TESTING FORWARD EXCHANGE RATE UNBIASEDNESS EFFICIENTLY: A SEMIPARAMETRIC APPROACH

Douglas J. Hodgson*, Oliver Linton†, and Keith Vorkink‡

University of Quebec at Montreal, London School of Economics, and Brigham Young University

December 3, 2003

Abstract

We apply semiparametric efficient estimation procedures for a seemingly unrelated regression model where the multivariate error density is elliptically symmetric to study the efficiency of the foreign exchange market. We consider both cointegrating regressions and standard stationary regressions. The elliptical symmetry assumption allows us to avoid the curse of dimensionality problem that typically arises in multivariate semiparametric estimation procedures, because the multivariate elliptically symmetric density function can be written as a function of a scalar transformation of the observed multivariate data.

*Address for correspondence: Douglas J. Hodgson, Department of Economics, University of Quebec at Montreal, P.O. Box 8888, Downtown Station, Montreal, PQ, Canada, H3C 3P8 ; hodgson.douglas-james@uqam.ca.

†Oliver Linton, Department of Economics, London School of Economics, Houghton Street, London WC2A 2AE, United Kingdom; lintono@lse.ac.uk.

‡Marriott School of Management, Brigham Young University, Provo, UT, USA, 84602; keith_vorkink@byu.edu.

TESTING FORWARD EXCHANGE RATE UNBIASEDNESS EFFICIENTLY: A SEMIPARAMETRIC APPROACH

JEL C22; Keywords - Forward exchange market, time series econometrics, nonparametric statistics

1 Introduction

There is a substantial empirical literature investigating both of the two distinct versions of what is called the forward exchange unbiasedness hypothesis, according to which forward exchange rates represent unbiased forecasts of future spot exchange rates. The basic procedure is to regress the spot rate (or its first difference) on the lagged future rate (or the lagged forward premium). [In fact, usually the analysis is carried out in terms of logarithms of the exchange rates]. The two approaches to formulating the unbiasedness hypothesis are complementary and have different interpretations.[1] The levels regression is a cointegrating one in which the long-run relationship between spot and forward rates is being characterized. The regression of first differences on lagged premia is a conventional stationary regression and characterizes the short-run dynamics in the foreign exchange market. Both formulations have received considerable attention from empirical researchers.

Regardless of how the unbiasedness hypothesis is posed, it is often the case that we will have data on several currencies and will wish to test the unbiasedness hypothesis for all of them. This implies the estimation of a number of equations equal to the number of currencies for which we have data. Several investigators have recognized the fact that it may be quite beneficial to estimate these equations together in a system, rather than estimating each one separately. The reasoning is that the integration of world financial markets, as well as the fact that most exchange rates are measured in terms of a common currency, viz., the U.S. dollar, both imply that the disturbances to the equations for the different countries will be correlated, so that systems estimation using Zellner's (1962) feasible GLS estimator for seemingly unrelated regressions (SUR) should produce more efficient estimates and more precise tests than would the equation-by-equation application of OLS.

The importance of efficient estimation has been well-recognized by empirical researchers in this field. In testing the levels formulation of the unbiasedness hypothesis, the studies that have estimated an SUR system of cointegrating regressions include Bailey, Baillie, and McMahon (1984), Barnhart and Szakmary (1991), and Evans and Lewis (1995). In the stationary first differences formulation, SUR techniques have been employed by, for example, Bilson (1981), Fama (1984), Cornell (1989), and Barnhart and Szakmary (1991). In related multi-country analyses of forward exchange pricing, Levine (1989 and 1991) has employed three-stage least squares, taking advantage of the correlation across currencies in the context of a simultaneous equations model. Many of these authors find significant changes in their results when the correlation across currencies is accounted for. Although

[1] The relationship between the two approaches is discussed in detail by Hakkio and Rush (1989) and Barnhart and Szakmary (1991).

non-normality is a pervasive characteristic of exchange rate data, only Bilson (1981) among the authors listed above attempts to gauge its affects on his results. He finds quite significant effects.

In taking account of the efficiency gains obtainable through the exploitation of the correlation structure of the errors, few researchers have recognized that significant additional efficiency gains are possible by exploiting the thick tails and multivariate non-normality of these errors' density functions[2]. We overcome this shortcoming in the existing empirical literature by adaptively estimating multivariate forward unbiasedness models. To do so, we make use of the adaptive estimator of stationary SUR models under elliptical symmetry developed by Hodgson, Linton, and Vorkink (2002) to test the first differences version of the hypothesis. In addition, to test the levels version of the hypothesis, the present paper extends the analysis of Hodgson, Linton, and Vorkink (2002) to allow for cointegrating regressions. We also implement a test of elliptical symmetry due to Beran (1979).

The adaptive estimators applied in this paper allow for an error density of unknown form. To overcome the curse of dimensionality, we focus on the restriction that the multivariate density is elliptically symmetric.[3] Elliptical symmetry is important for a number of reasons. First, it allows for leptokurtic marginals and hence is more consistent with commonly observed data distributions. Second, Chamberlain (1983) showed that a necessary and sufficient condition for mean-variance utility functions, and hence two parameter fund separation, is that the return distribution be elliptically symmetric. Similar semiparametric models have been explored previously in Bickel (1982), Jeganathan (1995) and Hodgson (1998). These authors defined adaptive estimators of the identifiable parameters in various regression models. However, their proposed estimates do not exploit the dimensionality reduction implied by elliptical symmetry and consequently suffer serious "small sample" costs. What is required here is estimation of a multidimensional density function and its first derivative. See Silverman (1986, page 94) for a dramatic illustration of the effects of dimensionality on estimating a normal density at the origin. Although the semiparametric theory says that asymptotically these effects disappear when the properties of the parameter estimates are being considered, in even quite large samples they do not.

In Section 2, we introduce the two versions of the unbiasedness hypothesis and the corresponding cointegrated and non-cointegrated SUR econometric models, along with a general modeling strategy that nests the two models. In Section 3, we provide a formula for computing the adaptive estimator developed by Hodgson, Linton, and Vorkink (2002) for stationary models and describe the extension

[2]The effects of thick tails in univariate tests of forward exchange market unbiasedness have been investigated by, for example, Steigerwald (1992), Phillips, McFarland, and McMahon (1996), and Hodgson (1998a, 1999).

[3]See also Fernández, Osiewalski, and Steel (1995) for some interesting generalizations of elliptical symmetry.

of this estimator to cointegrated models. Section 4 discusses issues involved in applying Beran's (1979) test for elliptical symmetry, as well as the issue of bandwidth selection in our model, and Section 5 reports the results of our exchange rate analysis. We use $\|A\| = \left(\mathrm{tr}A^T A\right)^{1/2}$ to denote the Euclidean norm of a vector or matrix A, while $\xrightarrow{P}$ denotes convergence in probability and $\Rightarrow$ signifies weak convergence of probability measures. We say that $X \sim MN(0, V)$ when X is mixed normal with (possibly) random covariance matrix V.

2 Econometric Models

In this section, we introduce the basic regression models through which the two versions of the forward unbiasedness hypothesis are generally implemented. We will first consider the stationary version of the hypothesis and associated econometric model, followed by a discussion of the cointegrated version of the hypothesis with the associated econometric model. We will then proceed to formulate a general econometric model which incorporates the two hypotheses within a unified econometric framework.

Suppose we observe a sequence of (logged) spot exchange rates $\{s_t^i\}$, $t = 1, \ldots, n+1$, $i = 1, \ldots, m$ and (logged) one-period ahead forward exchange rates $\{f_t^i\}$, $t = 1, \ldots, n$, $i = 1, \ldots, m$. In this formulation, s_t^i is the log spot exchange rate between the currency of country i and some control currency, such as the U.S. dollar. Suppose we have data for m different currencies, indexed by i, and for each currency we have observations for n consecutive time periods, indexed by t. The forward rate f_t^i is the log of the price paid in period t for the delivery of a unit of currency i in period $t+1$. For instance, if we have a sequence of monthly data, then f_t^i would denote the one-month forward rate prevailing at period t. More specific details about the data actually used in our empirical study are provided in Section 6.

What we have referred to as the stationary version of the unbiasedness hypothesis states that the forward exchange premium $f_t^i - s_t^i$ provides an unbiased forecast of the change in the exchange rate over the upcoming time period, i.e. that

$$E_t\left[s_{t+1}^i - s_t^i\right] = f_t^i - s_t^i,$$

where E_t is the conditional expectation formed on the basis of all information available as of time period t. This hypothesis can be tested empirically using estimates of the following set of m regression equations:

$$s_{t+1}^i - s_t^i = \alpha^i + \beta^i(f_t^i - s_t^i) + u_{i,t+1} \quad ; \quad t = 1, \ldots, n, \ i = 1, \ldots, m. \tag{1}$$

Within this framework, the unbiasedness hypothesis can be stated as follows:

$$\mathbf{H}_0: \ \alpha^i = 0, \ \beta^i = 1, \quad i = 1, \ldots, m \tag{1a}$$

versus the general alternative. Under the null, the forward rate provides an unbiased prediction of future spot rates and the market is informationally efficient. This hypothesis has been tested many times before; see Engel (1996) for a review. Our test is multivariate as we estimate the above regression equations as a seemingly unrelated regression taking account of the comovement (across i) that we expect to find in $u_{i,t+1}$.

The levels, or cointegrated, version of the hypothesis enquires as to whether or not the current forward rate is an unbiased predictor of the next period's spot rate, which we can write as follows:

$$E_t\left[s^i_{t+1}\right] = f^i_t.$$

The corresponding regressions we estimate are of the following form:

$$s^i_{t+1} = \alpha^i + \beta^i f^i_t + u_{i,t+1}. \tag{2}$$

We are then interested in testing the hypothesis

$$\mathbf{H}_0: \ \alpha^i = 0, \ \beta^i = 1, \quad i = 1, \ldots, m \tag{3}$$

versus the general alternative.

In both formulations of the hypothesis, we have a system of m regression equations to be estimated. Standard single-equation estimation methods such as ordinary least squares (OLS) can be used to estimate the parameters of the model and form valid asymptotically chi-squared Wald test statistics for both versions of the hypothesis. However, as discussed in the introduction, the single-equation approach can entail a substantial loss in estimation efficiency and testing power relative to an estimation strategy such as generalized least squares (GLS) which estimates all m equations jointly as a seemingly unrelated regressions (SUR) system. We claim that further efficiency gains may be obtainable by accounting for the possible presence of non-normality in the disturbances to the SUR system.

