Adaptive Estimation of Cointegrated Models: Simulation Evidence and an Application to the Forward Exchange Market

Douglas J. Hodgson

Journal of Applied Econometrics, Vol. 14, No. 6. (Nov. - Dec., 1999), pp. 627-650.

Stable URL:
http://links.jstor.org/sici?sici=0883-7252%28199911%2F12%2914%3A6%3C627%3AAEOCMS%3E2.0.CO%3B2-Y

Journal of Applied Econometrics is currently published by John Wiley & Sons.

Your use of the JSTOR archive indicates your acceptance of JSTOR's Terms and Conditions of Use, available at http://www.jstor.org/about/terms.html. JSTOR's Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Please contact the publisher regarding any further use of this work. Publisher contact information may be obtained at http://www.jstor.org/journals/jwiley.html.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

The JSTOR Archive is a trusted digital repository providing for long-term preservation and access to leading academic journals and scholarly literature from around the world. The Archive is supported by libraries, scholarly societies, publishers, and foundations. It is an initiative of JSTOR, a not-for-profit organization with a mission to help the scholarly community take advantage of advances in technology. For more information regarding JSTOR, please contact support@jstor.org.

http://www.jstor.org
Wed Mar 19 06:17:34 2008

JOURNAL OF APPLIED ECONOMETRICS
J. Appl. Econ. **14**: 627–650 (1999)

ADAPTIVE ESTIMATION OF COINTEGRATED MODELS: SIMULATION EVIDENCE AND AN APPLICATION TO THE FORWARD EXCHANGE MARKET

DOUGLAS J. HODGSON*
Department of Economics, University of Rochester, Rochester, NY 14627, USA

SUMMARY

The paper reports simulation and empirical evidence on the finite-sample performance of adaptive estimators in cointegrated systems. Adaptive estimators are asymptotically efficient, even when the shape of the likelihood function is unknown. We consider two representations of cointegrated systems—triangular cointegrating regressions and error correction models. The motivation for and advantages of adaptive estimators in such systems are discussed and their construction is described. We report results from the estimation of a forward exchange market unbiasedness regression using the adaptive and competing estimators, and provide related Monte Carlo simulation evidence on the performance of the estimators.
Copyright © 1999 John Wiley & Sons, Ltd.

1. INTRODUCTION

The past decade has seen the development of a fairly complete theory of estimation in cointegrated systems, both in their triangular representations (e.g. Phillips and Hansen, 1990) and their error correction representations (e.g. Johansen, 1988). Understandably, attention was initially focused on the development of estimators that would be asymptotically efficient under the assumption that the data were driven by an underlying sequence of iid innovations with a normal, or Gaussian, distribution. The leading estimators of triangular cointegrating regressions, such as FM-OLS (Phillips and Hansen, 1990), canonical cointegrating regressions (Park, 1992), and the 'leads and lags' estimators (Saikkonen, 1991; Phillips and Loretan, 1991; Stock and Watson, 1993), can all be thought of as Gaussian maximum likelihood estimators, at least asymptotically. Similarly, the theory of Gaussian maximum likelihood estimation of error correction models is well developed (Johansen, 1988; Ahn and Reinsel, 1990).[1]

Many of the economic time series we observe whose autoregressive representations are well approximated by a unit root model, and which we may desire to include in a cointegrating regression, are financial series such as interest rates, stock prices, and exchange rates. These

* Correspondence to: D. J. Hodgson, Department of Economics, University of Rochester, Rochester, NY 14627, USA. E-mail: dshn@troi.cc.rochester.edu

Contract/grant sponsor: Alfred P. Sloan Foundation.
Contract/grant sponsor: Social Sciences and Humanities Research Council of Canada.
Contract/grant sponsor: NSF; Contract/grant number: SBR-9701959.

[1] For further discussion of the estimators referred to in this paragraph, and to the topic of efficient Gaussian estimation in cointegrated systems, see Phillips (1991), Watson (1994), and Hamilton (1994).

CCC 0883-7252/99/060627–24$17·50
Received 12 September 1995
Revised 4 June 1998
Copyright © 1999 John Wiley & Sons, Ltd.

variables are known to be characterized by a high degree of non-Gaussianity in their first differences, especially when sampled at weekly or daily frequencies, with the non-Gaussianity being primarily attributable to heavy tails (i.e. excess kurtosis). The presence of non-Gaussianity in the innovations to a cointegrated system will render the estimators cited in the previous paragraph inefficient, asymptotically, and motivates the development and employment of estimators that are robust to thick tails and that are, ideally, fully efficient.

A recent literature on robust and efficient estimation in non-Gaussian cointegrated models has developed in response to the empirical phenomenon of excess kurtosis noted above. Phillips (1995) develops variants of robust LAD and M-estimators that are suitable for inference in triangular cointegrating regressions and which he calls 'fully modified' LAD (FM-LAD) and 'fully modified' M (FM-M) estimators. The FM-LAD estimator is implemented in an empirical study of the foreign exchange market by Phillips, McFarland, and McMahon (1996). Fully asymptotically efficient adaptive estimators have been developed for triangular cointegrated models by Jeganathan (1995) and Hodgson (1998a), and for error correction models by Hodgson (1998b). These estimators are asymptotically equivalent to the maximum likelihood estimator, but are constructed without an assumption of knowledge of the distribution of the iid innovations to the model, and hence of the shape of the likelihood function, on the part of the investigator. The estimators are semiparametric and utilize non-parametric kernel density estimates of the unknown density function of the innovations. The essential idea is to construct a one-step iterative estimator, beginning from some consistent preliminary estimator of the model's parameters, and to use the residuals from this preliminary estimator to compute the non-parametric density estimate. The idea of adaptive estimation is due to Stein (1956), and the basic methodology was developed by Stone (1975) for location models and extended to a fuller range of models in papers by Bickel (1982), Kreiss (1987b), Steigerwald (1992a), Linton (1993), and others.

The present paper attempts to assess the performance of the adaptive estimators of cointegrated models developed by Hodgson (1998a,b) in finite samples through a Monte Carlo simulation study and through an empirical application to the estimation of a forward exchange market unbiasedness model similar to that estimated by Phillips, McFarland, and McMahon (1996). In Section 2, we introduce and briefly describe the adaptive estimators to be employed in the subsequent empirical and simulation study. In Section 3, we describe our forward unbiasedness model and report the results of estimating this model by several methods, including adaptive estimation. Section 4 reports the results of a Monte Carlo simulation study and Section 5 concludes.

2. THE MODELS AND ADAPTIVE ESTIMATORS

This section summarizes the methodology for computing adaptive estimators as developed in Hodgson (1998a) for triangular cointegrating regressions and in Hodgson (1998b) for error correction models. In subsections 2.1 and 2.2, respectively, we present the models to be estimated and provide recipes for the construction of adaptive estimators. Subsection 2.3 discusses the asymptotic properties of the estimators and the computation of standard errors and test statistics.

Copyright © 1999 John Wiley & Sons, Ltd.

2.1. Triangular Cointegrating Regressions

We assume that a single cointegrating relationship exists among $m + 1$ observed time series, each of which is I(1), and that the deviations of the variables from the cointegrating relationship follow an ARMA(p, q) process. For every $t = 1, \ldots, n$, we have

$$Y_t = B_1 + B_0' X_t + u_t \tag{1}$$

$$X_t = X_{t-1} + v_t \tag{2}$$

$$u_t = \sum_{j=1}^{p} a_j u_{t-j} + \sum_{j=1}^{q} b_j \varepsilon_{t-j} + \varepsilon_t \tag{3}$$

where X_t and B_0 are m-vectors, B_1 is an intercept, and $(\varepsilon_t, v_t')'$ is an iid sequence from the unknown elliptically symmetric density $p(\varepsilon, v)$. We denote the vector of ARMA coefficients by $\eta = (a_1, \ldots, a_p;\ b_1, \ldots, b_q)'$, the cointegrating vector by $B = [B_1, B_0']'$, the $p + q + m + 1$-dimensional full parameter vector by $\theta = (\eta', B')'$, and a $p + q + m + 1$-dimensional scaling matrix by $\delta_n = \mathrm{diag}[n^{-1/2} I_{p+q+1}, n^{-1} I_m]$. We define $\psi(\varepsilon, v) = (\partial p(\varepsilon, v)/\partial\varepsilon)/p(\varepsilon, v)$ and $\lambda^2 = \iint \psi(\varepsilon, v)^2 p(\varepsilon, v)\, \mathrm{d}v\, \mathrm{d}\varepsilon$, and assume that $0 < \lambda^2 < \infty$. If one knew the functional form of the error density $p(\varepsilon, v)$, then it would be straightforward to compute an asymptotically efficient estimate of the parameter vector θ. One could use an iterative approach such as Newton's method, in which the score function $\psi(\varepsilon, v)$ and information λ^2 would enter into the calculations. The problem that adaptive estimation seeks to address is that the error density $p(\varepsilon, v)$ may be *unknown* to the investigator. If the density is unknown, then so will be the score $\psi(\varepsilon, v)$ and information λ^2, in which case we seemingly cannot compute an efficient iterative estimator of θ. However, Hodgson (1998a) shows that it generally is possible to efficiently, or *adaptively*, estimate θ. To do so, we must obtain consistent non-parametric estimators of $\psi(\varepsilon, v)$ and λ^2, which can then be utilized in the construction of an iterative estimator. We briefly provide a step-by-step algorithm for computing an adaptive estimator in the remainder of this subsection.

