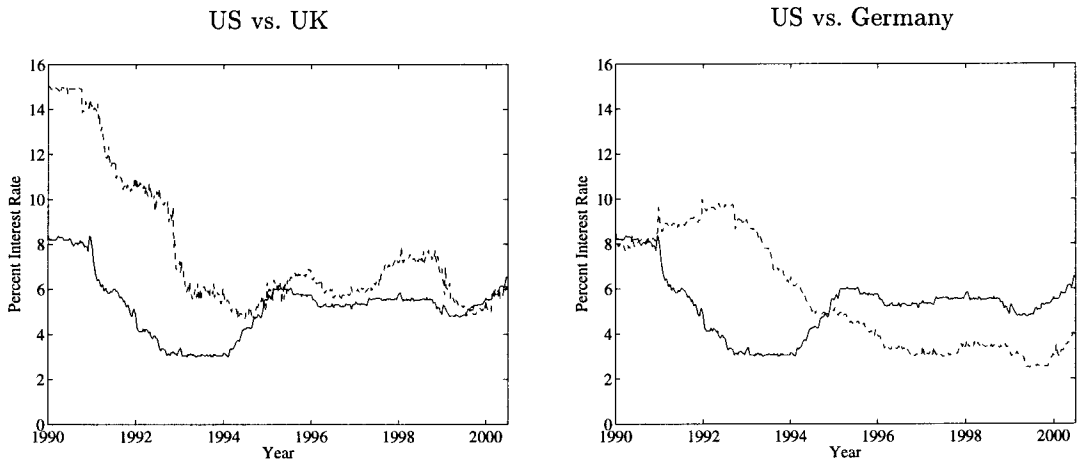
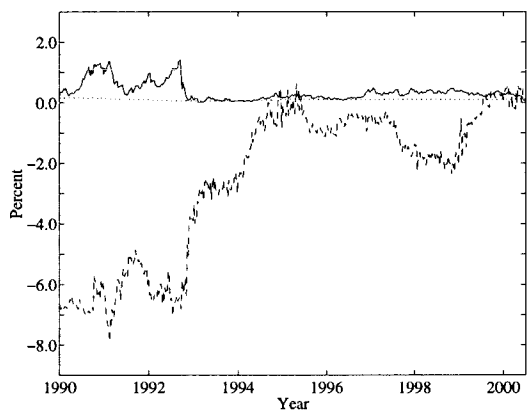


Figure 3
Exchange Rate Drift Decomposition

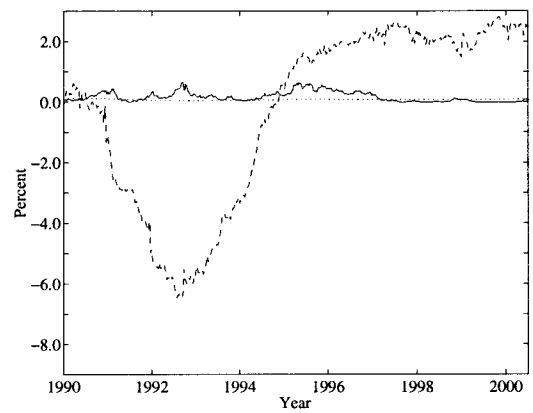
This figure decomposes the drift κ_t of the exchange rate. It plots as solid line the currency risk premium orthogonal to interest rate risk $\psi_t^2 - \rho_{zz^\star}\psi_t\psi_t^\star$, as dashed line the interest rate differential $r_t - r_t^\star$, and as dotted line the interest rate risk premium $\phi_t^2 - \rho_{ww^\star}\phi_t\phi_t^\star$. In Panel A, the market prices of risk ψ_t and $\psi_t^\star$ depend linearly on the interest rate differential and log exchange rate e_t. In Panel B, the market prices of risk also depend linearly on the exchange rate volatility v_t.

