TESTING FORWARD EXCHANGE RATE
UNBIASEDNESS EFFICIENTLY:
A SEMIPARAMETRIC APPROACH

Douglas J. Hodgson
University of Quebec at Montreal
Oliver Linton
London School of Economics
Keith Vorkink[*]
Brigham Young University

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. We test the unbiasedness hypothesis on both weekly and daily exchange rate data and strongly reject unbiasedness at the weekly horizon, but fail to reject the unbiasedness hypothesis on the daily data. Estimates of the semiparametric procedure in some cases differ substantially from traditional OLS estimates.

I. 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, the analysis is usually 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

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

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 non-normality is a pervasive characteristic of exchange rate data, only Bilson (1981) among the authors listed above attempts to gauge its effects 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

---

[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.

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 (1998a,b). These authors defined adaptive estimators of the identifiable parameters in various regression models. However, their proposed estimators 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, p. 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 II, 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 III, we provide a formula for computing the adaptive estimator developed by Hodgson, Linton, and Vorkink (2002) for stationary models and describe the extension of this estimator to cointegrated models, while Section IV reports the results of our exchange rate analysis. We use $\xrightarrow{P}$ to denote convergence in probability and $\xrightarrow{D}$ to denote convergence in distribution. We say that $X \sim MN(0, V)$ when X is mixed normal with (possibly) random covariance matrix V.

II. 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 $\{ f_t^i \}$, $t = 1,\ldots, n + 1$, $i = 1,\ldots, m$. 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 IV.

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:

$$\mathbf{H}_0 : \alpha^i = 0, \ \beta^i = 1, \quad i = 1,\ldots, m \tag{2}$$

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:

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

The corresponding regressions we estimate are of the following form:

$$s_{t+1}^i = \alpha^i + \beta^i f_t^i + u_{i,t+1}.$$

We are then interested in testing the hypothesis

$$\mathbf{H}_0 \colon \ \alpha^i = 0, \ \beta^i = 1, \quad i = 1, \ldots, m \tag{4}$$

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

$$y_t = \alpha + x_t \beta + u_t \coloneqq w_t \theta + u_t, \quad t = 1, \ldots, n, \tag{5}$$

where $y_t \in \Re^m, \alpha \in \Re^m,$

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

$w_t = [I_m \; x_t]$, $x_{it}^T \in \Re^{k_i}$ and $\beta_i \in \Re^{k_i}$ for every $i = 1, \ldots, m$, the full parameter vector is $\theta = \left[\alpha^T \beta^T \right]^T \in \Re^{m+k}$, where $k = k_1 + \ldots + k_m$, and $u_t \in \Re^m$ 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(.) 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 $\delta_n = \text{diag}\left[n^{-1/2} I_m, n^{-1} I_k \right]$ 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 $\bar{\theta}$ starting from a preliminary

δ_n^{-1} -consistent estimator $\hat{\theta}$, obtained from the Gaussian likelihood. Under general conditions, this will be first order asymptotically equivalent to the MLE, i.e.,

$$\delta_n^{-1}(\overline{\theta} - \theta_0) \xrightarrow{D} MN(0, \Omega^{-1}), \tag{6}$$

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

$$\Omega = \begin{bmatrix} \Phi_p & E\left[\Phi_p x_t\right] \\ E\left[x_t^T \Phi_p\right] & E\left[x_t^T \Phi_p x_t\right] \end{bmatrix},$$

while for the cointegrated model, it is

$$\Omega = \begin{bmatrix} \Phi_p & \Phi_p \int\limits_0^1 M(r)dr \\ \int\limits_0^1 M(r)^T dr \Phi_p & \int\limits_0^1 M(r)^T \Phi_p M(r)dr \end{bmatrix},$$

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 $i = 1,\dots, m$. Note that in the case of cointegration, Ω 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 $\tilde{\theta}$ will have the form

$$\tilde{\theta} = \hat{\theta} + \delta_n \hat{\Omega}_n^{-1}(\hat{\theta}) \, \hat{\Delta}_n(\hat{\theta}), \tag{7}$$

where $\hat{\Delta}_n$ and $\hat{\Omega}_n$ are estimates of the first and second standardized derivatives of L_n respectively. Their computation is described in section III below. In particular,

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

where $\hat{\varphi}_t(\hat{u}_t)$ is a consistent estimator of the m-dimensional score vector $\varphi(u_t)$, while $\hat{u}_t = y_t - w_t\hat{\theta}$. The standard approach to this problem is to use multivariate kernel estimates $\hat{p}$ and $\hat{p}'$ to construct $\hat{\varphi}$, 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 $\hat{\varphi}_t(.)$ 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.