The two regression models described above can be nested within a more general framework of multivariate regressions. Consider the m-equation seemingly unrelated regression model

$$y_t = \alpha + x_t\beta + u_t := w_t\theta + u_t, \quad t = 1, \ldots, n, \tag{4}$$

where $y_t \in \mathbb{R}^m$, $\alpha \in \mathbb{R}^m$,

$$x_t = \begin{bmatrix} x_{1t}^T & & & 0 \\ & x_{2t}^T & & \\ & & \ddots & \\ 0 & & & x_{mt}^T \end{bmatrix}, \quad \beta = \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_m \end{bmatrix}, \quad u_t = \begin{pmatrix} u_{1t} \\ \vdots \\ u_{mt} \end{pmatrix},$$

$w_t = [I_m \; x_t]$, $x_{it}^T \in \mathbb{R}^{k_i}$ and $\beta_i \in \mathbb{R}^{k_i}$ for every $i = 1, \ldots, m$, the full parameter vector is $\theta = \left[\alpha^T, \beta^T\right]^T \in R^{m+k}$, where $k = k_1 + \cdots + k_m$, and $u_t \in \mathbb{R}^m$ are i.i.d., mean zero innovations with $E(u_t u_t^T) = \Sigma_u$ and with density $p(u)$. Here, the regressors x_t may be either integrated of order one (I(1)) or stationary and ergodic. In either case, we assume that x_t and u_t are independent. When the regressors are I(1), each of the m regressions is cointegrating and the framework is suitable for the analysis of the second form of the unbiasedness hypothesis stated above. When the regressors are stationary, the regressions are standard and are suitable for the analysis of the first version of the unbiasedness hypothesis.

We consider two different assumptions about p. Firstly, that p is unrestricted. Secondly, we restrict p to be elliptically symmetric.

Definition. *An m-dimensional density function $p(u)$ is elliptically symmetric if it can be written in the form $(\det \Sigma)^{-1/2} g(u^T \Sigma^{-1} u)$ for some scalar density generating function $g(\cdot)$ and matrix Σ.*

The practical content of the elliptical symmetry restriction arises from the fact that the function g has only a scalar argument.

Assuming that p were known, the log-likelihood for the data would be

$$L_n(\theta) = \sum_{t=1}^{n} \ln p(y_t - w_t\theta),$$

and estimation of θ proceeds by maximizing $L_n(\theta)$. We define the weighting matrix δ_n, where $\delta_n = n^{-1/2} I_{m+k}$ if x_t are stationary and $\delta_n = \operatorname{diag}\left[n^{-1/2} I_m, n^{-1} I_k\right]$ if x_t are integrated. These structures for δ_n are associated with the fact that the rate of consistency of estimators in non-cointegrated models is $n^{1/2}$, whereas in cointegrating regressions it is $n^{1/2}$ for intercept parameters and n for slope parameters. One estimation strategy here is to use a two-step Newton-Raphson estimator $\overline{\theta}$ starting from a preliminary δ_n^{-1}-consistent estimator $\widehat{\theta}$, obtained from the Gaussian likelihood. Under general conditions, this will be first order asymptotically equivalent to the MLE, i.e.,

$$\delta_n^{-1}(\overline{\theta} - \theta_0) \Rightarrow MN(0, \mathcal{I}^{-1}), \tag{4}$$

where the asymptotic information matrix $\mathcal{I}$ is such that $\delta_n \left(\partial^2 L_n \left(\theta_0\right) / \partial\theta\partial\theta'\right) \delta_n \Rightarrow \mathcal{I}$. In order to derive an expression for $\mathcal{I}$, we define $\varphi(u) = \frac{\partial p(u)/\partial u}{p(u)}$, the m-dimensional score vector of p, and ${}_p = \int \varphi(u)\varphi(u)^T p(u)du$, the information matrix of p. For the stationary model, the asymptotic information matrix is

$$\mathcal{I} = \begin{bmatrix} {}_p & E\left[\ {}_p x_t\right] \\ E\left[x_t^T\ {}_p\right] & E\left[x_t^T\ {}_p x_t\right] \end{bmatrix},$$

while for the cointegrated model, it is

$$\mathcal{I} = \begin{bmatrix} {}_p & {}_p \int_0^1 M(r)dr \\ \int_0^1 M(r)^T dr\ {}_p & \int_0^1 M(r)^T\ {}_p M(r)dr \end{bmatrix},$$

where

$$M(r) = \begin{bmatrix} M_1^T(r) & & & 0 \\ & M_2^T(r) & & \\ & & \ddots & \\ 0 & & & M_m^T(r) \end{bmatrix}$$

and $M_i(r)$ is a k_i-dimensional Brownian motion with covariance matrix equal to the long run covariance matrix of Δx_{it}, for every $i = 1, \ldots, m$. Note that in the case of cointegration, $\mathcal{I}$ is random, hence the mixed normal limit theory.

The estimation strategy employed in the present paper follows Hodgson, Linton, and Vorkink (2002) in that we also use a Newton-Raphson iterative approach to estimation but must replace the unknown density p by a nonparametric estimator; thus our adaptive estimator $\widetilde{\theta}$ will have the form

$$\widetilde{\theta} = \widehat{\theta} + \delta_n \widehat{\mathcal{I}}_n^{-1}(\widehat{\theta})\widehat{\Delta}_n(\widehat{\theta}), \tag{5}$$

where $\widehat{\Delta}_n$ and $\widehat{\mathcal{I}}_n$ are estimates of the first and second standardized derivatives of L_n respectively. Their computation is described in Section 3 below. In particular,

$$\widehat{\Delta}_n(\widehat{\theta}) = -\delta_n \sum_{t=1}^{n} w_t' \widehat{\varphi}_t(\widehat{u}_t),$$

where $\widehat{\varphi}_t(\widehat{u}_t)$ is a consistent estimator of the m-dimensional score vector $\varphi(u_t)$, while $\widehat{u}_t = y_t - w_t\widehat{\theta}$. The standard approach to this problem is to use multivariate kernel estimates $\widehat{p}$ and $\widehat{p}'$ to construct $\widehat{\varphi}$, with some observations possibly being trimmed, see Bickel (1982). Unfortunately, if m is large such estimates will have poor performance due to the curse of dimensionality, see Härdle and Linton (1994). We follow Hodgson, Linton, and Vorkink (2002) in using a construction of $\widehat{\varphi}_t(.)$ that takes advantage of our elliptical symmetry assumption and employs only one-dimensional smoothing operations.[4] We should note that Hodgson, Linton, and Vorkink (2002) only consider estimation of the stationary SUR model, with the extension of the estimator to cointegrating regressions in the present paper being new.

3 Estimation

Our argument in the preceding section implies that the finite sample performance of an adaptive estimator can be significantly improved if, in computing a nonparametric score estimator $\widehat{\varphi}$, we use a direct kernel estimate of the density of the univariate random variable $v = u^T\Sigma^{-1}u$ to *indirectly* estimate the density of the m-vector u, rather than directly estimating the latter with a multivariate kernel. The adaptive estimator described below does indeed use such an indirect approach to estimating p'/p, but does so at *two* removes rather than one. In other words, our estimate of the density of v is itself an indirect estimate, derived from a kernel estimate of the univariate density of the transformed random variable $z = \tau(v)$, where $\tau(\cdot)$ is some transformation. Of course, the identity is a valid transformation, so that direct estimation of the density of v is allowed by our theory; however, certain other transformations may yield estimators with better finite sample performance. Hodgson, Linton, and Vorkink (2002) consider a general class of transformations given by $\tau(v;\zeta) = \left(v^{\zeta} - 1\right)/\zeta$, where selection of the parameter ζ is left to the discretion of the investigator and is discussed by Hodgson, Linton, and Vorkink (2002).

Before introducing our estimator, we must introduce some preliminary notation. Recall that

$$p(u) = (\det \Sigma)^{-1/2} g(u^T\Sigma^{-1}u)$$

for some function g and matrix Σ. Note that in this formulation, the magnitude of the matrix Σ is left indeterminate, as multiplying it by a constant can be accommodated by changing the definition

[4] As shown in Stute and Werner (1991) these procedures ensure density estimators whose pointwise rate of convergence is the one-dimensional rate.

of $g(\cdot)$ to absorb the constant. This matrix is determined up to multiplication by a scalar, however, and will generally be proportional to the covariance matrix $\Sigma_u = E\left[uu^T\right]$ and to the inverse of the information matrix $\quad_p = E\left[\varphi(u)\varphi(u)^T\right]$. Following Hodgson, Linton, and Vorkink (2002), we tie down the value of Σ by defining it such that $\det(\Sigma) = 1$. Note that we do this without any loss of generality. It follows that $\Sigma_u = c\Sigma$, where $c = (\det \Sigma_u)^{1/m}$, i.e., $\Sigma = \Sigma_u/(\det \Sigma_u)^{1/m}$. To simplify our analysis we define the spherically symmetric m-dimensional random variable $\varepsilon = \Sigma^{-1/2}u$, which is just the standardized innovation vector. Note that its density function $f(\varepsilon)$ is directly proportional to our innovation density $p(u)$, as shown by the following relation:

$$f(\varepsilon) = p(u)\left|\frac{du}{d\varepsilon}\right| = g(\varepsilon^T\varepsilon).$$

Defining the transformed random variable $z \equiv \tau(\varepsilon^T\varepsilon) \equiv \tau(v)$, let its density function be denoted by $\gamma(z)$.