Step 1: Estimate equation (1) by one of the usual cointegration estimators (e.g. OLS, FM-OLS, CCR, etc.) to obtain parameter estimates B_1^* and B_0^* and compute the associated residuals u_t^* for $t = 1, \ldots, n$. Now fit an ARMA model to $\{u_t^*\}$ and obtain $\sqrt{n}$-consistent estimates $\eta^* = (a_1^*, \ldots, a_p^*;\ b_1^*, \ldots, b_q^*)'$, using, for example, Gaussian MLE. Define the vector of preliminary estimates by $\theta^* = (\eta^{*'}, B_1^*, B_0^{*'})'$. For every $t = 1, \ldots, n$, compute the estimated innovations

$$\varepsilon_t^* = u_t^* - \sum_{j=1}^{p} a_j^* u_{t-j}^* - \sum_{j=1}^{q} b_j^* \varepsilon_{t-j}^*$$

(set $\varepsilon_s^* = u_s^* = 0$ for $s \leqslant 0$).

Step 2: Run an OLS regression of ε_t^* on $v_t = \Delta X_t$ and define the OLS residuals z_t^*. Now, for every $t = 1, \ldots, n$, compute the following kernel density estimate:

$$\hat{f}_{\sigma,t}(z_t^*, v_t) = (1/2(n-1)) \sum_{i=1, i\neq t}^{n} \{\pi(z_t^* + z_i^*, v_t + v_i, \sigma) + \pi(z_t^* - z_i^*, v_t + v_i, \sigma)\}$$

where we use the Gaussian kernel

$$\pi(z, v, \sigma) = (1/(\sigma\sqrt{2\pi})^{m+1})\exp(-(|z|^2 + |v|^2)/2\sigma^2)$$

where $|\cdot|$ denotes the Euclidean norm. (Note that σ is a bandwidth, or smoothing, parameter.) For every $t = 1, \ldots, n$, we also compute the following kernel estimate of the first derivative of this density:

$$\hat{f}'_{\sigma,t}(z_t^*, v_t) = (1/2(n-1)) \sum_{i=1, i\neq t}^{n} \{\pi'(z_t^* + z_i^*, v_t + v_i, \sigma) + \pi'(z_t^* - z_i^*, v_t + v_i, \sigma)\}$$

where $\pi'(z, v, \sigma) = (-z/\sigma^2)\pi(z, v, \sigma)$ is the first derivative of $\pi(z, v, \sigma)$ with respect to z.

Step 3: For every $t = 1, \ldots, n$, compute the following ratio:

$$\hat{\psi}_t(z_t^*, v_t) = \begin{cases} \dfrac{\hat{f}'_{\sigma,t}(z_t^*, v_t)}{\hat{f}_{\sigma,t}(z_t^*, v_t)} & \text{if } \begin{cases} \hat{f}_{\sigma,t}(z_t^*, v_t) \geqslant m_n \\ |(z_t^*, v_t)| \leqslant \alpha_n \\ |\hat{f}'_{\sigma,t}(z_t^*, v_t)| \leqslant c_n \hat{f}_{\sigma,t}(z_t^*, v_t) \end{cases} \\ 0 & \text{otherwise} \end{cases}$$

where $c_n \to \infty$, $\alpha_n \to \infty$, $\sigma \to 0$, and $m_n \to 0$. We discuss the selection of these smoothing and trimming constants in Sections 3 and 4. Now also compute

$$\hat{\lambda}^2 = n^{-1} \sum_{t=1}^{n} \hat{\psi}_t(z_t^*, v_t)^2$$

Step 4: For every $s = 0, 1, \ldots, n-1$, compute γ_s^* by the following formula:

$$\gamma_s^* + b_1^*\gamma_{s-1}^* + \cdots + b_q^*\gamma_{s-q}^* = 0 \; \forall s \geqslant 1$$

setting $\gamma_s^* = 0 \; \forall s < 0$ and $\gamma_0^* = 1$. Now, for every $t = 1, \ldots, n$, compute the following vectors:

$$Z_{t-1}^* = \sum_{k=0}^{t-1} \gamma_k^*(u_{t-1-k}^*, \ldots, u_{t-p-k}^*; \varepsilon_{t-1-k}^*, \ldots, \varepsilon_{t-q-k}^*)'$$

$$\Gamma_{t-1}^* = \sum_{j=0}^{t-1} \gamma_j^* \left[X_{t-j} - \sum_{k=1}^{p} a_k^* X_{t-j-k} \right] - v_t$$

$$H_{t-1}^* = (Z_{t-1}^{*\prime}, \Gamma_{t-1}^{*\prime})'$$

$$W_n^* = -\sum_{t=1}^{n} \delta_n H_{t-1}^* \hat{\psi}_t(z_t^*, v_t)$$

and compute the matrix

$$S_n^* = \hat{\lambda}^2 \sum_{t=1}^{n} \delta_n H_{t-1}^* H_{t-1}^{*\prime} \delta_n$$

Copyright © 1999 John Wiley & Sons, Ltd.

Step 5: Now compute the following adaptive estimator:

$$\tilde{\theta} = \theta^* + \delta_n^{-1} S_n^{*-1} W_n^*$$

2.2. The Error Correction Model

The procedure employed in obtaining an adaptive estimator for the error correction model will be seen to be similar to that for the triangular model, the chief differences being in the form of the model's score function and information matrix, and in the fact that we must non-parametrically estimate the entire vector-valued score function for the multivariate innovation density, rather than just its first element. The development in this subsection is based on Hodgson (1998b).

Allowing $\{X_t\}_{t=1}^n$ to be a q-vector of I(1) time series with r cointegrating relationships, and assuming it has a vector autoregressive representation of known order k, we may write the following error correction representation:

$$\Delta X_t = \pi_0 + ABX_{t-1} + \sum_{j=1}^{k-1} \Phi_j \Delta X_{t-j} + \varepsilon_t \tag{4}$$

where A is a $q \times r$ matrix of error correction coefficients, B is an $r \times q$ matrix whose rows are cointegrating vectors, and $\{\varepsilon_t\}$ are iid from the unknown density $p(\varepsilon)$, the negative of whose q-dimensional score vector we denote by $\psi(\varepsilon) = (\partial p(\varepsilon)/\partial\varepsilon)/p(\varepsilon)$, and whose finite, positive-definite $q \times q$ information matrix we denote by $\Omega = \int \psi(\varepsilon)\psi(\varepsilon)' p(\varepsilon)\, \mathrm{d}\varepsilon$. We also assume that $\pi_0 = -AB_1$, where B_1 is the r-vector of intercepts in the cointegrating vectors.

We assume that the model is identified, and follow Ahn and Reinsel (1990) by partitioning X_t as $[X'_{1t}, X'_{2t}]'$, where X_{1t} has r elements, X_{2t} has $q-r$ elements with $q-r$ unit roots, $B = [I_r, -B_0]$, and the $r \times (q-r)$ matrix B_0 contains the model's cointegrating coefficients. We can rewrite equation (4) as

$$\Delta X_t = A[X_{1,t-1} - B_1 - B_0 X_{2,t-1}] + \Phi Z_{t-1} + \varepsilon_t \tag{5}$$

where $\Phi = [\Phi_1, \ldots, \Phi_{k-1}]$ and $Z_{t-1} = [\Delta X'_{t-1}, \ldots, \Delta X'_{t-k+1}]'$. We define $\alpha = \mathrm{vec}(A)$, $\varphi = \mathrm{vec}(\Phi)$, and $\beta = \mathrm{vec}(B_0)$ (where the vec operator stacks the transposed rows of a matrix into a vector), which we then gather into the m-dimensional full parameter vector $\theta = [\alpha', \varphi', B'_1, \beta']'$, where $m = 2qr - r^2 + q^2(k-1) + r$. Defining $s = qr + q^2(k-1)$, the number of parameters in α and φ, the stationary component of the model, we can then introduce the scaling matrix $\delta_n = \mathrm{diag}[n^{-1/2} I_s, n^{-1/2} I_r, n^{-1} I_{m-s}]$. We now describe the adaptive estimation of θ.