III. 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 $\hat{\varphi}$, 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

---

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

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 $\tau (.)$ 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 $\tau(v;\zeta) = (v^{\zeta} - 1)/\zeta$, 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 of $g(.)$ 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 $\Phi_p = E\left[\varphi(u)\varphi(u)^T \right]$. 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 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 $J_\tau(z) = \left| \partial \tau^{-1}(z) / \partial z \right|$. Thus, $g(v) = c_m^{-1} J_\tau^{-1}\{\tau(v)\} v^{1-m/2} \gamma\{\tau(v)\}$, This gives 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 (7) above presupposed the existence of consistent score and information estimators $\hat{\varphi}_t$ and $\hat{\Omega}_n$. 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 Ω proceeds according to the following steps:

1. First obtain $\hat{\theta}$ and define the residuals $\{\hat{u}_t\}_{t=1}^n$ and the standardized residuals $\{\hat{\varepsilon}_t\}_{t=1}^n$, where $\hat{\varepsilon}_t = \hat{\Sigma}^{-1/2} \hat{u}_t$, $\hat{\Sigma} = \hat{c}^{-1} \hat{\Sigma}_u$, $\hat{\Sigma}_u = n^{-1} \sum_{t=1}^n \hat{u}_t \hat{u}_t^T$ and $\hat{c} = \left[\det \hat{\Sigma}_u \right]^{1/m}$.

Then compute the transformed sequence $\{\hat{z}_t\}_{t=1}^n$, where $\hat{z}_t = \tau(\hat{v}_t)$ with $\hat{v}_t = \hat{\varepsilon}_t^T \hat{\varepsilon}_t$.

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

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

3. Introduce the following trimming conditions: (i) $\hat{\gamma}_t(\hat{z}_t) \geq d_n$; (ii) $|\hat{z}_t| \leq e_n$; (iii) $|\lambda(\hat{z}_t)| \leq b_n$; (iv) $|\rho^{1/2}(\hat{z}_t)\hat{\gamma}_t'(\hat{z}_t)| \leq c_n \hat{\gamma}_t(\hat{z}_t)$, 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:

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

where $\quad s(v) = (1 - m/2)v^{-1} - \dfrac{J_\tau'}{J_\tau}\{\tau(v)\}\tau'(v)$, and $\quad \hat{\Phi}_p = \dfrac{1}{n}\sum_{t=1}^n \hat{\varphi}_t(\hat{u}_t)\hat{\varphi}_t(\hat{u}_t)^T$.

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

$$\hat{\Delta}_n(\hat{\theta}) = -\delta_n \sum_{t=1}^n w_t' \hat{\varphi}_t(\hat{u}_t); \quad \hat{\Omega}_n(\hat{\theta}) = \delta_n \sum_{t=1}^n w_t^T \hat{\Phi}_p w_t \delta_n.$$

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

---

[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 (1998a) 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.

Proposition 1: *Suppose that* Φ_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,*

$$\delta_n^{-1}(\tilde{\theta} - \theta) \xrightarrow{D} MN(0, \Omega^{-1}). \tag{9}$$

i.e., the estimator $\tilde{\theta}$ *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 $\tau(.)$ that we use and can be more or less restrictive for different selections of $\tau(.)$. For example, when the transformation is $\tau(v; \zeta) = (v^\zeta - 1)/\zeta$, with either $\zeta = 0$, $\zeta = 1$, or $\zeta = 1/2m$, the condition implies that $E\left[(\varepsilon^T \varepsilon)^{m/2-2}\right] < \infty$. However, when $\zeta = m/2$, there is no restriction on the moments of u.

(b) Note that the information matrix estimator $\hat{\Omega}_n(\hat{\theta})$ defined in (8) is a consistent estimator of the asymptotic covariance matrix, so that $\hat{\Omega}_n(\hat{\theta}) - \Omega = o_p(1)$. This result is true even for cointegrated models, in which case Ω is random. We can therefore use $\hat{\Omega}_n(\hat{\theta})$ in the construction of *t*-ratios and Wald statistics, which will have respective standard normal and chi-squared asymptotic distributions. Let θ_l and $\tilde{\theta}_l$ be the lth elements of the θ and $\tilde{\theta}$ vectors, respectively. Now suppose we wish to test the null hypothesis that $\theta_t = r$, where *r* is some constant. Then we can compute the usual *t*-ratio, as follows:

$$\frac{(\delta_n^{-1})_{ll}(\tilde{\theta}_l - r)}{\sqrt{(\hat{\Omega}_n^{-1}(\hat{\theta}))_{ll}}} \xrightarrow{D} N(0,1)$$