The construction of the estimator described below is motivated by the fact that we can derive a mathematical relationship between the univariate density $\gamma(z)$ and the multivariate density $p(u)$, so that a nonparametric estimate of the latter can be derived from a nonparametric estimate of the former. The following relationships will be useful to keep in mind when considering the computation of the estimator described below. We begin by considering the transformation $z = \tau(\varepsilon^T\varepsilon)$. We are particularly interested in deriving an expression for its density $\gamma(z)$ and characterizing the relationship between $\gamma(z)$ and $f(\varepsilon)$ (and hence between $\gamma(z)$ and $p(u)$). Suppose that the m-vectors ε_t are i.i.d. from the density $f(\varepsilon) = g(\varepsilon^T\varepsilon) \equiv g(v)$ where $v = \varepsilon^T\varepsilon$. From Muirhead (1982), the density of v, which we shall denote $h(v)$, is

$$h(v) = c_m v^{m/2-1} g(v),$$

where $c_m = \pi^{m/2}/\Gamma(m/2)$. By Theorem 2.1.2 of Casella and Berger (1990) we have

$$\gamma(z) = h(\tau^{-1}(z)) \cdot \left|\frac{\partial\tau^{-1}(z)}{\partial z}\right| = c_m\left[\tau^{-1}(z)\right]^{m/2-1} g(\tau^{-1}(z)) \cdot J_\tau(z),$$

where $J_\tau(z) = |\partial\tau^{-1}(z)/\partial z|$. Thus, $g(v) = c_m^{-1} J_\tau^{-1}\{\tau(v)\}\, v^{1-m/2}\gamma\{\tau(v)\}$. This gives us our desired expression for $g(v)$ - and hence for $f(\varepsilon)$ and $p(u)$ - in terms of $\gamma(z)$.

The formula for an adaptive estimator given in (5) above presupposed the existence of consistent score and information estimators $\widehat{\varphi}_t$ and $\widehat{\mathcal{I}}_n$. With the notation developed in the preceding section, we can now provide procedures for computing these consistent nonparametric estimates. In particular, we show how we can use direct kernel estimates of $\gamma(z)$ to indirectly obtain consistent estimates of the score and information of p. This construction is due to Hodgson, Linton, and Vorkink (2002), following Bickel (1982), and is justified theoretically in those papers.

Our algorithm for estimating φ and $\mathcal{I}$ proceeds according to the following steps:

1. First obtain $\widehat{\theta}$ and define the residuals $\{\widehat{u}_t\}_{t=1}^n$ and the standardized residuals $\{\widehat{\varepsilon}_t\}_{t=1}^n$, where $\widehat{\varepsilon}_t = \widehat{\Sigma}^{-1/2}\widehat{u}_t$, $\widehat{\Sigma} = \widehat{c}^{-1}\widehat{\Sigma}_u$, $\widehat{\Sigma}_u = n^{-1}\sum_{t=1}^n \widehat{u}_t\widehat{u}_t^T$, and $\widehat{c} = \left[\det \widehat{\Sigma}_u\right]^{1/m}$. Then compute the transformed sequence $\{\widehat{z}_t\}_{t=1}^n$, where $\widehat{z}_t = \tau(\widehat{v}_t)$ with $\widehat{v}_t = \widehat{\varepsilon}_t^T\widehat{\varepsilon}_t$.

2. Denoting by $K_{h_n}(\cdot)$ a kernel with bandwidth h_n, form the following estimates of the density of $\widehat{z}_t$ and its first derivative:

$$\widehat{\gamma}_t(z) = \frac{1}{n-1}\sum_{\substack{s=1\\ s\neq t}}^n K_{h_n}(z-\widehat{z}_s) \quad ; \quad \widehat{\gamma}_t'(z) = \frac{1}{n-1}\sum_{\substack{s=1\\ s\neq t}}^n K_{h_n}'(z-\widehat{z}_s).$$

3. Introduce the following trimming conditions: (i) $\widehat{\gamma}_t(\widehat{z}_t) \geq d_n$; (ii) $|\widehat{z}_t| \leq e_n$; (iii) $|\lambda(\widehat{z}_t)| \leq b_n$; (iv) $|\rho^{1/2}(\widehat{z}_t)\widehat{\gamma}_t'(\widehat{z}_t)| \leq c_n\widehat{\gamma}_t(\widehat{z}_t)$, where $\rho(z) = v\tau'(v)J_\tau^{-1}(z)$ [recall that $v = \tau^{-1}(z)$] and $\lambda(z) = (d/dz)^{-1}\rho^{1/2}(z)$.[5] Then estimate the score and information of $p(u)$ as follows:

$$\widehat{\varphi}_t(\widehat{u}_t) = \begin{cases} \widehat{\Sigma}^{-1/2}\widehat{\varepsilon}_t\left[s(\widehat{v}_t) + \tau'(\widehat{v}_t)\frac{\widehat{\gamma}_t'}{\widehat{\gamma}_t}(\widehat{z}_t)\right] & \text{if (i)} - \text{(iv) all hold} \\ 0 & \text{otherwise,} \end{cases}$$

where $s(v) = (1-m/2)v^{-1} - \frac{J_\tau'}{J_\tau}\{\tau(v)\}\,\tau'(v)$, and

$$\widehat{\ }_p = \frac{1}{n}\sum_{t=1}^n \widehat{\varphi}_t(\widehat{u}_t)\widehat{\varphi}_t(\widehat{u}_t)^T.$$

4. Then define the score and information estimators for the model as

$$\widehat{\Delta}_n(\widehat{\theta}) = -\delta_n\sum_{t=1}^n w_t'\widehat{\varphi}_t(\widehat{u}_t) \quad ; \quad \widehat{\mathcal{I}}_n(\widehat{\theta}) = \delta_n\sum_{t=1}^n w_t^T\widehat{\ }_p w_t\delta_n. \tag{6}$$

[5] These trimming conditions ensure consistency of our score estimator when a Gaussian kernel is being used, i.e. when K_{h_n} is a Gaussian kernel. For other kernels often employed in the literature (e.g. Schick's (1987) logistic kernel and the bi-quartic kernel used in the applications reported below), the necessary trimming conditions, if they differed at all from these, would be less stringent, so that these conditions will still be sufficient for consistency but may not be necessary. Simulation work reported by Hsieh and Manski (1987) and Hodgson (1998) finds that, for a Gaussian kernel, the adaptive point estimate is not very sensitive to variation in the value of the trimming parameters, and that good results are obtained in practice when we trim as little as 1% of the observations.

We can state the following Proposition, which is a straightforward extension of Theorem 1 of Hodgson, Linton, and Vorkink (2002) to our model.

PROPOSITION 1: *Suppose that* $_p$ *is finite and positive definite; that* $\int_0^\infty v^{m/2} s(v)^2 g(v) dv < \infty$; *that the error distribution is absolutely continuous with respect to Lebesgue measure with Lebesgue density* $p(u)$, *that the regressors* x_t *are strictly exogenous, and that the constants in (i)-(iv) satisfy* $c_n \to \infty$, $e_n \to \infty$, $b_n \to \infty$, $h_n \to 0$, $d_n \to 0$, $h_n c_n \to 0$, $e_n h_n^{-3} = o(n)$, *and* $b_n h_n^{-3} = o(n)$. *Then,*

$$\delta_n^{-1}(\widetilde{\theta} - \theta) \Rightarrow MN(0, \mathcal{I}^{-1}), \tag{7}$$

i.e., the estimator $\widetilde{\theta}$ *is adaptive.*

REMARKS. (a) The moment condition $\int_0^\infty v^{m/2} s(v)^2 g(v) dv < \infty$ will depend on the transformation $\tau(\cdot)$ which we use and can be more or less restrictive for different selections of $\tau(\cdot)$. For example, when the transformation is $\tau(v;\zeta) = \left(v^\zeta - 1\right)/\zeta$, with either $\zeta = 0$, $\zeta = 1$, or $\zeta = 1/2m$, the condition implies that $E[(\varepsilon^T \varepsilon)^{m/2-2}] < \infty$. However, when $\zeta = m/2$, there is no restriction on the moments of u.

(b) Note that the information matrix estimator $\widehat{\mathcal{I}}_n(\widehat{\theta})$ defined in (6) is a consistent estimator of the asymptotic covariance matrix, so that $\widehat{\mathcal{I}}_n(\widehat{\theta}) - \mathcal{I} = o_p(1)$. This result is true even for cointegrated models, in which case $\mathcal{I}$ is random. We can therefore use $\widehat{\mathcal{I}}_n(\widehat{\theta})$ in the construction of t-ratios and Wald statistics which will have respective standard normal and chi-squared asymptotic distributions. Let θ_ℓ and $\widetilde{\theta}_\ell$ be the ℓ^{th} elements of the θ and $\widetilde{\theta}$ vectors, respectively. Now suppose we wish to test the null hypothesis that $\theta_\ell = r$, where r is some constant. Then we can compute the usual t-ratio, as follows:

$$\frac{(\delta_n^{-1})_{\ell\ell}\left(\widetilde{\theta}_\ell - r\right)}{\sqrt{\left(\widehat{\mathcal{I}}_n^{-1}(\widehat{\theta})\right)_{\ell\ell}}} \xrightarrow{d} N(0,1)$$

under the null, where $(\delta_n^{-1})_{\ell\ell}$ and $\left(\widehat{\mathcal{I}}_n^{-1}(\widehat{\theta})\right)_{\ell\ell}$ are the ℓ^{th} elements along the diagonals of δ_n^{-1} and $\widehat{\mathcal{I}}_n^{-1}(\widehat{\theta})$, respectively. If we want to test the joint hypothesis $\theta = r$ for the entire vector θ, where r is now a known $(m+k)-$vector of constants, we can compute the Wald statistic

$$\left[\delta_n^{-1}\left(\widetilde{\theta} - r\right)\right]' \widehat{\mathcal{I}}_n(\widehat{\theta}) \left[\delta_n^{-1}\left(\widetilde{\theta} - r\right)\right] \xrightarrow{d} \chi^2_{m+k}.$$

Note that these convergence results will hold regardless of whether the model is stationary or cointegrated.