Step 1: Obtain a δ_n^{-1}-consistent estimator θ^*. There are many possible ways to do this, including the Gaussian ML approach of Johansen (1988) or Ahn and Reinsel (1990). Perhaps the simplest way in practice is to estimate B_1 and B_0 through an OLS regression of X_{1t} on a constant vector and X_{2t}, plug the residuals $\Upsilon^*_{t-1} = X_{1,t-1} - B_1^* - B_0^* X_{2,t-1}$ from this regression into equation (5), and then estimate A and Φ through an OLS regression of ΔX_t on Υ^*_{t-1} and Z_{t-1}. In any event, once we have computed θ^*, plug the estimates into equation (5) to obtain the estimated residuals ε_t^*.

Step 2: For every $t = 1, \ldots, n$, compute the density estimate

$$\hat{p}_{\sigma,t}(\varepsilon_t^*) = (1/2(n-1)) \sum_{i=1, i\neq t}^{n} \{\pi(\varepsilon_t^* + \varepsilon_i^*, \sigma) + \pi(\varepsilon_t^* - \varepsilon_i^*, \sigma)\}$$

Copyright © 1999 John Wiley & Sons, Ltd.

where

$$\pi(\varepsilon) = 1/(\sigma\sqrt{2\pi})^q \exp(-|\varepsilon|^2/2\sigma^2).$$

Also compute the following derivative estimate for every $j = 1, \ldots, q$:

$$\hat{p}^j_{\sigma,t}(\varepsilon^*_t) = (1/2(n-1)) \sum_{i=1, i\neq t}^{n} \{\pi^j(\varepsilon^*_t + \varepsilon^*_i, \sigma) + \pi^j(\varepsilon^*_t - \varepsilon^*_i, \sigma)\}$$

where $\pi^j(\varepsilon) = (-\varepsilon^j/\sigma^2)\pi(\varepsilon)$ is the jth element of vector of partial derivatives of $\pi(\varepsilon)$ with respect to ε and ε^j is the jth element of ε. Now, for every $j = 1, \ldots, q$, compute the following ratio:

$$\hat{\psi}^j_t(\varepsilon^*_t) = \begin{cases} \dfrac{\hat{p}^j_{\sigma,t}(\varepsilon^*_t)}{\hat{p}_{\sigma,t}(\varepsilon^*_t)} & \text{if} \begin{cases} \hat{p}_{\sigma,t}(\varepsilon^*_t) \geqslant m^j_n \\ |\varepsilon^*_t| \leqslant \alpha^j_n \\ |\hat{p}^j_{\sigma,t}(\varepsilon^*_t)| \leqslant c^j_n \hat{p}_{\sigma,t}(\varepsilon^*_t) \end{cases} \\ 0 & \text{otherwise} \end{cases}$$

where $c^j_n \to \infty$, $\alpha^j_n \to \infty$, and $m^j_n \to 0$ for every $j = 1, \ldots, q$. We then have $\hat{\psi}_t(\varepsilon^*_t) = (\hat{\psi}^1_t(\varepsilon^*_t), \ldots, \hat{\psi}^q_t(\varepsilon^*_t))'$. Use this estimated score vector to compute the following information matrix estimate:

$$\hat{\Omega}_n = n^{-1} \sum_{t=1}^{n} \hat{\psi}_t(\varepsilon^*_t)\hat{\psi}_t(\varepsilon^*_t)'$$

Step 3: Compute the following matrices:

$$H^*_{t-1} = [(I_q \otimes \Upsilon^*_{t-1})', (I_q \otimes Z_{t-1})', -A^*, (-A^* \otimes X_{2,t-1})']'$$

$$W^*_n = -\sum_{t=1}^{n} \delta_n H^*_{t-1} \hat{\psi}_t(\varepsilon^*_t)$$

and

$$S^*_n = \sum_{t=1}^{n} \delta_n H^*_{t-1} \hat{\Omega}_n H^{*\prime}_{t-1} \delta_n$$

Finally, compute the adaptive estimate

$$\tilde{\theta} = \theta^* + \delta_n^{-1} S_n^{*-1} W^*_n$$

2.3. Properties and Tests

For both of the models discussed above, the resulting estimator $\tilde{\theta}$ is fully asymptotically efficient and has the asymptotic distribution

$$\delta_n^{-1}(\tilde{\theta} - \theta) \Rightarrow MN(0, S(\theta)^{-1})$$

Copyright © 1999 John Wiley & Sons, Ltd.

where $\Rightarrow$ denotes weak convergence of probability measures, MN denotes mixed normal, and $S(\theta)$ is the (random) asymptotic information matrix, which can be explicitly computed for either of the above models by noting that $S_n^* \Rightarrow S(\theta)$ for both models. Note that although the asymptotic covariance matrix will be random, we can still use our covariance matrix estimator S_n^* in the usual fashion to compute Wald tests and t-ratios which will have the usual chi-squared and standard normal limit theory, respectively.

This mixed normal limit theory for the efficient estimators in both of our models is associated with the fact that the asymptotic covariance matrices are random. This randomness is a standard feature of cointegrated models — the familiar Gaussian pseudo-MLEs, or their equivalents, such as FM-OLS (Phillips and Hansen, 1990) for the triangular model, and reduced rank regression (Johansen, 1988) for the error correction model, also have random covariance matrices. This randomness suggests that it would be difficult, if not impossible, to make efficiency comparisons between the MLE and the Gaussian pseudo-MLE by considering a ratio of asymptotic variances (or of determinants of asymptotic covariance matrices), as is easily done in standard models with deterministic asymptotic covariance matrices. However, note that, for both of the models considered here, the asymptotic covariance matrices of both the MLE and the Gaussian pseudo-MLE can be written as the product of a random and a non-random component, and that, furthermore, the random component is the same for both estimators, so that an asymptotic efficiency ratio *can* be obtained simply by comparing the non-random components of the respective estimators (Hodgson, 1998a,b). In fact, the asymptotic efficiency ratio between the MLE and the Gaussian pseudo-MLE for both the triangular and the error correction models essentially reduces to comparing the inverse of the information matrix of the innovation density with the covariance matrix of the innovations, so that the ratio turns out to be identical to that which would prevail if we wanted to compare the MLE with the sample mean in the estimation of the most basic multivariate location parameter model. A detailed analysis of this model when the innovation density is elliptically symmetric is provided by Mitchell (1989).

A natural question concerns the behaviour of the adaptive estimators when the leptokurtosis present in the unconditional density of the uncorrelated innovation process is due to unmodelled dependence or heterogeneity in higher-order moments of the process. For example, autoregressive conditional heteroscedasticity will induce thick tails in the unconditional density of an uncorrelated sequence of random variables even if the conditional densities are Gaussian (Engle, 1982). We would intuitively think that an adaptive estimator should control for the thick tails present in such a sequence, even if the sequence is not iid, since it employs a non-parametric kernel density estimator which should consistently estimate the unconditional density of the process. We could therefore 'guess' that the adaptive estimator would have a distribution theory that was the same as that of the pseudo-maximum likelihood estimator that we would compute if we assumed that the innovations were iid from their true unconditional density. These conjectures are found in Hodgson (1998, unpublished manuscript) to be correct, under certain conditions.

Hodgson (1998, unpublished manuscript) analyses a variety of stationary and non-stationary time series models with uncorrelated but possibly dependent innovations. The central result of the paper implies that if the innovation process at any point in time has a density that is symmetric about zero, conditional on the past of the process, then the asymptotic variance of the semiparametric iterative 'adaptive' estimator will be the same as that of the parametric iterative pseudo-MLE that assumes that the innovations are iid from their unconditional density. For most time series models, this asymptotic covariance matrix will have the 'sandwich' structure common to pseudo-ML estimators (White, 1982). Note that this situation also occurs for OLS in

Copyright © 1999 John Wiley & Sons, Ltd.

many time series models (Kuersteiner, 1997, unpublished manuscript). For estimators such as those described above, it would therefore be desirable to compute robust semiparametric 'sandwich'-type asymptotic covariance matrix estimates. However, such estimators have not yet been developed, as they require us to compute consistent non-parametric versions not only of the outer-product of gradients version of the information matrix of the innovation density but also of the Hessian of this density. Although estimators of the former are available, we do not know of any results on the estimation of the latter. The development of such an estimator is a matter for further research.