under the null, where $(\delta_n^{-1})_{ll}$ and $(\hat{\Omega}_n^{-1}(\hat{\theta}))_{ll}$ are the lth elements along the diagonals of δ_n^{-1} and $\hat{\Omega}_n^{-1}(\hat{\theta})$, respectively. If we want to test the joint hypothesis $\theta_t = r$ for the entire vector θ, where *r* is now a known $(m + k)$ – vector of constants, we can compute the Wald statistic

$$\left[\delta_n^{-1}(\tilde{\theta} - r)\right]' \hat{\Omega}_n(\hat{\theta})\left[\delta_n^{-1}(\tilde{\theta} - r)\right] \xrightarrow{D} \chi_{m+k}^2.$$

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 that 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 $\hat{\varphi}$ and $\hat{\Phi}$ 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) The estimator is applied using a method developed by Schuster (1985) to account for bias that is present in our nonparametric density estimator in the neighborhood of the origin, and a correction similar to that of Linton (1995) for bias in our nonparametric information estimator. In addition, we choose optimal Silverman (1986) rule-of-thumb bandwidths separately for our nonparametric density and first derivative estimators $\hat{\gamma}$ and $\hat{\gamma}'$. More detail on these technical issues, as well as on the application of the Beran (1979) elliptical symmetry test, can be found in Hodgson, Linton, and Vorkink (2002), or in a technical appendix available from the authors upon request.

IV. 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 1 and 2 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 0.05 level with most rejections at the 0.01 level.[6]

Table 1. Summary Statistics for Cointegrated Model

Variable	Mean	Std. dev.	Min.	Max	ADF
Daily (n = 806)					
JPY/USD (s_t)	4.762	0.086	4.621	4.986	-1.348
GBP/USD (s_t)	0.433	0.061	0.317	.0537	-0.892
CAD/USD (s_t)	0.408	0.028	0.344	0.472	-1.322
JPY/USD (f_t)	4.762	0.086	4.621	4.992	-0.902
GBP/USD (f_t)	0.433	0.061	0.317	0.537	-1.084
CAD/USD (f_t)	0.408	0.028	0.345	0.473	-0.984
Weekly (n = 468)					
JPY/USD (s_t)	4.717	0.107	4.415	4.985	-2.514
GBP/USD (s_t)	0.443	0.050	0.321	0.536	-1.351
CAD/USD (s_t)	0.350	0.059	0.218	0.478	-2.316
JPY/USD (f_t)	4.716	0.107	4.411	4.986	-2.512
GBP/USD (f_t)	0.443	0.050	0.321	0.535	-1.458
CAD/USD (f_t)	0.350	0.059	0.219	0.478	-2.738

Note: These are summary statistics of the logged spot and forward rates used in the empirical analysis. ADF stands for Augmented Dickey-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.

─────────────

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

Table 2. Summary Statistics for Stationary Model

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

Notes: These are summary statistics of the logged spot and forward rates used in the empirical analysis. ADF stands for Augmented Dickey-Fuller unit root test where 20 lagged difference terms and a constant are included in the test. See the note to Table 1 for critical values for the ADF statistics on the daily and weekly data.

Table 3 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 4. 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 5% level for any of the series (although rejections at the 10% level would occur for the daily data).

Table 3. Properties of Model Residuals

Panel A: Cointegrated Model			
	JPY/USD	GBP/USD	CAD/USD
Daily Data			
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			
	JPY/USD	GBP/USD	CAD/USD
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)

Notes: The test statistics are Box-Pierce tests of residual serial correlation. P-values are in parentheses and * indicates a *p*-value less than 0.01.

Table 4. Multivariate Tests of Normality and Elliptical Symmetry

Panel A: Multivariate kurtosis test	
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 (S_n)	
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	1.288 (0.25)

Notes: The test statistics 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 0.01.

Tables 5 and 6 provide the estimation results, for the levels and differences regressions, respectively, while Wald statistics of the unbiasedness null hypothesis as stated in (2) and (4) are reported in Table 7. We note that the adaptive estimates are computed using a Gaussian kernel with Schuster's (1985) correction and the Box-Cox transformation $z = \tau(v) = (v^{\zeta} - 1)/\zeta$, with $\zeta = 1/2m$.[7]

A. Cointegrated Model. Results

Our estimates of the cointegrated model and associated unbiasedness test statistics are reported in Table 5, and in the second half of Table 7, respectively. Before analyzing our results, we will consider some of the previous evidence

---

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

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 nor 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 modeling 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-normality in exchange rate data. Phillips, MacFarland, and McMahon (1996) and Phillips and MacFarland (1997) apply the robust fully modified least absolute deviations 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) accept the hypothesis using FM-OLS but strongly reject

Table 5. Results of Spot-Futures Estimation, Cointegrated Model:
$s_{t+1} = \alpha + \beta f_t + u_{t+1}$

Exchange rate	α		β	
	Estimate	Std. error	Estimate	Std. error
Panel A: Daily data, OLS estimates				
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

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 5 and 7, 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 existing literature, although more unambiguously supportive of the hypothesis than much of it.