(c) It is natural to ask how the present estimator will behave if the thick tails in the unconditional density of the errors are induced by some sort of conditional dependence, such as a multivariate GARCH model. A related question has been addressed in Hodgson (2000) within the context of adaptively estimating univariate time series regression models, and the following conjectures are based on Hodgson's (2000) findings. It should be possible to extend these findings to obtain a useful robustness result for our estimator in the case where the error process $\{u_t\}$ is uncorrelated but not necessarily independent over time, and has an *unconditional* density which is elliptically symmetric. This would happen, for example, if the errors followed a multivariate GARCH process, had a conditional density which was elliptically symmetric, and had a conditional covariance matrix whose magnitude changed over time but whose covariance structure remained unchanged. In any event, if the unconditional density is elliptically symmetric, then the nonparametric score and information estimators $\widehat{\varphi}$ and $\widehat{\ }$ described above and used in our computation of the adaptive estimator should still consistently estimate the score and information of the unconditional density of the errors. Our one-step estimator will then have the same asymptotic distribution as the one-step iterative pseudo-MLE based on the true unconditional density of the errors. When the regressors are strictly exogenous, as we have assumed above, then the resulting estimator will have an asymptotic distribution which is identical to that which it would have if the i.i.d. assumption on the errors was correct. In other words, the distribution depends only on the unconditional density of the errors and is completely invariant to the presence of conditional heteroskedasticity. Furthermore, the standard error estimates and test statistics described in the preceding remark will be robust to the presence of conditional heteroskedasticity. When the strict exogeneity assumption on the regressors is relaxed, this robustness property no longer holds. It is still true that our one-step semiparametric estimator will have the same distribution as the one-step fully parametric estimator based on the true unconditional density, but it will now be the case that the latter estimator's asymptotic covariance matrix will have the "sandwich" structure characteristic of pseudo-MLE's in misspecified models (cf. White (1982)). To construct robust standard errors in this case, we would require a consistent nonparametric estimator of the Hessian of the innovation density, since both the Hessian and OPG versions of the information will enter the asymptotic covariance matrix. The derivation of such a consistent Hessian estimator has not yet been considered in the literature and is a topic for future research.

(d) Some technical issues relating to the empirical implementation of our estimator and of a test for elliptical symmetry are discussed in the Appendix.

4 Forward Market Unbiasedness Tests

Like many economic theories, spot-futures parity does not purport a specific forecast horizon to which the theory applies. We will use data sets of two different frequencies of spot and future exchange rate in our empirical tests: 1) daily data ranging from January 1998 through December 2001; 2) weekly data ranging from January 1993 through December 2001. We collect spot and futures rates for three currencies (each expressed in terms of U.S dollars) for each of these frequencies: the Japanese yen (JPY/USD), the British pound (GBP/USD), and the Canadian dollar (CAD/USD). We obtain the data for both of the spot and futures rates from Bloomberg Inc. The futures rates provided by Bloomberg are taken from futures quotes on the Chicago Mercantile Exchange and Bloomberg's spot rates are New York Composite quotes, or average rates across the large institutional currency traders. We are careful to match the horizon of the futures with the sampling frequency of the data so that our residuals should be uncorrelated through time, i.e.- the daily data include futures prices with one-day horizons while the weekly data include futures prices with one-week horizons. We do find a number of dates, for both the daily and the weekly data, where the futures price is missing and as a result we delete these dates from the data set. Our final daily (weekly) frequency data set has 806 (468) observations.

Tables I and II provide the summary statistics for the two data sets. We see that the Augmented Dickey-Fuller (ADF) test fails to reject a unit root for all of the logged spot and forward rates. However, when these rates are converted to percent changes for the stationary model, the ADF test rejects the presence of a unit root for all of the series at a .05 level with most rejections at the .01 level.[6] Table III reports results of Box-Pierce tests applied to the OLS residuals from each of our regressions. There is generally little evidence of serial correlation, with the exception of the levels regression for GBP with weekly data. Multivariate distributional tests, as applied to the OLS residuals, are reported in Table IV. The Beran test statistic, S_n, reported in Panel B, sets $k = l = 7$, with sensitivity analysis on these choices finding that the statistic varies little for small changes in these values. The Mardia (1970) kurtosis test finds evidence of significant excess kurtosis in all series, whereas the Beran (1979) statistic does not lead to rejections of the null of elliptical symmetry at the

[6] We repeated these tests using an adaptive unit root test developed by Beelders (1998) and came to identical conclusions.

5% level for any of the series (although rejections at the 10% level would occur for the daily data).

Tables V and VI provide the estimation results, for the levels and differences regressions, respectively, while Wald statistics of the unbiasedness null hypothesis as stated in (1a) and (3) are reported in Table VII. We note that the adaptive estimates are computed using a Gaussian kernel with Schuster's (1985) correction and the Box-Cox transformation $z = \tau(v) = \left(v^{\zeta} - 1\right)/\zeta$, with $\zeta = 1/2m$.[7]

4.1 Cointegrated Model: Results

Our estimates of the cointegrated model and associated unbiasedness test statistics are reported in Table V, and in the second half of Table VII, respectively. Before analyzing our results, we will consider some of the previous evidence that has been obtained in recent years, highlighting in addition the evolution of the econometric methodology in this area. Note that all exchange rates are assumed to be taken with respect to the US dollar, unless stated otherwise.

Two studies completed shortly after the introduction of the concept of cointegration to econometricians are Baillie and Bollerslev (1989) and Hakkio and Rush (1989). The former study uses daily data for the U.K., Germany, Japan, France, Italy, Switzerland, and Canada for the period 1980-1985, whereas the latter considers the monthly exchange rates of the U.K and Germany for 1975-86. Both consider 30-day forward rates and estimate the levels cointegrating regression by OLS, but, due the lack of availability of distribution theory for cointegration estimators at the time, neither compute standard errors or formally test the unbiasedness hypothesis. Their point estimates are fairly close to those suggested by the hypothesis, however, except that Baillie and Bollerslev (1989) find intercept and slope estimates for Japan to be -0.83 and 0.85, respectively. Barnhart and Szakmary (1991) exploit the efficiency gains in estimation that may be available in modelling cross-currency correlations by using a seemingly unrelated regression estimator in a system of four currencies (U.K., Germany, Japan, Canada), using monthly data and 30-day forward rates for 1974-88. They accept the unbiasedness hypothesis for all four currencies.

The estimator developed in the present paper combines Barnhart and Szakmary's (1991) systems approach with the robust and semiparametric estimators developed in a number of studies in the late 90's to address the issue of non-normailty in exchange rate data. Phillips, MacFarland, and McMahon (1996) and Phillips and MacFarland (1997) apply the robust fully modified least absolute deviations

[7] See Hodgson, Linton, and Vorkink (2002) for a motivation for the choice of the transformation function.

estimator (FM-LAD) of Phillips (1995) to study, in the first case, the daily exchange rates and one-month forward rates of Belgium, France, Italy, and U.S.A. (vis-à-vis the U.K.) for 1922-25 and, in the second case, the Australian dollar for 1984-91, considering both 30- and 90-day forward rates. In both papers, comparison is made of the inferences obtained using FM-LAD as opposed to the non-robust (to non-normality) FM-OLS estimator of Phillips and Hansen (1990). Phillips, MacFarland, and McMahon (1996) strongly reject the unbiasedness hypothesis for all countries except the U.S., regardless of estimation methodology, whereas Phillips and MacFarland (1997) accpet the hypothesis using FM-OLS but strongly reject it with FM-LAD. The studies of Hodgson (1998a,1999) apply the fully efficient semiparametric adaptive estimators developed in Hodgson (1998a,b) for cointegrating regressions and error correction models, respectively, using daily data and 3-month forward rates for the Canadian dollar for 1990-93. Results using the semiparametric estimators are compared with those obtained using FM-OLS and Johansen's (1988) error correction model estimator, and are generally found to lend stronger support to the unbiasedness hypothesis.

Note that none of the estimators used in these four latter studies exploit cross-currency dependencies in the manner of a Gaussian SUR estimator, whereas the present paper does so while still attempting to account optimally for the possible presence of non-normality. In addition, we depart from the papers listed above in using a more recent data set and in considering much shorter forward maturities. As can be seen from Tables V and VII, we generally obtain very strong support for the unbiasedness hypothesis, regardless of currency, estimator, forward horizon, or frequency of observation. This is broadly consistent with the exisiting literature, although more unambiguously supportive of the hypothesis than much of it.

4.2 Stationary Model: Results

The existing literature analyzing this model is much vaster and stretches back farther in time than the corresponding literature for the cointegrated model. We will therefore not attempt anything even resembling an overview of the literature (for which the reader is referred to, for example, Baillie and McMahon (1989) or Engel (1996)), but will merely reference a small handful of representative papers in order to highlight the contributions of the present one.

A number of studies have proceeded using "overlapping" data - i.e. data for which the frequency of observation is higher than the length of the futures contract (for example, the use of weekly data with a 30-day forward rate). This practice introduces a moving average autocorrelation structure to the regression disturbances in equation (1), which complicates the estimation theory. In our data, we have matched the forward horizon with the frequency of observation, so that the disturbances should be uncorrelated, which simplifies the econometric analysis and allows the application of the estimator developed in this paper. We proceed therefore to first discuss our results within the context of existing studies that use non-overlapping data. We then briefly discuss the possibilities of extending our analysis to allow for overlapping data, and consider some further extensions of our analysis in the light of some recent developments in the empirical literature.

Among the many papers that estimate (1) using non-overlapping data, we will reference here Bilson (1981), Fama (1984), and Barnhart and Szakmary (1991) as representative examples. Working with monthly data and 30-day forward rates for a collection of nine major OECD currencies over the period 1974-80, Bilson (1981), in estimating (1) by OLS, finds point estimates far from those predicted by the unbiasedness hypothesis, with estimates of the slope coefficients β generally being well below one, but fails to reject the hypothesis for most countries due to high standard errors. He subsequently groups the nine currencies into a system, which is estimated using the more efficient SUR-GLS estimator introduced by Zellner (1962), and obtains much stronger rejections of the null. Fama (1984), using a similar data set for nine countries for 1973-82, also compares OLS and SUR estimates, and although the latter produce substantially smaller standard errors, both estimators generally lead to rejections of the null, again due to slope estimates well below one. Barnhart and Szakmary (1991) obtain similar results using an SUR estimator and the data set described above. In fact, it has become something of a "stylized fact" in the literature that slope estimates are generally found to be less than one, and, in many cases, significantly negative.