The possibility of deriving semiparametric estimators that have desirable robustness properties when the innovations are *not* conditionally symmetric around zero (which would be the case, for example, if there were unmodelled autocorrelation present) has not yet been fully investigated. The problem that our adaptive estimators face in this context arises from the fact that they utilize non-parametric density estimators which are symmetric about zero by construction, so that the non-parametric score estimators $\hat{\psi}$ are anti-symmetric about zero. This anti-symmetry greatly facilitates the proof that the $\hat{\psi}$ are consistent estimators of the true scores ψ. However, when the innovation density is not conditionally symmetric, then the score estimators $\hat{\psi}$ based on symmetrized nonparametric density estimators will not consistently estimate ψ; furthermore, they will generally not have zero mean, so that adaptive estimators constructed using them will be inconsistent. It should be possible to derive consistent score estimators that do not rely on a symmetrized density estimator and so will be robust to conditional asymmetry in the data (see Kreiss, 1987a). This is a topic of further research.

3. FORWARD EXCHANGE MARKET UNBIASEDNESS

The extensive literature on forward market unbiasedness attempts to determine whether or not forward exchange rates are unbiased predictors of future spot rates. The question is of interest in terms of both its bearing on the question of the efficiency of speculative markets and its implications for the treatment of exchange rates in macroeconomic models. (The former aspect of the question is emphasized by many analysts, including, for example, Fama, 1984, Froot and Frankel, 1989, and Hansen and Hodrick, 1980; the later aspect is emphasized by McCallum, 1994; both aspects are discussed by Baillie and McMahon, 1989).

There are two alternative approaches to modelling and estimating forward exchange market unbiasedness models. One involves a levels cointegrating regression of the spot rate on the lagged forward rate, while the second involves specifying a single-equation error correction model in which the first difference of the spot rate is regressed on the lagged spread between the forward and spot rates. The latter approach typically proceeds under the *a priori* assumption that the slope parameter in the cointegrating regression is unity. The relationship between these two alternative (actually complementary) approaches is analysed by Hakkio and Rush (1989) and Barnhart and Szakmary (1991).

In this section our concern is with the estimation of the first model described in the preceding paragraph, i.e. the cointegrating regression of the logarithm of the spot rate on the logarithm of the lagged forward rate. The model can be written as follows:

$$s_{t+w} = B_1 + B_0 f_{t,w} + u_{t+w} \tag{6}$$

Copyright © 1999 John Wiley & Sons, Ltd.

where s_{t+w} is the logarithm of the spot exchange rate between two currencies at time $t+w$, $f_{t,w}$ is the logarithm of the w-period-ahead forward exchange rate observed at time t, and u_{t+w} is a possibly autocorrelated error term. The forward unbiasedness hypothesis posits zero intercept and unit slope, i.e. that $(B_1, B_0) = (0, 1)$.

In empirical applications, this model is typically estimated using ordinary least squares or some variant (see, for example, McCallum, 1984; Hakkio and Rush, 1989; Corbae, Lim and Ouliaris, 1992; Baillie and Bollerslev, 1989; Baillie, Lippens and McMahon, 1983; Barnhart and Szakmary, 1991). However, it is a well-documented fact that returns to speculative price series are poorly approximated by a normal distribution, and are usually found to be leptokurtic *vis-à-vis* the normal. The lack of robustness of least squares procedures in the presence of significant deviations from Gaussianity motivated the study of Phillips, McFarland and McMahon (1996), who investigate the application of the robust FM-LAD estimator of Phillips (1995) to the estimation of a forward unbiasedness regression.

In this section, we report the results of the estimation of the unbiasedness model (6) for the Canadian dollar's exchange rate with respect to the US dollar using data from the early 1990s. Our data consist of daily noon observations on the spot rate and 90-day forward rate. In the notation of model (6), we therefore have $w = 90$. We have 658 observations on each variable, with t running from 19 November 1990 to 30 June 1993. The data were obtained from the Bank of Canada and the appropriate spot rate was matched with each day's forward rate according to the procedure described in Cornell (1989). The data are plotted in Figure 1. We note here that our use of data sampled more frequently than the forecast interval of 90 days implies that the error term u_{t+w} in model (6) will be autocorrelated, following an MA($w-1$) process. Due to the existence of weekends and holidays, w will not equal 90 but will generally be some smaller number (on average, it will be 66, since the number of business days in a month averages 22).

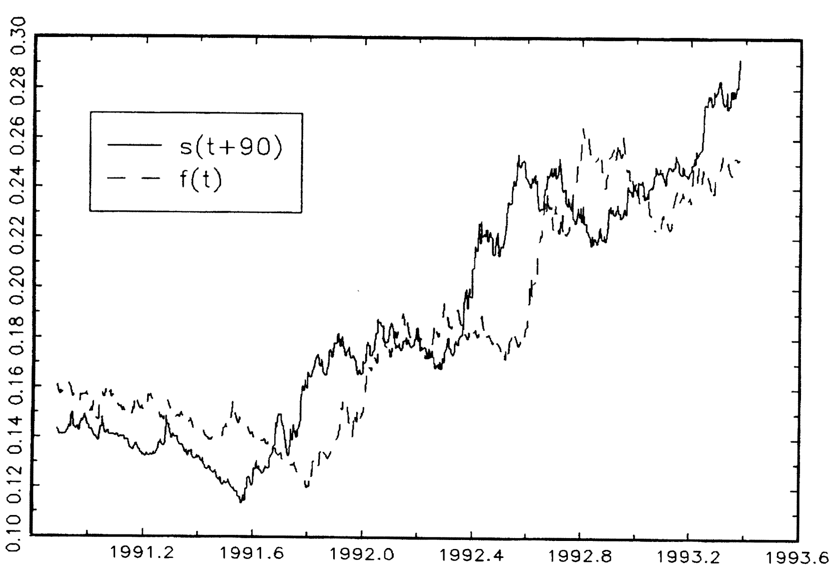

Figure 1. Logarithms of spot and forward rates

Copyright © 1999 John Wiley & Sons, Ltd.

We estimate the cointegration vector within both a triangular representation and an error correction model, with an emphasis on comparing results obtained through the employment of the leading Gaussian pseudo-ML estimators with those obtained using the adaptive estimators described in the preceding section. We hope that the illustration of adaptive estimation in action will be of interest in its own right. Adaptive estimators have appeared mainly in the theoretical statistics and econometrics literature, an exception being Steigerwald (1992b). Several questions arise regarding the empirical implementation of adaptive estimators. These questions are addressed, discussed, and illustrated in the analysis reported below.

3.1. Error Correction Model

In this subsection, we discuss the results of estimating the forward unbiasedness equation (6) within the context of an error correction representation. We estimate the cointegration parameters using the reduced rank regression estimator of Johansen (1988) and the adaptive estimator of Hodgson (1998b). These estimators have the advantage over the triangular estimators considered in the following subsection of being full information maximum likelihood estimators when the appropriate distributional assumptions on the innovations hold—Gaussianity for the Johansen (1988) estimator and symmetry for the Hodgson (1998b) estimator—whereas the triangular model estimators are limited information MLEs and will hence not be fully efficient.

We shall discuss some details of the implementation of the estimators before turning to our results. First, there is the familiar issue of specifying the lag length k in the VAR model from which equation (4) is derived. The presence of moving average effects in the error correction term u_{t+w} implies that k should be infinite, so we must technically think of k as some truncation parameter that increases to infinity with the sample size (although the theory in Hodgson, 1998b, does not allow for this, it would be a straightforward extension to allow it to do so—see also Steigerwald, 1992a, and Jeganathan, 1997). The value of k that minimizes the BIC criterion for our data set is $k = 1$, so in our empirical application we shall estimate the model using this setting; we shall also set $k = 8$, as a check on the robustness of our results to specification.