Panel A: $\psi_t = \psi(r_t - r_t^\star, e_t)$ and $\psi_t^\star = \psi^\star(r_t - r_t^\star, e_t)$

US vs. UK

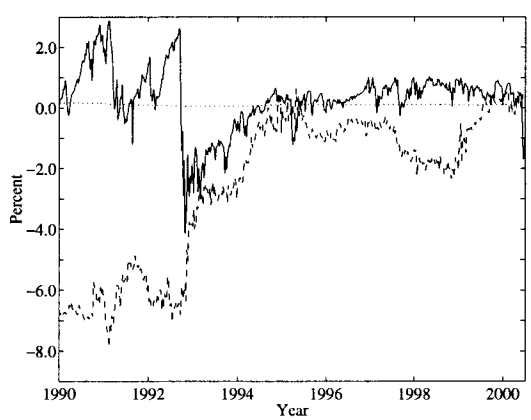


US vs. Germany



Panel B: $\psi_t = \psi(r_t - r_t^\star, e_t, v_t)$ and $\psi_t^\star = \psi^\star(r_t - r_t^\star, e_t, v_t)$

US vs. UK



US vs. Germany

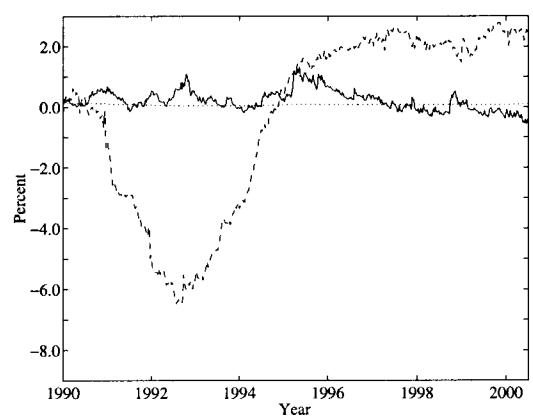


Figure 4
Constrained and Unconstrained Currency Risk Premium

This figure plots the currency risk premium $\psi_t^2 - \rho_{zz^\star}\psi_t\psi_t^\star$. The market prices of risk ψ_t and $\psi_t^\star$ depend linearly on the interest rate differential $r_t - r_t^\star$, log exchange rate e_t, and exchange rate volatility v_t. The solid line is the unconstrained risk premium. The dashed line is the constrained risk premium that satisfies ${v_t}^2 \geq ({\phi_t}^2 + {\phi_t^\star}^2 - 2\,\rho_{ww^\star}\phi_t\phi_t^\star) + ({\psi_t}^2 + {\psi_t^\star}^2 - 2\rho_{zz^\star}\psi_t\psi_t^\star)$ for all t. The dotted line is the interest rate differential $r_t - r_t^\star$.

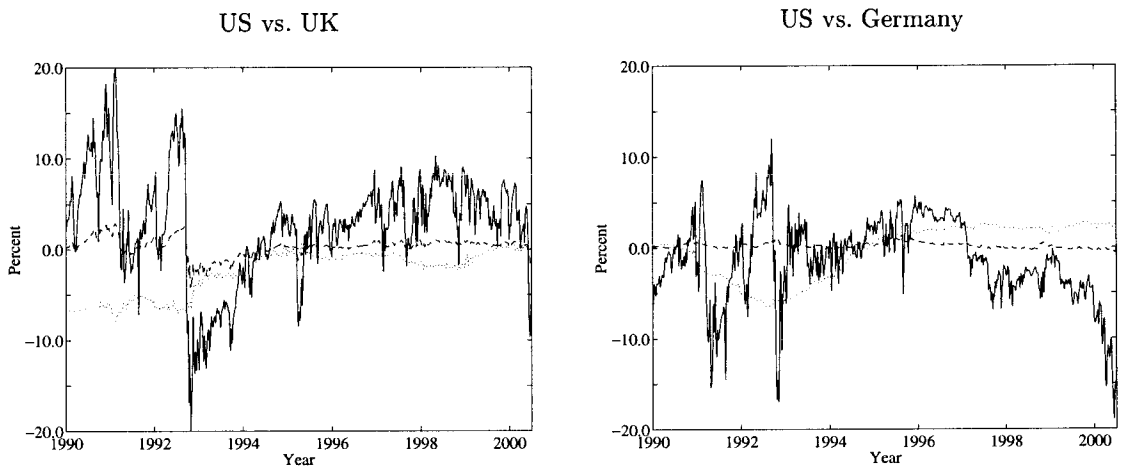
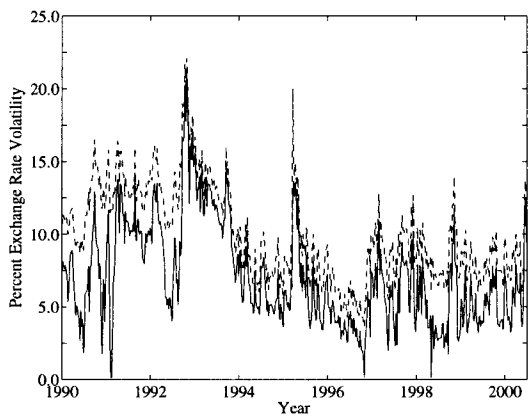