As described above, we bring new evidence to bear based on the recent period covered by our data and the shorter forecast horizon. In addition, our estimator builds on the intuition of the aforementioned papers in increasing efficiency of estimation by modelling the currencies in a system, while allowing nonparametrically for the possible presence of non-normality in the data. Our results for the estimation of (1) are presented in Table VI and in the first two panels of Table VII, where we compare the results obtained using OLS and the semiparametric adaptive estimator developed here. Regarding the basic inference as reported in Table VII, we can see that the null hypothesis is actually accepted for both estimators with daily data, and is rejected with weekly data. Nevertheless, a look at the point estimates in table VI reveals that the acceptance is due mainly to the wildly inaccurate point estimates, somewhat less imprecise for the adaptive estimator. When moving from OLS to the adaptive estimates, there are huge changes in the slope estimates, from well below zero to well above unity. In the weekly data, the adaptive estimates are substantially more precise than OLS, judging by the standard errors, but both estimates yield slope estimates that are significantly less than one, and even, for certain currencies, significanlty negative. At least for the weekly data, our results are consistent with previous studies.

As mentioned above, several studies work with overlapping data. Although our methodology cannot be directly applied to such a situation, a brief consideration of previous econometric approaches may suggest extensions of our methodology that could lead to efficiency improvements for these models. Beginning with the paper of Hansen and Hodrick (1980), several investigators have modelled spot and forward rates for individual currencies as bivariate vector autoregressions (VAR's), which are then estimated by Gaussian MLE, possibly under parameter restrictions, from which inferences can be made regarding the unbiasedness hypothesis (see also, for example, Hakkio (1981) and Baillie, Lippens, and McMahon (1983)). Although questions of parameterization would probably forbid the inculsion of several currencies into a large joint VAR, there is no reason, in principle, why the individual-country bivariate VAR could not be estimated adaptively or semiparametrically efficiently, using an extension of the procedures used in this paper.

A final possibility for extensions would be in the area of fractionally integrated models of the forward premium. Baillie and Bollerslev (1994) compute Gaussian ML estimates of fractionally integrated ARFIMA models, finding evidence of fractional integration in the forward premium in a number of major currencies. Maynard and Phillips (2001) obtain similar results, and investigate their consequences for the estimation of models such as that of equation (1). As a suggestion for future work, it may be worth investigating the possiblity of efficiency gains in ARFIMA models through the specification of joint likelihoods for several currencies, and/or the specification of semiparametric likelihoods and the derivation of semiparametric efficiency bounds.

5 Appendix - Some Technical Issues

We discuss here some issues that arise in the implementation of our estimator. Subsection 1 describes Schuster's correction, a modification of our basic estimator which is undertaken to correct for the poor properties of the nonparametric density estimator in the neighborhood of the origin which is due to the fact that we do not directly estimate the density of the innovations u_t but rather estimate the density of a transformation z_t whose support is only on the positive portion of the real line (see Stute and Werner (1991), for a discussion of this problem, known as the "volcano effect"). In subsection 2, we discuss the issue of bandwidth selection, in subsections 3 and 4 we discuss issues arising in the implementation of a test for elliptical symmetry due to Beran (1979), while subsection 5 considers a method for correcting bias in our nonparametric information matrix estimator.

5.1 Schuster's correction

The construction of $\widehat{\varphi}$ imposing elliptical symmetry uses one dimensional kernel estimates of the transformed variable z. For several transformations $z = \tau\left(\varepsilon^T\varepsilon\right)$, the support will have the restriction $z \geq 0$. This additional information is not incorporated in the standard Parzen-Rosenblatt kernel estimator, $f_n(z) = n^{-1}h_n^{-1}\sum_{i=1}^{n} K((z-z_i)/h_n)$, which generates a downward bias in the density estimate at this boundary. For most standard choice of symmetric kernel, the density estimator $f_n(z)$ typically performs poorly on the right neighborhood of zero. This bias arises because for points z_i in the right neighborhood of 0, the contribution of z_i given by $n^{-1}h_n^{-1}K((z-z_i)/h_n)$ to $f_n(z)$ extends to points $z \leq 0$ where $f(z) = 0$. A similar bias arise in the multivariate density estimates which imposes the elliptical symmetry restriction. Ignoring the additional information of the restricted support of the transformed variable generates a similar downward bias around the mean. This bias creates a volcano like contour in the bivariate density estimate. The overflow in weights beyond the lower support of 0 can be corrected by using an estimator which incorporates this additional support constraint information into $f_n(x)$.

Schuster (1985) offers a correction that incorporates this overflow to the region $z < c$, for finite c,back into the region $z \geq c$ by adding a mirror image term $n^{-1}h_n^{-1}K((z-2c+z_i)/h_n)$ to $n^{-1}h_n^{-1}K((z-z_i)/h_n)$. The resulting estimator for $z \geq c$ is given by

$$\widetilde{f}_n(z) = \frac{1}{nh_n}\sum_{i=1}^{n}\left[K\left(\frac{z-z_i}{h_n}\right) + K\left(\frac{z-2c+z_i}{h_n}\right)\right].$$

In our case, $c = 0$. Schuster (1985) also proves consistency and asymptotic normality results for this estimator.

5.2 Bandwidth Selection

The smoothing parameter used in the kernel estimation is chosen in two ways. The first involves the standard approach for density estimation suggested in Silverman (1986). We shall compute the Silverman (1986) rule-of-thumb (ROT) bandwidth for estimation of the density $\gamma(z)$ of the transformed random variable $z = \tau\left(\varepsilon^T\varepsilon\right)$. This bandwidth is optimal in terms of minimizing the weighted (using a χ_p^2 density) mean integrated square error (MISE) of the density estimate if the underlying density is normal. Of course, this density is unknown, but the ROT provides an easy-to-compute benchmark. The ROT bandwidth will depend on both the transformation used and the kernel.

The scoring equation used in our procedure requires that we estimate the ratio γ'/γ of the transformed variable z. The kernel estimate of the derivative of the density $\widehat{\gamma}'$ in the numerator will have a convergence rate which is slower compared to that of the density estimate $\widehat{\gamma}$ in the denominator. This slower convergence rate is likely to dominate in the estimate of this ratio. The second approach in parameterizing h_{opt} involves calculating the optimal smoothing parameter that minimizes the approximation of the mean integrated squared error to the kernel estimate of the derivative of the density $\gamma'(z)$. This alternative criterion is appropriate given that the slower rate of convergence in the derivative estimate is likely to dominate that of the density estimate.

5.3 Test for elliptical symmetry

In this section, we describe a test for elliptical symmetry developed by Beran (1979) and discuss its implementation. Suppose we have a series of standardized regression residuals $\widehat{\varepsilon}_t = \widehat{\Sigma}^{-1/2}\widehat{u}_t$ for $t = 1, \ldots, n$, where $\widehat{u}_t$ are OLS residuals and $\widehat{\Sigma}$ is a consistent estimator of Σ. These residuals will be used in the construction of a test of the null hypothesis that the true underlying innovations $\{\varepsilon_t\}$ are i.i.d. draws from a spherically symmetric density. The test utilizes a couple of distinctive features of spherically symmetric random variables. The first is that the standardized random variable $\varepsilon_t/\left\|\varepsilon_t\right\|$ is uniformly distributed on the $m-1$-dimensional unit hypersphere. The second is that this standardized random vector, which we refer to as the "direction" of ε_t, is independent of the vector's "length", viz. $\left\|\varepsilon_t\right\|$.

We now describe the construction of the test, provide some intuition as to how it incorporates the aforementioned characteristics, and then state the test's asymptotic distribution under the null. We begin by ranking the distances $\{\|\widehat{\varepsilon}_t\|\}_{t=1}^n$ and dividing these ranks by $n+1$. Let $\{R_t\}_{t=1}^n$ denote these ranks. Note that the directional vector $\varepsilon/\left\|\varepsilon\right\|$ can be represented in terms of its polar coordinates $\Xi = \left(\xi_1, \ldots, \xi_{m-1}\right)$ as follows:

$$\frac{\varepsilon}{\|\varepsilon\|} = \left(\cos\left(\xi_1\right), \sin\left(\xi_1\right)\cos\left(\xi_2\right), \ldots, \sin\left(\xi_1\right)\sin\left(\xi_2\right), \ldots, \sin\left(\xi_{m-1}\right)\right).$$

Define the coordinates of $\widehat{\varepsilon}_t/\left\|\widehat{\varepsilon}_t\right\|$ by Ξ_t. Let $\{a_k : k \geq 1\}$ be the family of functions orthonormal with respect to the Lebesgue measure on $[0,1]$ and orthogonal to the constant function on $[0,1]$. Furthermore, let $\{b_\ell : \ell \geq 1\}$ denote another family of orthonormal functions with respect to the uniform measure on $[0,\pi]^{p-2} \times [0,2\pi)$ and orthogonal to the constant function on this domain [we use Legendre's polynomials described below]. Beran (1979) proposed a statistic of the form

$$S_n = \sum_{k=1}^{K_n} \sum_{\ell=1}^{L_n} \left[\frac{1}{\sqrt{n}} \sum_{t=1}^{n} a_k(R_t)\, b_m(\Xi_t) \right]^2.$$

If the innovations $\{\varepsilon_t\}$ have a spherically symmetric distribution, then S_n should be close to zero; otherwise, it should be far from zero. Using the fact that, under the null, R_t and Ξ_t both have uniform distributions, it follows from our assumptions on $\{a_k\}$ and $\{b_\ell\}$ that $E[a_k] = E[b_\ell] = 0$ for all k, ℓ. The independence of R_t and Ξ_t under the null furthermore implies that $E[a_k(R_t)\, b_m(\Xi_t)] = 0$ for all k, ℓ. The following proposition gives the asymptotic distribution of S_n.

Proposition 1 (Beran (1979)) . *Suppose that the functions $\{a_k : k \geq 1\}$ and $\{b_\ell : \ell \geq 1\}$ are differentiable and that:*

1. $\lim_{n\to\infty} K_n = \lim_{n\to\infty} L_n = \infty$
2. $\lim_{n\to\infty} n^{-1/2} K_n^{-1/2} L_n^{1/2} \sum_{k=1}^{K_n} \|a_k'\| = \lim_{n\to\infty} n^{-1/2} K_n^{1/2} L_n^{-1/2} \sum_{\ell=1}^{L_n} \|b_\ell'\| = 0$
3. $\lim_{n\to\infty} \frac{1}{n} M_n K_n = 0$.

Then, the null limiting distribution of $(2M_nK_n)^{-\frac{1}{2}} [S_n - M_nK_n]$ as $n \to \infty$ is $N(0,1)$.