In all cases, the relevant Johansen (1988) reduced rank regression estimator serves as the preliminary estimator of the cointegrating vector in the computation of the adaptive estimator, with OLS being used to compute the preliminary estimates of all other parameters. In order to implement the adaptive estimator, we must specify values for the bandwidth parameter σ and the trimming parameters c_n, α_n, and m_n (we use the same trimming values for each equation, since the standard deviations of the two residual sequences are similar in our data, and so omit the j superscripts used above). We reduce the trimming and parameter selection to a univariate one in a manner similar to that described in Hsieh and Manski (1987). Denoting by $\hat{\mu}^2$ the average of the estimated variances of our residual sequences, and defining $h = \tilde{h}\hat{\mu}$, where $\tilde{h}$ is some constant chosen by the investigator, we set $\alpha_n = h$, $c_n = h/\hat{\mu}^2$, and $m_n = \exp\{-h^2/\hat{\mu}^2\}$. Using this approach, each parameter trims at roughly $\tilde{h}$ standard deviations from the origin. In practice, we set $\tilde{h} = 10$. We do not investigate the sensitivity of our results to variation in $\tilde{h}$ here, as results in Hsieh and Manski (1987) and Hodgson (1995) indicate little sensitivity in the point estimates to such variation (although there can be sensitivity in the standard error estimates).

Two methods are employed to select a value for the bandwidth parameter σ. First, we compute the rule-of-thumb (ROT) bandwidth described by Silverman (1986). This is an automatic procedure which selects that value of the bandwidth which would minimize the mean integrated

Copyright © 1999 John Wiley & Sons, Ltd.

squared error of the non-parametric kernel estimator of the innovation density if this density were Gaussian. This optimality criterion is admittedly different from the one that is of interest to us, which is to produce an estimator of the regression parameters that is in some sense optimal (according to a mean-squared error criterion, for example). However, since this bandwidth will generally deliver a fairly good density estimate, we have reason to think that it will deliver a reasonably good adaptive estimate and that it is in some neighbourhood of the bandwidth that would be optimal according to our criterion. This consideration, along with its ease of computation, makes the ROT an appealing choice, at least as a preliminary benchmark.

A second approach is based on the bootstrap method advocated by Hsieh and Manski (1987). They generate N bootstrap samples, and compute an adaptive estimate of the parameter of interest for each sample using each of a grid of J preselected bandwidths. For each bandwidth, they then compute the MSE over the N samples, and finally select an optimum bandwidth by choosing the one which is associated with the minimum MSE in the grid or the minimum MSE obtained through fitting a quadratic to the grid, whichever is smaller. They use no systematic criterion for selecting the range over which to form their grid. We modify the Hsieh and Manski (1987) approach principally through the addition of such a criterion. In particular, we choose a grid which is evenly spaced over a neighbourhood of the ROT bandwidth ranging from approximately half the ROT value to approximately double it. For our model, with $k = 1$ and no intercept, the ROT bandwidth is 0·000864. We considered a grid of candidate bandwidths ranging from 0·00044 to 0·00164, spaced by 0·0001. Hence, we have $J = 13$. For our bootstrap Monte Carlo, we set $B_0 = 1$, set other parameters equal to their estimated values, and drew the Monte Carlo innovations from the empirical distribution of the estimated residuals. The number of Monte Carlo samples was $N = 500$. (It took approximately 4 days on a Pentium 150 to conduct this simulation.) Over this grid, the MSE was minimized at 0·00104, and was monotonically increasing in either direction, suggesting that we may get a better-performing estimator, for our sample at any rate, by selecting a bandwidth that is greater than the ROT by approximately 0·0002. We follow this course in our empirical and simulation analysis, throughout adaptively estimating the model using bandwidths of ROT and ROT + 0·0002.

The results of our empirical analysis are presented in Tables I and II and in Figures 2–5.[2] Before discussing the estimates reported in Table I, we shall make some comments regarding the specification of the model. First, the Box and Pierce (1970) autocorrelation diagnostics reported in Table II suggest that our setting of $k = 1$ as the number of lags in the vector error correction model adequately captures the autocorrelation present in the sample. None of the reported statistics are close to being significant at the 10% level, most are not even significant at the 25% level, and extending the lag length to $k = 8$ affects the value of the statistics only very slightly. This information seems to be confirmed by Figures 2 and 3, which plot the residuals from an estimate of the error correction model with $k = 1$.[3] The Jarque–Bera (1980) statistics reported in Table II strongly reject the null hypothesis of Gaussianity for both residual series, a finding which is again completely insensitive to specification of k. We find excess kurtosis to be present in each series, a finding reflected in Figures 4 and 5, where non-parametric Gaussian

[2] The Johansen estimates were computed in GAUSS using the routine 'SJ' in the COINT package of Ouliaris and Phillips (1994), while all other computations were carried out using GAUSS routines programmed by the author. The latter are available from the author upon request.

[3] The residuals used in the computation of the statistics in Table II and the plotting of the graphs in Figures 2–5 were obtained by estimating the cointegrating parameters (including intercept) by reduced rank regression, and estimating all other parameters by OLS.

Copyright © 1999 John Wiley & Sons, Ltd.

Table I. Parameter estimates — ECM

	Model including intercept				Model excluding intercept	
	$k = 1$		$k = 8$		$k = 1$	$k = 8$
Estimator	$\hat{B}_1$	$\hat{B}_0$	$\hat{B}_1$	$\hat{B}_0$	$\hat{B}_0$	$\hat{B}_0$
Johansen	−0·0103	0·985	−0·00769	0·977	0·931	0·937
	(0·034)	(0·184)	(0·028)	(0·149)	(0·041)	(0·033)
Adaptive (ROT)	0·0214	0·885	0·00838	0·962	0·993	1·003
	(0·029)	(0·155)	(0·023)	(0·122)	(0·034)	(0·027)
Adaptive	0·0194	0·910	0·0160	0·916	1·009	0·997
(ROT + 0·0002)	(0·032)	(0·172)	(0·026)	(0·138)	(0·038)	(0·031)

Notes:
(1) The estimated standard errors are given in parentheses.
(2) The Silverman (1986) rule-of-thumb bandwidth is denoted ROT. When $k = 1$, ROT = 0·000864. When $k = 8$, ROT = 0·000861 when intercept included and 0·000869 when excluded.

Table II. Diagnostics — ECM

	$k = 1$		$k = 8$	
Statistic	$\hat{\varepsilon}_1$	$\hat{\varepsilon}_2$	$\hat{\varepsilon}_1$	$\hat{\varepsilon}_2$
Jarque–Bera kurtosis	53·51	64·93	52·11	64·22
Jarque–Bera normality	141·66	180·07	135·92	176·39
Box–Pierce ($q = 1$)	0·63	0·02	0·61	0·03
Box–Pierce ($q = 5$)	2·85	5·10	2·84	5·36
Box–Pierce ($q = 10$)	11·41	13·08	11·32	12·89
Box–Pierce ($q = 20$)	23·08	25·28	22·86	25·32

Notes:
(1) $\hat{\varepsilon}_1$ and $\hat{\varepsilon}_2$ denote the residuals to the first and second equations, respectively, where the cointegration parameters are estimated by the Johansen (1988) methodology with intercept, and all other parameters are estimated by OLS.
(2) The null distributions of the Jarque–Bera kurtosis and normality tests are N(0, 24) and chi-squared with two degrees of freedom, respectively, while for the Box–Pierce tests they are chi-squared with q degrees of freedom.

kernel density estimates are plotted (using the Silverman, 1986, ROT bandwidth) together with a Gaussian density whose variance equals the sample variance of the residuals. Looking at Figures 2 and 3, it appears as if there is conditional heteroscedasticity present in these series (volatility is relatively low early in the sample, but there seems to be a bout of high volatility in mid- to late 1992), which would certainly contribute to our finding of thick tails in the unconditional density. As mentioned in Section 2, it would be desirable to obtain standard error estimates that were robust to the conditional heteroscedasticity, but a way to compute such estimates has not yet been developed.

The first four columns of Table I report our estimates of the cointegrating vector when an intercept is included, and the slope estimates when the intercept is constrained to equal zero are reported in the final two columns. The first observation to make about the former set of estimates is that the intercepts are all quite close to zero, being approximately 0·02 or less in absolute value and always being well within one standard deviation of zero. The slope estimates are also all within one standard error of unity, but these standard errors are quite large, ranging from 0·122

Copyright © 1999 John Wiley & Sons, Ltd.