B. 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 (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 modeling 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 6 and in the first two panels of Table 7, where we compare the results obtained using OLS and the semiparametric adaptive estimator developed here. Regarding the basic inference as reported in Table 7, 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 6 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

Table 6. Results of Spot-Futures Estimation, Stationary Model:

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

Exchange rate	α		β	
	Estimate	Std. error	Estimate	Std. error
Panel A: Daily data, OLS estimates				
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

currencies, significantly 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

Table 7. Forward Rate Unbiased Tests, H$_0$: $\alpha^i = 0, \beta^i = 1, \ i = 1, \ldots, m$

	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)

Notes: 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 0.01.

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 modeled 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 inclusion 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 possibility 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.

References

Bailey, Ralph W., Richard T. Baillie, and Patrick C. McMahon (1984), "Interpreting Econometric Evidence on Efficiency in the Foreign Exchange Market," *Oxford Economic Papers* **36**: 67-85.

Baillie, Richard T., Robert E. Lippens, and Patrick C. McMahon (1983), "Testing Rational Expectations and Efficiency in the Foreign Exchange Market," *Econometrica* **51**: 553-563.

Baillie, Richard, and Patrick C. McMahon (1989), *The Foreign Exchange Market: Theory and Econometric Evidence*, Cambridge, Cambridge University Press.

Baillie, Richard T., and Tim Bollerslev (1989), "Common Stochastic Trends in a System of Exchange Rates," *Journal of Finance* **64**: 167-181.

Baillie, Richard T., and Tim Bollerslev (1994), "The Long Memory of the Forward Premium," *Journal of International Money and Finance* **13**: 565-571.

Barnhart, Scott W., and Carol Szakmary (1991), "Testing the Unbiased Forward Rate Hypothesis: Evidence on Unit Roots, Cointegration, and Stochastic Coefficients," *Journal of Financial and Quantitative Analysis* **26**: 245-267.

Beelders, Owen (1998), "Adaptive Unit Root Tests," Working Paper, Emory University.

Beran, Rudolph (1979), "Testing for Ellipsoidal Symmetry of a Multivariate Density," *Annals of Statistics* **7**: 150-162.

Bickel, Peter J. (1982), "On Adaptive Estimation," *Annals of Statistics* **10**: 647-671.

Bilson, John F.O. (1981), "The 'Speculative Efficiency' Hypothesis," *Journal of Business* **54**: 435-451.

Casella, George, and Roger L. Berger (1990), *Statistical Inference*, Belmont, CA, Duxbury Press.

Chamberlain, Gary (1983), "A Characterization of the Distributions that Imply Mean Variance Utility Functions," *Journal of Economic Theory* **29**: 185-201.

Cornell, Bradford (1989), "The Impact of Data Errors on Measurement of the Foreign Exchange Risk Premium," *Journal of International Money and Finance* **8**: 147-157.

Engel, Charles (1996), "The Forward Discount Anomaly and the Risk Premium: A Survey of Recent Evidence," *Journal of Empirical Finance* **3**: 123-192.

Evans, Martin D., and Karen K. Lewis (1995), "Do Long-term Swings in the Dollar Affect Estimates of the Risk Premia?," *Review of Financial Studies* **8**: 709-742.

Fama, Eugene F. (1984), "Forward and Spot Exchange Rates," *Journal of Monetary Economics* **14**: 319-338.

Fernández, Carmen, Jacek Osiewalski, and Mark F.J. Steel (1995), "Modelling and Inference with V-spherical Distributions," *Journal of the American Statistical Association* **90**: 1331-1340.

Hakkio, Craig S. (1981), "Expectations and the Forward Exchange Rate," *International Economic Review* **22**: 663-678.

Hakkio, Craig S., and Mark Rush (1989), "Market Efficiency and Cointegration: An Application to the Sterling and Deutschemark Exchange Markets," *Journal of International Money and Finance* **8**: 75-88.

Hansen, Lars Peter, and Robert J. Hodrick (1980), "Forward Exchange Rates as Optimal Predictors of Future Spot Rates: An Econometric Analysis," *Journal of Political Economy* **88**: 829-853.