This test of Beran (1979) is based on Fourier series expansion density estimators. Considering the fact that our estimation routine employs *kernel*-based density estimation, it would seem natural to employ a kernel-based test for elliptical symmetry. We are not immediately aware of the existence of such a test, although it would presumably be feasible to develop one, perhaps employing results on kernel-based goodness-of-fit tests for density functions as developed by Fan (1994).

5.4 Orthonormal Polynomials

The orthonormal polynomials used for our set of functions arise from solutions to what is called Legendre's differential equations of the form,

$$(1-x^2)y'' - 2xy' + n(n+1)y = 0.$$

The general solution is given by

$$y = c_1 P_n(x) + c_2 Q_n(x),$$

where the polynomials used are given by

$$P_n(x) = \frac{1}{2^n n!} \frac{d^n}{dx^n} (x^2 - 1)^n$$

where $x \in [-1, 1]$. These polynomials are suitably transformed to ensure orthogonality in the appropriate domain specified by the test statistics.

The family of differentiable orthonormal polynomials $\{a(z)_k : k \geq 1\}$ on the $[0, 1]$ domain for z are as follow:

$$\begin{aligned}
a(z)_1 &= \sqrt{3}(2z - 1) \\
a(z)_2 &= \sqrt{5}(6z^2 - 6z + 1) \\
a(z)_3 &= \sqrt{7}\,(2z - 1)\left(10z^2 - 10z + 1\right) \\
a(z)_4 &= \sqrt{9}(70z^4 - 140z^3 + 90z^2 - 20z + 1) \\
a(z)_5 &= \sqrt{\frac{11}{64}}\left(63\,(2z - 1)^5 - 70\,(2z - 1)^3 + 15\,(2z - 1)\right) \\
a(z)_6 &= \frac{\sqrt{13}}{16}\left(231\,(2z - 1)^6 - 315\,(2z - 1)^4 + 105\,(2z - 1)^2 - 5\right) \\
a(z)_7 &= \frac{\sqrt{15}}{16}\left(429\,(2z - 1)^7 - 693\,(2z - 1)^5 + 315\,(2z - 1)^3 - 35\,(2z - 1)\right)
\end{aligned}$$

The family of orthonormal polynomials $\{b(\xi)_m : m \geq 1\}$ are as follow:

$$
\begin{aligned}
b(\xi)_1 &= \sqrt{\frac{3}{\lambda\pi}}(\frac{2\xi}{\lambda\pi}-1) \\
b(\xi)_2 &= \sqrt{\frac{5}{\lambda\pi}}\left(6\left(\frac{\xi}{\lambda\pi}\right)^2-\frac{6\xi}{\lambda\pi}+1\right) \\
b(\xi)_3 &= \sqrt{\frac{7}{\lambda\pi}}\left(\frac{2\xi}{\lambda\pi}-1\right)\left(10\left(\frac{\xi}{\lambda\pi}\right)^2-\frac{10\xi}{\lambda\pi}+1\right) \\
b(\xi)_4 &= \sqrt{\frac{9}{\lambda\pi}}\left(70\left(\frac{\xi}{\lambda\pi}\right)^4-140\left(\frac{\xi}{\lambda\pi}\right)^3+90\left(\frac{\xi}{\lambda\pi}\right)^2-20\left(\frac{\xi}{\lambda\pi}\right)+1\right) \\
b(\xi)_5 &= \sqrt{\frac{11}{64\lambda\pi}}\left(63\left(\frac{2\xi}{\lambda\pi}-1\right)^5-70\left(\frac{2\xi}{\lambda\pi}-1\right)^3+15\left(\frac{2\xi}{\lambda\pi}-1\right)\right) \\
b(\xi)_6 &= \sqrt{\frac{13}{256\lambda\pi}}\left(231\left(\frac{2\xi}{\lambda\pi}-1\right)^6-315\left(\frac{2\xi}{\lambda\pi}-1\right)^4+105\left(\frac{2\xi}{\lambda\pi}-1\right)^2-5\right) \\
b(\xi)_7 &= \sqrt{\frac{15}{256\lambda\pi}}\left(429\left(\frac{2\xi}{\lambda\pi}-1\right)^7-693\left(\frac{2\xi}{\lambda\pi}-1\right)^5+315\left(\frac{2\xi}{\lambda\pi}-1\right)^3-35\left(\frac{2\xi}{\lambda\pi}-1\right)\right)
\end{aligned}
$$

where $\lambda = 1$ for the range $\xi \in [0, \pi]$ and $\lambda = 2$ for the range $\theta \in [0, 2\pi]$.

5.5 Bias Correction for Information Matrix

Our estimator of the information matrix, although consistent, has a finite sample upwards bias that therefore biases downwards our standard error estimates. In our empirical application, we employ a simple degrees of freedom correction. Write $\widehat{\gamma}'_t = \sum_s \omega'_{nts}$ and $\widehat{\gamma}_t = \sum_s \omega_{nts}$ for some weights ω'_{nts} and ω_{nts} implicitly defined in our estimation algorithm. We replace $\left(\widehat{\gamma}'_t\right)^2$ and $\left(\widehat{\gamma}_t\right)^2$ in (6) by $\left(\widehat{\gamma}'_t\right)^2 - \sum_s \left(\omega'_{nts}\right)^2$ and $\left(\widehat{\gamma}_t\right)^2 - \sum_s \left(\omega_{nts}\right)^2$ respectively. The correction terms $\sum_s \left(\omega'_{nts}\right)^2$ and $\sum_s \left(\omega_{nts}\right)^2$ consistently estimate the degrees of freedom bias terms (see Linton (1995)).

Acknowledgements

We would like to thank Pedro Gozalo and a referee for helpful comments, Hyungsik Moon and Eugene Choo for research assistance and the National Science Foundation, Social Sciences and Humanities Research Council of Canada, and Montreal Institute of Mathematical Finance for financial support.

References

[1] Bailey, R.W., Baillie, R.T., and McMahon, P.C. 1984. Interpreting econometric evidence on efficiency in the foreign exchange market. *Oxford Economic Papers* 36:67-85.

[2] Baillie, R.T., Lippens, R.E., and McMahon, P.C. 1983. Testing rational expectations and efficiency in the foreign exchange market. *Econometrica* 51:553-563.

[3] Baillie, R. and McMahon, P. 1989. *The Foreign Exchange Market: Theory and Econometric Evidence.* Cambridge; Cambridge Univ. Press.

[4] Baillie, R. and Bollerslev, T. 1989. Common stochastic trends in a system of exchange rates. *Journal of Finance* 64: 167-181.

[5] Baillie, R. and Bollerslev, T. 1994. The long memory of the forward premium. *Journal of International Money and Finance* 13:565-571.

[6] Barnhart, S.W. and Szakmary, A.C. 1991. Testing the unbiased forward rate hypothesis: Evidence on unit roots, cointegration, and stochastic coefficients. *Journal of Financial and Quantitative Analysis* 26:245-267.

[7] Beelders, O. 1998. Adaptive unit root tests. working paper. Emory University.

[8] Beran, R. 1979. Testing for ellipsoidal symmetry of a multivariate density. *Annals of Statistics* 7:150-162.

[9] Bickel, P.J. 1982. On adaptive estimation. *Annals of Statistics* 10:647-671.

[10] Bilson, J.F.O. 1981. The "speculative efficiency" hypothesis. *Journal of Business* 54:435-451.

[11] Casella, G. and Berger, R.L. 1990. *Statistical Inference*. Duxbury Press: Belmont, California.

[12] Chamberlain, G. 1983. A characterization of the distributions that imply mean-variance utility functions. *Journal of Economic Theory* 29:185-201.

[13] Cornell, B. 1989. The impact of data errors on measurement of the foreign exchange risk premium. *Journal of International Money and Finance* 8:147-157.

[14] ENGEL, C. 1996. The forward discount anomaly and the risk premium: A survey of recent evidence. *Journal of Empirical Finance* 3:123-192.

[15] EVANS, M.D.D. AND LEWIS, K.K. 1995. Do long-term swings in the dollar affect estimates of the risk premia? *Review of Financial Studies* 8:709-742.

[16] FAMA, E.F. 1984. Forward and spot exchange rates. *Journal of Monetary Economics* 14:319-338.

[17] FAN, Y. 1994. Testing the goodness-of-fit of a parametric density function by kernel methods. *Econometric Theory* 10:316-356.

[18] FERNÁNDEZ, C., OSIEWALSKI, J. AND STEEL, M.F.J. 1995. Modelling and inference with v-spherical distributions. *Journal of the American Statistical Association* 90:1331-1340.

[19] HÄRDLE, W., AND O.B. LINTON. 1994. Applied nonparametric methods. In *The Handbook of Econometrics*, Vol. IV, pp. 2295-2339, eds D.F. McFadden and R.F. Engle III. North Holland.

[20] HAKKIO, C.S. 1981. Expectations and the forward exchange rate. *International Economic Review* 22:663-678.

[21] HAKKIO, C.S. AND RUSH, M. 1989. Market efficiency and cointegration: an application to the sterling and deutschemark exchange markets. *Journal of International Money and Finance* 8:75-88.

[22] HANSEN, L.P. AND HODRICK, R.J. 1980. Forward exchange rates as optimal predictors of future spot rates: An econometric analysis. *Journal of Political Economy* 88:829-853.

[23] HODGSON, D.J. 1998a. Adaptive estimation of cointegrating regressions with ARMA errors. *Journal of Econometrics* 85:231-268.

[24] HODGSON, D.J. 1998b. Adaptive estimation of error correction models. *Econometric Theory* 14:44-69.

[25] HODGSON, D.J. 1999. Adaptive estimation of cointegrated models: Simulation evidence and an application to the forward exchange market. *Journal of Applied Econometrics* 14:627-650.

[26] HODGSON, D.J. 2000. Partial maximum likelihood estimation and adaptive estimation in the presence of conditional heterogeneity of unknown form. *Econometric Reviews* 19:175-206.

[27] Hodgson, D.J., Linton, O., and Vorkink, K. 2002. Testing the capital asset pricing model efficiently under elliptical symmetry: A semiparametric approach. In press, *Journal of Applied Econometrics.*

[28] Hsieh, D.A. and Manski, C.F. 1987. Monte Carlo evidence on adaptive maximum likelihood estimation of a regression. *Annals of Statistics* 15:541-551.