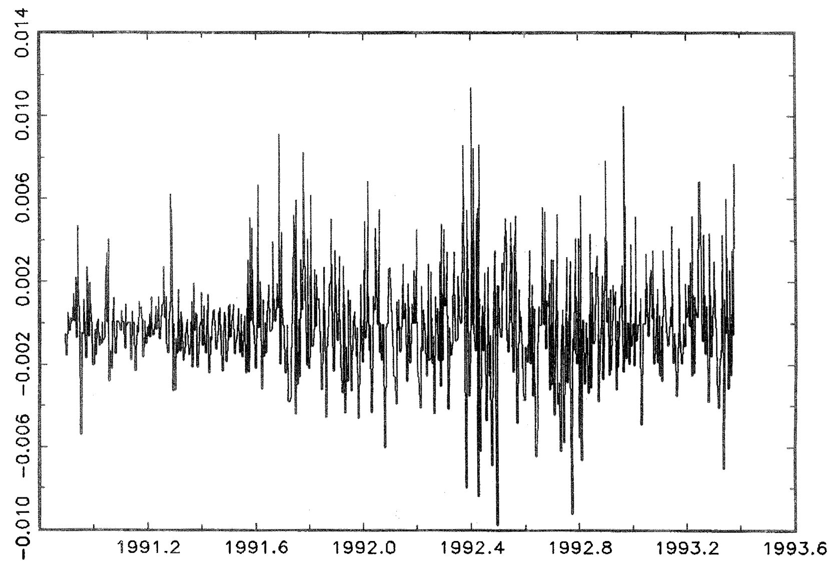

Figure 2. Residuals, ECM equation (1)

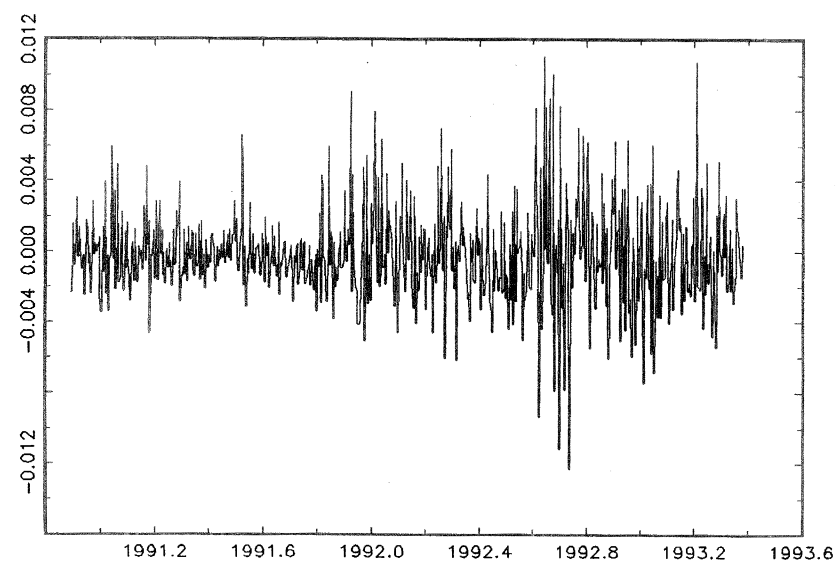

Figure 3. Residuals, ECM equation (2)

to 0·184, depending upon the estimator and the settings of k and σ. The Johansen estimate is quite invariant to lag length and is fairly close to unity, while the adaptive estimate varies in response to changing either k or σ, although the variation is small in magnitude relative to the estimated standard errors. The adaptive estimates are generally further from unity than the

Copyright © 1999 John Wiley & Sons, Ltd.

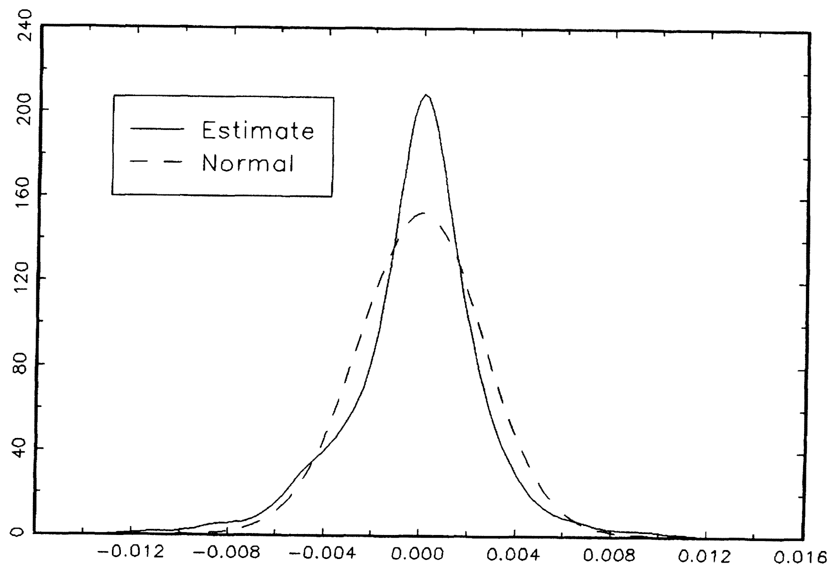

Figure 4. Density estimate of residuals, ECM equation (1)

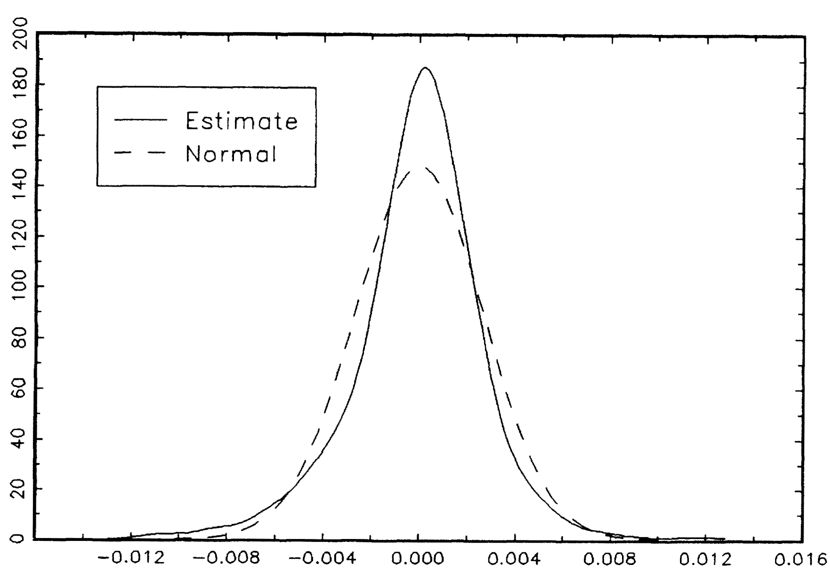

Figure 5. Density estimate of residuals, ECM equation (2)

Johansen estimates. When we exclude a constant from the regression, the estimated standard errors all fall considerably and the adaptive estimates are insensitive to variation in k and σ. The Johansen procedure produces slope estimates in the 0·93 range, nearly but not quite two standard errors from unity, while the adaptive estimates are almost right on unity.

3.2. Triangular Model

As mentioned earlier, the error process u_{t+w} to the cointegrating regression will follow a moving average process of high order (approximately 65) for our model. In the literature on efficient Gaussian estimation of the triangular representation of a cointegrated model, the possible autocorrelation present in the errors is generally handled in a non-parametric manner, it not being necessary to specify a parametric model of the model's stationary dynamics. In adaptively estimating a model, or in efficient estimation generally in the presence of non-Gaussianity, we know of no way to proceed other than to fit some parametric model to these dynamics. Our procedure in this section will be to fit an ARMA(1,1) model to the errors (a much more parsimonious parameterization than an MA(65)) and apply the estimator of Hodgson (1998a) described in Section 2. Most of the comments made in the preceding subsection concerning implementation of the adaptive estimator in error correction models will apply here as well. We shall compute the adaptive estimator for two bandwidths, ROT and ROT + 0·0002. We have not conducted a separate bootstrap exercise here to justify the choice of ROT + 0·0002.

Table III. Parameter estimates — triangular model

	Model with intercept		Without intercept
Estimator	$\hat{B}_0$	$\hat{B}_1$	$\hat{B}_0$
OLS	0·00289	1·009	1·024
FM-OLS	−0·0126	1·051	0·987
	(0·00635)	(0·0339)	(0·00741)
Adaptive (ROT)	0·00476	1·030	0·975
	(0·0108)	(0·0179)	(0·0167)
Adaptive (ROT + 0·0002)	0·00845	1·031	0·978
	(0·0120)	(0·0186)	(0·0186)
LAD	−0·041	1·213	1·008
FM-LAD			0·987
			(0·0689)

Notes:
(1) Standard errors are in parentheses.
(2) The Silverman (1986) ROT bandwidth was 0·00108 for the model with intercept and 0·00105 for the one without.
(3) For further details on computation of the estimators, see text.