Härdle, Wolfgang, and Oliver B. Linton (1994), "Applied Nonparametric Methods," in D.F. McFadden and R.F. Engle III, eds., *The Handbook of Econometrics*, vol. 4, Amsterdam, North Holland.

Hodgson, Douglas J. (1998a), "Adaptive Estimation of Cointegrating Regressions with ARMA Errors," *Journal of Econometrics* **85**: 231-268.

Hodgson, Douglas J. (1998b), "Adaptive Estimation of Error Correction Models," *Econometric Theory* **14**: 44-69.

Hodgson, Douglas J. (1999), "Adaptive Estimation of Cointegrated Models: Simulation Evidence and an Application to the Forward Exchange Market," *Journal of Applied Econometrics* **14**: 627-650.

Hodgson, Douglas J. (2000), "Partial Maximum Likelihood Estimation and Adaptive Estimation in the Presence of Conditional Heterogeneity of Unknown Form," *Econometric Reviews* **19**: 175-206.

Hodgson, Douglas J., Oliver Linton, and Keith Vorkink (2002), "Testing the Capital Asset Pricing Model Efficiently under Elliptical Symmetry: A Semiparametric Approach," *Journal of Applied Econometrics* **17**: 617-639.

Hsieh, David A., and Charles F. Manski (1987), "Monte Carlo Evidence on Adaptive Maximum Likelihood Estimation of a Regression," *Annals of Statistics* **15**: 541-551.

Jeganathan, Pradeep (1995), "Some Aspects of Asymptotic Theory with Applications to Time Series Models," *Econometric Theory* **11**: 818-887.

Johansen, Søren (1988), "Stochastic Analysis of Cointegration Vectors," *Journal of Economic Dynamics and Control* **12**: 231-254.

Levine, Ross (1989), "The Pricing of Forward Exchange Rates," *Journal of International Money and Finance* **8**: 163-179.

Levine, Ross (1991), "An Empirical Inquiry into the Nature of the Forward Exchange Rate Bias," *Journal of International Economics* **30**: 359-369.

Linton, Oliver (1995), "Second Order Approximations in a Partially Linear Regression Model," *Econometrica* **63**: 1079-1113.

Mardia, Kanti V. (1970), "Measures of Multivariate Skewness and Kurtosis with Applications," *Biometrika* **57**: 519-530.

Maynard, Alex, and Peter C.B. Phillips (2001), "Rethinking an Old Empirical Puzzle: Econometric Evidence on the Forward Discount Anomaly," *Journal of Applied Econometrics* **16**: 671-708.

Muirhead, Robb J. (1982), *Aspects of Multivariate Statistical Theory*, New York, Wiley.

Phillips, Peter C.B. (1995), "Robust Nonstationary Regression," *Econometric Theory* **11**: 912- 951.

Phillips, Peter C.B., and Bruce Hansen (1990), "Statistical Inference in

Instrumental Variables Regression with I(1) Processes," *Review of Economic Studies* **57**: 99-125.

Phillips, Peter C.B., and James W. McFarland (1997), "Forward Exchange Market Unbiasedness: The Case of the Australian Dollar since 1984," *Journal of International Money and Finance* **16**: 885-907.

Phillips, Peter C.B., James W. McFarland, and Patrick C. McMahon (1996), "Robust Tests of Forward Exchange Market Efficiency with Empirical Evidence from the 1920's," *Journal of Applied Econometrics* **11**: 1-22.

Schick, Anton (1987), "A Note on the Construction of Asymptotically Linear Estimators," *Journal of Statistical Planning and Inference* **16**: 89-105.

Schuster, Eugene F. (1985), "Incorporating Support Constraints into Nonparametric Estimators of Densities," *Communications in Statistics - Theory and Methods* **14**: 1123-1136.

Silverman, Bernie W. (1986), *Density Estimation for Statistics and Data Analysis*, London, Chapman & Hall.

Steigerwald, Douglas (1992), "Adaptive Estimation in Time Series Regression Models," *Journal of Econometrics* **54**: 251-275.

Stute, Winfried, and Wendelin Werner (1991), "Nonparametric Estimation of Elliptically Contoured Densities," in G. Roussas, ed., *Nonparametric Functional Estimation and Related Topics*, Dordrect, Kluwer Academic Publishers.

White, Halbert (1982), "Maximum Likelihood Estimation of Misspecified Models," *Econometrica* **50**: 1-25.

Zellner, Arnold (1962), "An Efficient Method for Estimating Seemingly Unrelated Regressions and Tests for Aggregation Bias," *Journal of the American Statistical Association* **57**: 348-368.