[29] Jeganathan, P. 1995. Some aspects of asymptotic theory with applications to time series models. *Econometric Theory* 11:818-887.

[30] Johansen, S. 1988. Stochastic analysis of cointegration vectors. *Journal of Economic Dynamics and Control* 12:231-254.

[31] Kreiss, J.-P. 1987. On adaptive estimation in stationary ARMA processes. *Annals of Statistics* 15:112-133.

[32] Levine, R. 1989. The pricing of forward exchange rates. *Journal of International Money and Finance* 8:163-179.

[33] Levine, R. 1991. An empirical inquiry into the nature of the forward exchange rate bias. *Journal of International Economics* 30:359-369.

[34] Linton, O. 1995. Second order approximations in a partially linear regression model. *Econometrica* 63:1079-1113.

[35] Mardia, K.V. 1970. Measures of multivariate skewness and kurtosis with applications. *Biometrika* 57:519-530.

[36] Maynard, A. and Phillips, P.C.B. 2001. Rethinking an old empirical puzzle: Econometric evidence on the forward discount anomaly. *Journal of Applied Econometrics* 16:671-708.

[37] McCallum, B. 1994. A reconsideration of the uncovered interest parity relationship. *Journal of Monetary Economics* 33:105-132.

[38] Muirhead, R.J. 1982. *Aspects of Multivariate Statistical Theory*. New York; Wiley.

[39] Phillips, P.C.B. 1995. Robust nonstationary regression. *Econometric Theory* 11:912-951.

[40] Phillips, P.C.B. and Hansen, B. 1990. Statistical inference in instrumental variables regression with I(1) processes. *Review of Economic Studies* 57:99-125.

[41] Phillips, P.C.B. and McFarland, J.W. 1997. Forward exchange market unbiasedness: The case of the Australian dollar since 1984. *Journal of International Money and Finance* 16:885-907.

[42] Phillips, P.C.B., McFarland, J.W., and McMahon, P.C. 1996. Robust tests of forward exchange market efficiency with empirical evidence from the 1920's. *Journal of Applied Econometrics* 11:1-22.

[43] Schick, A. 1987. A note on the construction of asymptotically linear estimators. *Journal of Statistical Planning and Inference* 16:89-105.

[44] Schuster, E. 1985. Incorporating support constraints into nonparametric estimators of densities. *Communications in Statistics - Theory and Methods* 14:1123-1136.

[45] Silverman, B.W. 1986. *Density Estimation for Statistics and Data Analysis.* London; Chapman & Hall.

[46] Steigerwald, D. 1992a. Adaptive estimation in time series regression models. *Journal of Econometrics* 54:251-275.

[47] Steigerwald, D. 1992b. On the finite sample behaviour of adaptive estimators. *Journal of Econometrics* 54:371-400.

[48] Stone, C. 1975. Adaptive maximum likelihood estimation of a location parameter. *Annals of Statistics* 3:267-284.

[49] Stute, W. and Werner, U. 1991. Nonparametric estimation of elliptically contoured densities. In Roussas, G. (ed.), *Nonparametric Functional Estimation and Related Topics*, Kluwer Academic Publishers, pp. 173-190.

[50] White, H. 1982. Maximum likelihood estimation of misspecified models. *Econometrica* 50:1-25.

[51] Zellner, A. 1962. An efficient method for estimating seemingly unrelated regressions and tests for aggregation bias. *Journal of the American Statistical Association* 57:348-368.

Table I

Summary Statistics Cointegrated Model

Below are summary statistics of the logged spot and forward rates used in the empirical analysis. ADF stands for Augmented Dicky-Fuller unit root test where 20 lagged difference terms and a constant are included in the test. Critical values for the ADF statistics on the daily data are -3.4395, -2.8648, and -2.5685 at the 1%, 5%, and 10% respectively. Critical values for the ADF statistics on the weekly data are -3.9817, -3.4213, and -3.1331 at the 1%, 5%, and 10% respectively.

Variable	Mean	Std. Dev.	min	max	ADF
Daily ($n = 806$)					
s_t					
JPY/USD	4.762	0.086	4.621	4.986	-1.348
GBP/USD	0.433	0.061	0.317	.0537	-0.892
CAD/USD	0.408	0.028	.0344	0.472	-1.322
f_t					
JPY/USD	0.408	0.028	0.345	0.473	-0.902
GBP/USD	0.433	0.061	0.317	0.537	-1.084
CAD/USD	0.408	0.028	0.345	0.473	-0.984
Weekly ($n = 468$)					
s_t					
JPY/USD	4.717	0.107	4.415	4.985	-2.514
GBP/USD	0.443	0.050	0.321	0.536	-1.351
CAD/USD	0.350	0.059	0.218	0.478	-2.316
f_t					
JPY/USD	4.716	0.107	4.411	4.986	-2.512
GBP/USD	0.443	0.050	0.321	0.535	-1.458
CAD/USD	0.350	0.059	0.219	0.478	-2.738

Table II

Summary Statistics Stationary Model

Below are summary statistics of the logged spot and forward rates used in the empirical analysis. ADF stands for Augmented Dicky-Fuller unit root test where 20 lagged difference terms and a constant are included in the test. See the note to Table I for critical values for the ADF statistics on the daily and weekly data.

Variable	Mean	Std. Dev.	min	max	ADF
Daily ($n = 806$)					
$s_{t+1} - s_t$					
JPY/USD	0.000	0.008	-0.069	0.033	-7.266
GBP/USD	0.000	0.005	-0.016	0.019	-7.671
CAD/USD	0.000	0.003	-0.016	0.011	-7.612
$f_t - s_t$					
JPY/USD	0.000	0.000	-0.001	0.000	-5.669
GBP/USD	0.000	0.000	-0.001	0.001	-3.682
CAD/USD	0.000	0.000	-0.001	0.001	-4.629
Weekly ($n = 468$)					
$s_{t+1} - s_t$					
JPY/USD	0.000	0.017	-0.150	0.059	-5.597
GBP/USD	0.000	0.012	-0.035	0.039	-5.576
CAD/USD	0.001	0.007	-0.027	0.021	-10.186
$f_t - s_t$					
JPY/USD	0.000	0.001	-0.001	0.001	-3.861
GBP/USD	0.000	0.001	-0.013	0.023	-7.869
CAD/USD	0.000	0.001	-0.007	0.022	-3.856

Table III

Properties of Model Residuals

The test statistics below are Box-Pierce tests of residual serial correlation. P-values are in parentheses and *indicates a p-value less than .001.

Panel A: Cointegrated Model

Daily Data	JPY/USD	GBP/USD	CAD/USD
Box-Pierce (q=1)	1.58(0.21)	0.72(0.40)	0.93(0.34)
Box-Pierce (q=5)	3.52(0.62)	3.44(0.63)	4.54(0.48)
Box-Pierce (q=10)	11.76(0.30)	7.18(0.71)	5.61(0.85)
Box-Pierce (q=20)	28.62(0.10)	14.25(0.82)	14.79(0.79)

Weekly Data			
Box-Pierce (q=1)	0.00(0.97)	20.43(0.00*)	2.27(0.13)
Box-Pierce (q=5)	8.03(0.15)	28.71(0.00*)	7.41(0.19)
Box-Pierce (q=10)	9.96(0.44)	36.65(0.00*)	9.71(0.47)
Box-Pierce (q=20)	25.08(0.20)	57.85(0.00*)	19.34(0.50)

Panel B: Stationary Model

Daily Data			
Box-Pierce (q=1)	1.60(0.21)	0.85(0.36)	1.16(0.28)
Box-Pierce (q=5)	3.49(0.63)	3.88(0.57)	4.96(0.42)
Box-Pierce (q=10)	10.94(0.36)	7.52(0.68)	6.16(0.80)
Box-Pierce (q=20)	29.53(0.09)	15.11(0.77)	15.42(0.75)

Weekly Data			
Box-Pierce (q=1)	1.14(0.29)	0.00(0.96)	6.73(0.02)
Box-Pierce (q=5)	7.13(0.21)	4.72(0.45)	12.66(0.03)
Box-Pierce (q=10)	8.65(0.57)	7.60(0.67)	14.34(0.16)
Box-Pierce (q=20)	22.82(0.30)	24.78(0.21)	23.91(0.25)

Table IV

Multivariate Tests of Normality and Elliptical Symmetry

The test statistics below are Mardia's (1970) multivariate kurtosis measure and Beran's (1979) elliptical symmetry measure, S_n. Tests are constructed using residuals from OLS estimation of stated model. P-values are in parentheses and *indicates a p-value less than .001.

Panel A: Multivariate Kurtosis Test

	Test Statistic
Stationary Model, Daily Data	34.268(0.00*)
Cointegrated Model, Daily Data	35.172(0.00*)
Stationary Model, Weekly Data	27.340(0.00*)
Cointegrated Model, Weekly Data	25.904(0.00*)

Panel B: Elliptical Symmetric Test ($\mathbf{S}_n$)

	Test Statistic
Stationary Model, Daily Data	3.275(0.07)
Cointegrated Model, Daily Data	3.634(0.06)
Stationary Model, Weekly Data	0.901(0.31)
Cointegrated Model, Weekly Data	1.228(0.25)

Table V

Results of Spot-Futures Estimation

Cointegrated Model

$$s_{t+1} = \alpha + \beta f_t + u_{t+1}$$

Panel A: Daily Data, OLS Estimates

	α		β	
Exchange Rate	Estimate	Std. Error	Estimate	Std. Error
JPY/USD	0.0172	0.0154	0.9964	0.0032
GBP/USD	0.0022	0.0012	0.9950	0.0028
CAD/USD	0.0024	0.0018	0.9942	0.0043