Our results are reported in Tables III and IV and Figures 6–10.[4] Some of these results and a discussion of them can also be found in Hodgson (1998a). Figure 6 plots the FM-OLS residuals $\{\hat{u}_t\}$, illustrating the strong autocorrelation we know to be present. The estimated innovations $\{\hat{\varepsilon}_t\}$ obtained by fitting an ARMA(1,1) model to $\{\hat{u}_t\}$ are analysed — along with the regressor first differences $\{v_t\}$ — in Table IV, and illustrated in Figures 7–10. The Box–Pierce statistics in Table IV contain no evidence to contradict our specification of both of these series as being uncorrelated, nor does an inspection of the plots of the series in Figures 7 and 8. The Jarque–

[4] The FM-OLS estimates and standard errors were computed in GAUSS using the 'FM' routine (with Parzen kernel and automatic bandwidth) in the COINT package of Ouliaris and Phillips (1994), the LAD estimate for the model with intercept was computed using the 'qreg' routine in STATA (my gratitude to Michael Wolkoff for allowing me to use his computer for this calculation), and all remaining computations were carried out using GAUSS routines programmed by the author and available upon request. FM-LAD was computed using a Bartlett kernel and user-specified bandwidth.

Copyright © 1999 John Wiley & Sons, Ltd.

Table IV. Diagnostics — triangular model

Statistic	$\hat{\varepsilon}_1$	ν
Jarque–Bera kurtosis	37·34	67·93
Jarque–Bera normality	58·15	217·78
Box–Pierce ($q = 1$)	0·0248	$2{\cdot}04 \times 10^{-5}$
Box–Pierce ($q = 5$)	6·54	5·63
Box–Pierce ($q = 10$)	11·49	13·89
Box–Pierce ($q = 20$)	21·65	24·50

Notes:
(1) $\hat{\varepsilon}_1$ denotes the estimated innovations to the ARMA(1,1) error process, while ν denotes the regressor first differences.
(2) The test statistics have the same null distributions as stated in the notes to Table II.

Bera test statistics in Table IV and the density estimates plotted in Figures 9 and 10 also indicate the presence of thick tails in both series, although the magnitude of the non-Gaussianity is considerably less in $\{\hat{\varepsilon}_t\}$ than in $\{\nu_t\}$. A comparison of the tail thicknesses of these series with those of the two innovation series in the error correction model may be of interest; while the ARMA innovations here are *less* thick-tailed than either of the latter two series, the regressor first differences appear to be *more* thick-tailed than the latter.

The estimates reported in Table III lend fairly strong support to the unbiasedness hypothesis. For the model with intercept, the adaptive estimates are closer to the hypothesized values than the FM-OLS estimates; the intercept point estimates are closer to zero for both bandwidth settings and the slope estimate is adjusted from 1·05 to 1·03. We have not reported an FM-LAD estimate for the model with intercept because we found the estimates to be extremely sensitive to selection of the lag truncation parameter. As in the error correction model, the estimated standard errors decrease when the intercept is constrained to equal zero. The point estimates

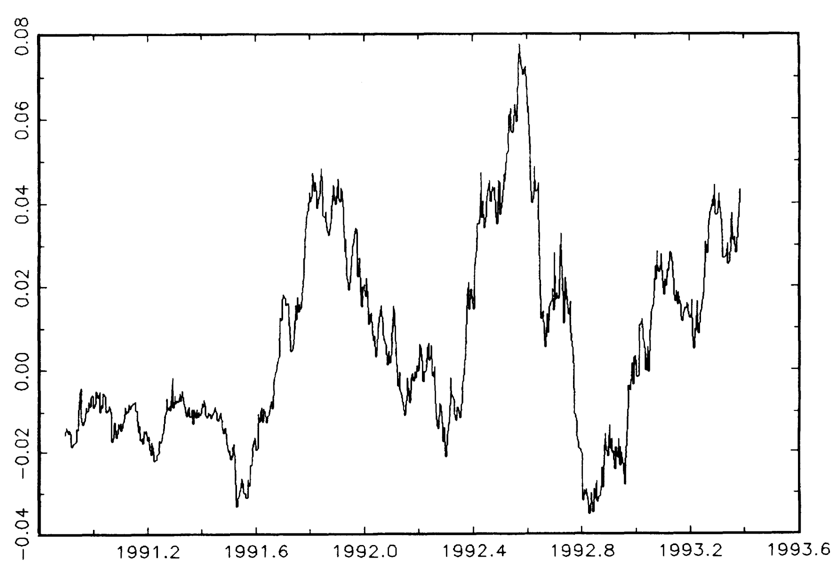

Figure 6. FM-OLS residuals

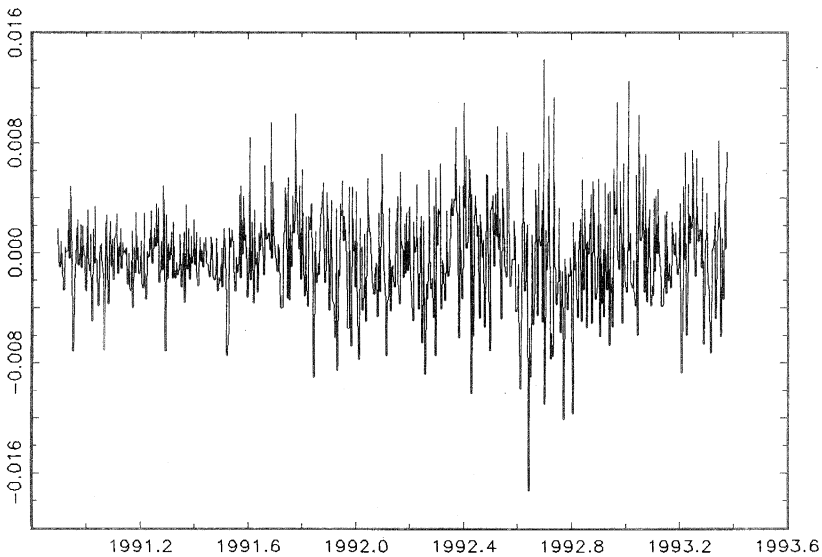

Figure 7. Residuals, ARMA(1,1) model

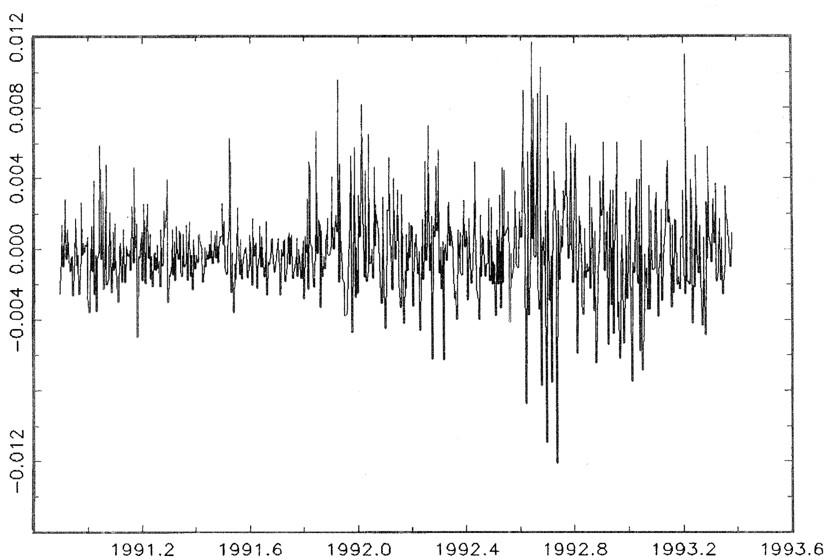

Figure 8. Regressor first differences, triangular model

themselves are quite similar for all estimators, including FM-LAD (the reported estimate uses a lag truncation parameter of 10). The relative insensitivity of our point estimates to the estimator employed (especially in comparison to the results we obtain for the error correction model) is likely due to the fact, alluded to above, that the innovations to the cointegrating regression are less thick-tailed than the innovations in the error correction model.

Copyright © 1999 John Wiley & Sons, Ltd.