Figure 5
Exchange Rate Volatility and Excess Volatility

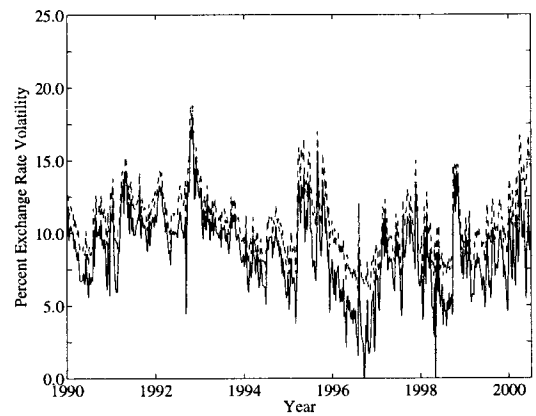
This figure shows as dashed line the exchange rate volatility v_t and as solid line the excess volatility $|\epsilon_t| = [{v_t}^2 - ({\phi_t}^2 + {\phi_t^\star}^2 - 2\,\rho_{ww^\star}\phi_t\phi_t^\star) - ({\psi_t}^2 + {\psi_t^\star}^2 - 2\rho_{zz^\star}\psi_t\psi_t^\star)]^{1/2}$. In Panel A, the market prices of risk ψ_t and $\psi_t^\star$ depend linearly on the interest rate differential $r_t - r_t^\star$ and log exchange rate e_t. In Panel B, the market prices of risk also depend linearly on the exchange rate volatility v_t.

Panel A: $\psi_t = \psi(r_t - r_t^\star, e_t)$ and $\psi_t^\star = \psi^\star(r_t - r_t^\star, e_t)$

US vs. UK

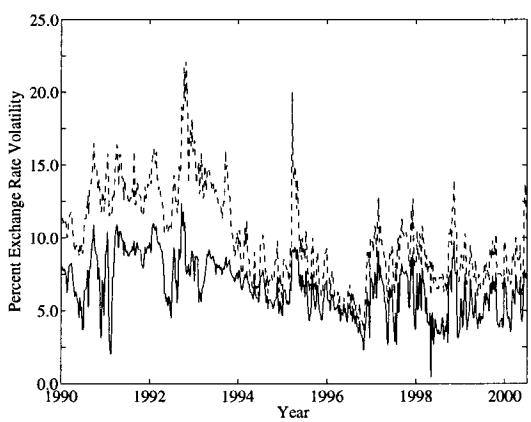


US vs. Germany

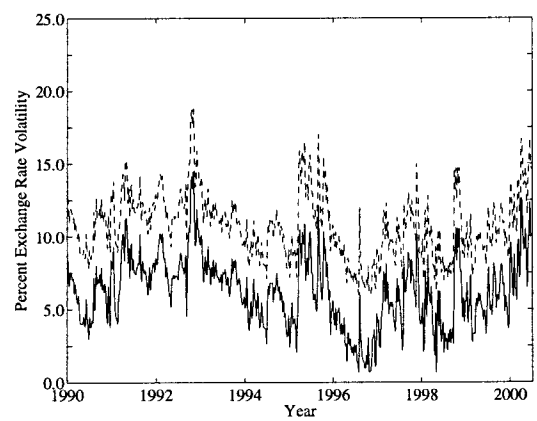


Panel B: $\psi_t = \psi(r_t - r_t^\star, e_t, v_t)$ and $\psi_t^\star = \psi^\star(r_t - r_t^\star, e_t, v_t)$

US vs. UK



US vs. Germany



THE JOHN E. ANDERSON
GRADUATE
SCHOOL OF MANAGEMENT
AT UCLA

110 WESTWOOD PLAZA
BOX 951481
LOS ANGELES, CA 90095-1481
WEB: http://www.anderson.ucla.edu/