Panel B: Daily Data, Adaptive Estimates

JPY/USD	0.0146	0.0147	0.9971	0.0031
GBP/USD	0.0002	0.0009	0.9989	0.0021
CAD/USD	0.0017	0.0018	0.9961	0.0044

Panel C: Weekly Data, OLS Estimates

JPY/USD	0.0613	0.0294	0.9872	0.0062
GBP/USD	0.0041	0.0035	0.9921	0.0074
CAD/USD	0.0001	0.0013	1.0001	0.0038

Panel D: Weekly Data, Adaptive Estimates

JPY/USD	0.0093	0.0272	0.9986	0.0057
GBP/USD	0.0022	0.0029	0.9945	0.0063
CAD/USD	0.0001	0.0012	1.0000	0.0034

Table VI

Results of Spot-Futures Estimation

Stationary Model

$$s_{t+1} - s_t = \alpha + \beta(f_t - s_t) + u_{t+1}$$

Panel A: Daily Data, OLS Estimates

	α		β	
Exchange Rate	Estimate	Std. Error	Estimate	Std. Error
JPY/USD	-0.0007	0.0005	-3.4552	1.8148
GBP/USD	-0.0001	0.0002	-0.2931	2.3153
CAD/USD	0.0000	0.0001	-2.0860	2.6759

Panel B: Daily Data, Adaptive Estimates

JPY/USD	0.0002	0.0005	0.1558	1.7122
GBP/USD	-0.0002	0.0002	1.7852	1.9943
CAD/USD	0.0002	0.0001	1.6924	2.8459

Panel C: Weekly Data, OLS Estimates

JPY/USD	-0.0004	0.0008	-0.2680	0.3309
GBP/USD	-0.0001	0.0005	-0.5446	0.1682
CAD/USD	0.0005	0.0003	0.1362	0.1897

Panel D: Weekly Data, Adaptive Estimates

JPY/USD	0.0008	0.0003	-0.2376	0.1250
GBP/USD	0.0000	0.0004	-0.6187	0.1007
CAD/USD	0.0006	0.0002	0.1183	0.1236

Table VII

Forward Rate Unbiased Tests

$\mathbf{H_0}$: $\alpha^i = 0,\ \beta^i = 1, \quad i = 1, \ldots, m$

Under the null, J is distributed asymptotically χ^2_6. P-values are in parentheses following the test statistics. *Indicates a p-value less than .001.

	J (p-value)
Stationary Model, Daily Returns	
OLS	8.247(0.22)
Adaptive	5.946(0.43)
Stationary Model, Weekly Returns	
OLS	117.864(0.00)
Adaptive	469.881(0.00)
Cointegrated Model, Daily Returns	
OLS	6.916(0.33)
Adaptive	7.454(0.28)
Cointegrated Model, Weekly Returns	
OLS	8.573(0.20)
Adaptive	15.007(0.02)

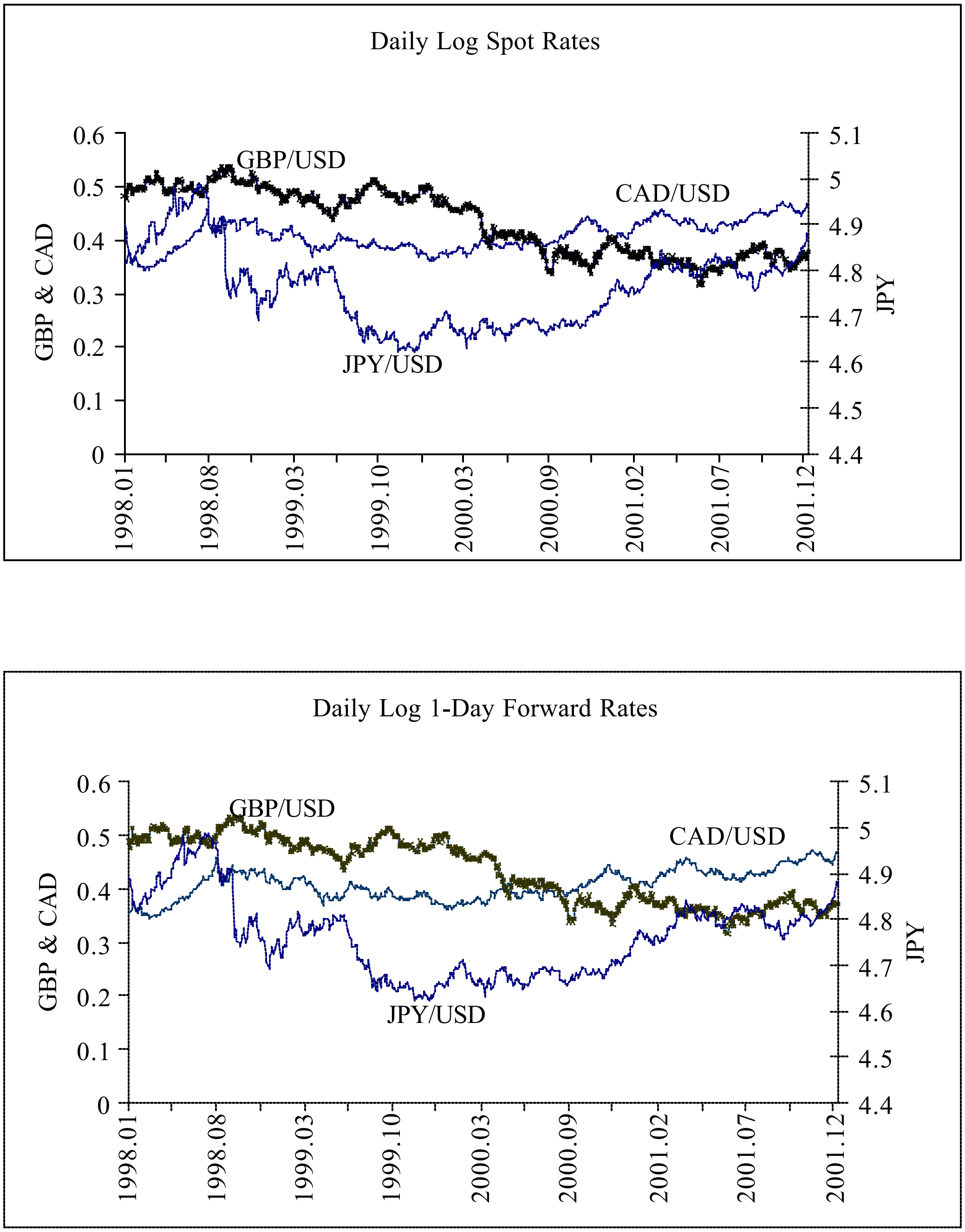
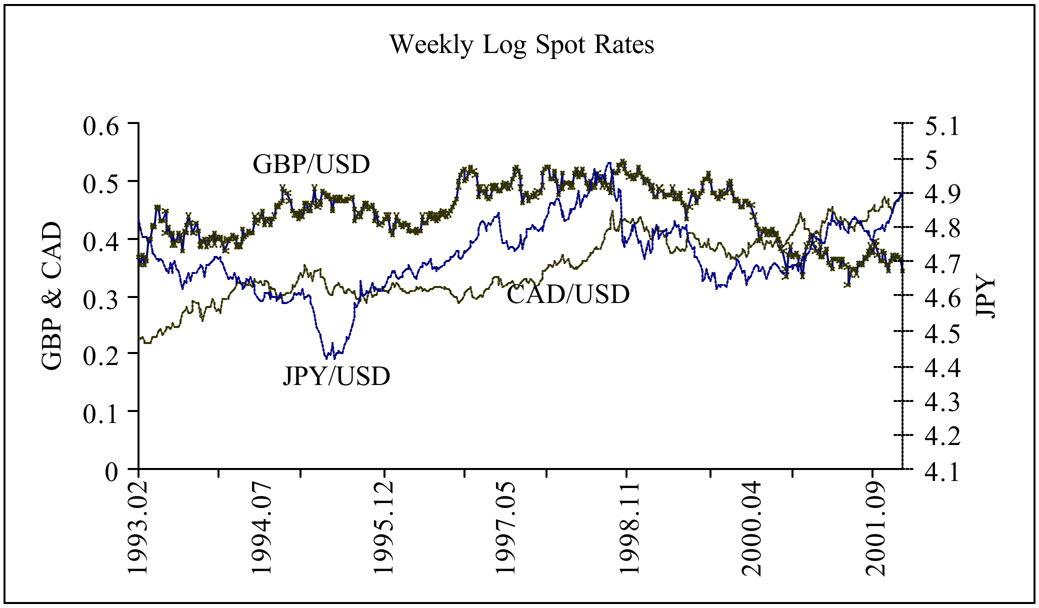


Figure 1: Daily Log Spot and Forward Rates



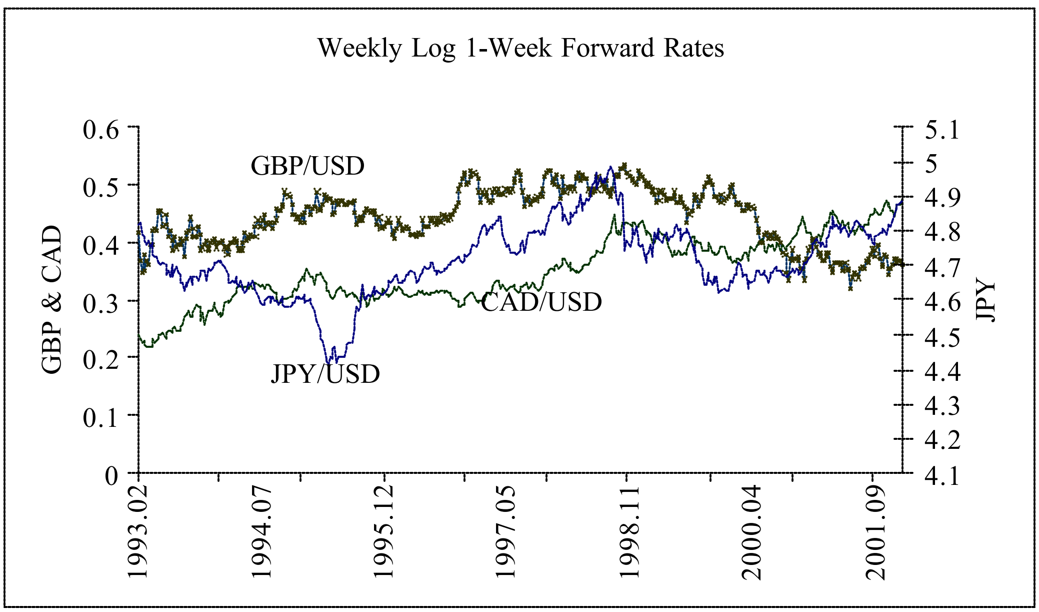


Figure 2: Weekly Log Spot and Forward Rates