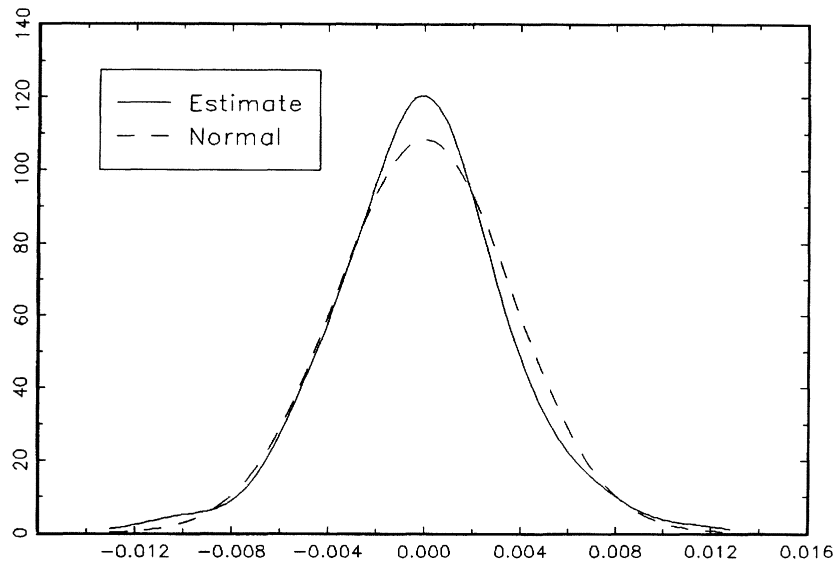

Figure 9. Density estimate, ARMA residuals

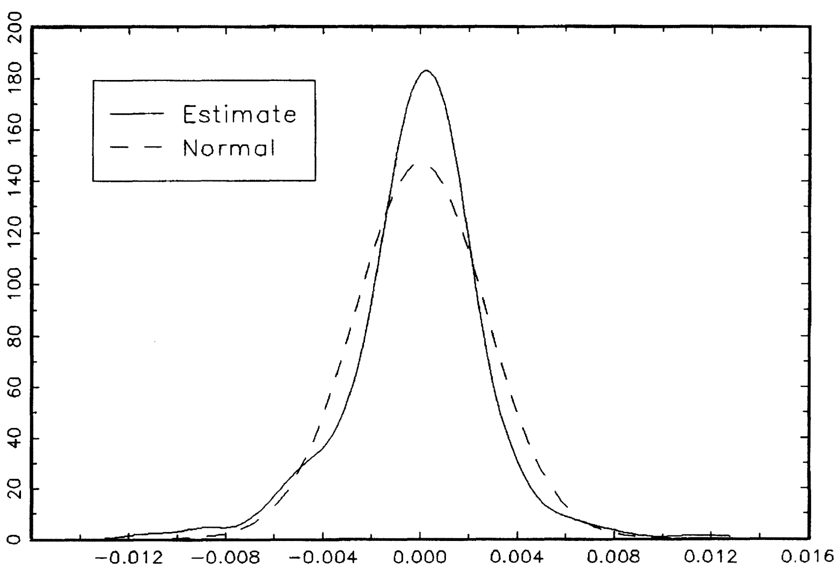

Figure 10. Density estimate, regressor first differences

4. MONTE CARLO SIMULATION EVIDENCE

We report the results of a Monte Carlo study of the finite-sample behaviour of the adaptive estimator in Tables V–VII. Since our empirical results indicate little change in inference for the triangular model when going from the Gaussian to the adaptive estimator, and to save space, we

Table V. Monte Carlo results — ECM. Student's t_ν errors ($k = 1$), estimation of slope

Estimator	Bias	MSE	Interdecile	Interquartile
Model with intercept ($\nu = 3$)				
Johansen	−0·0169	0·0275	0·710–1·452	0·835–1·129
Adaptive (ROT)	−0·0147	0·0264	0·721–1·446	0·837–1·128
Adaptive (ROT + 0·0002)	−0·0147	0·0264	0·721–1·451	0·837–1·127
Model without intercept ($\nu = 3$)				
Johansen	0·00749	0·00472	0·896–1·203	0·943–1·070
Adaptive (ROT)	0·00835	0·00455	0·900–1·200	0·947–1·067
Adaptive (ROT + 0·0002)	0·00842	0·00455	0·901–1·201	0·946–1·069
Model without intercept ($\nu = 5$)				
Johansen	0·00838	0·00193	0·935–1·113	0·967–1·050
Adaptive (ROT)	0·00837	0·00192	0·936–1·113	0·966–1·050
Adaptive (ROT + 0·0002)	0·00845	0·00192	0·936–1·113	0·967–1·050

Notes:
(1) 5000 iterations were conducted for each DGP.
(2) Bias and MSE are computed after discarding the least 500 and greatest 500 estimates in each case.
(3) The adaptive estimator uses the Johansen (1988) estimate as the preliminary estimate.
(4) The Silverman (1986) rule-of-thumb bandwidth is denoted by ROT.

Table VI. Monte Carlo results — ECM. Normal errors ($k = 1$), estimation of slope

Estimator	Bias	MSE	Interdecile	Interquartile
Johansen	0·00863	0·00510	0·894–1·194	0·943–1·073
Adaptive (ROT)	0·00859	0·00513	0·894–1·194	0·941–1·073
Adaptive (ROT + 0·0002)	0·00863	0·00510	0·894–1·195	0·942–1·073

Notes: See Table V.

confine our study to the estimation of the error correction model and consider a simulation DGP which mimics the characteristics of our sample. Further simulation evidence on the behaviour of the adaptive estimators in both the triangular and the error correction models can be found in Hodgson (1995, 1998a).

In all simulations, we focus our attention on estimation of the slope parameter in the cointegrating regression, comparing the properties of the Johansen (1988) reduced rank estimator with those of the adaptive estimator described above. In all simulations, we generate a bivariate sample of 658 observations from a DGP with $k = 1$, a cointegrating vector that has zero intercept and unit slope, an error correction matrix set equal to our OLS estimates from the actual sample, and an iid bivariate innovation sequence with covariance matrix equal to the sample covariance of the OLS residuals from the actual sample. We use starting values for the variables equal to those in the actual sample. We consider the following classes of innovation densities: (1) Student's t with 3 and 5 degrees of freedom (Table V); (2) Gaussian (Table VI); and (3) variance contaminated mixture of normals, $\gamma N(0, \sigma_1^2\hat{\Sigma}) + (1-\gamma)N(0, \sigma_2^2\hat{\Sigma})$, where $\hat{\Sigma}$ is the sample covariance matrix, $\gamma = (0{\cdot}9, 0{\cdot}75, 0{\cdot}6)$, $\sigma_1^2 = 0{\cdot}1/\gamma$, and $\sigma_2^2 = 0{\cdot}9/(1-\gamma)$ (Table VII).

We conduct 5000 simulations with each set-up, estimating the model by reduced rank regression and by adaptive estimation, the latter using the former as the preliminary estimate and using bandwidths equal to the Silverman (1986) rule of thumb (ROT) and ROT + 0·0002, the choice of the latter bandwidth being motivated by the discussion in the previous section. In all

Copyright © 1999 John Wiley & Sons, Ltd.

Table VII. Monte Carlo results — ECM. Mixed normal errors ($k = 1$), estimation of slope

Estimator	Bias	MSE	Interdecile	Interquartile
Model with intercept ($\gamma = 0{\cdot}9$)				
Johansen	0·0107	0·00496	0·899–1·199	0·945–1·074
Adaptive (ROT)	0·0127	0·00434	0·909–1·191	0·953–1·070
Adaptive (ROT + 0·0002)	0·0137	0·00452	0·908–1·192	0·953–1·073
Model without intercept ($\gamma = 0{\cdot}75$)				
Johansen	0·0107	0·00511	0·896–1·199	0·945–1·074
Adaptive (ROT)	0·0112	0·00454	0·905–1·199	0·952–1·069
Adaptive (ROT + 0·0002)	0·0114	0·00462	0·904–1·200	0·951–1·068
Model without intercept ($\gamma = 0{\cdot}9$)				
Johansen	0·00870	0·00497	0·893–1·198	0·943–1·071
Adaptive (ROT)	0·00990	0·00461	0·899–1·196	0·948–1·068
Adaptive (ROT + 0·0002)	0·01020	0·00462	0·900–1·196	0·948–1·070

Note: See Table V.

cases, k is set equal to one. For each estimator, we throw out both the greatest and the least 10% of the estimates, and compute the bias and MSE figures reported in Tables V–VII based on the remaining truncated set of estimates. The reason for this truncation procedure is to eliminate the extremely inaccurate estimates that will arise with non-negligible probability when using the reduced rank regression procedure (Phillips, 1994, shows that the finite sample distribution of this estimator lacks finite moments). We also report interdecile and interquartile ranges for the estimators.

The results for t_3 and t_5 errors are reported in Table V. For the case of three degrees of freedom, we estimate the model both with and without an intercept included in the model. For all other innovation densities, we only consider estimation of a model with no intercept included. Table V suggests that the efficiency gain obtained using the adaptive estimator is quite small for the parameterization of the model considered here when the errors are Student's t. The fact that the efficiency improvements reported in Hodgson (1995, 1998a) for models with t_3 errors but with parameter values different from ours are so much larger than those reported here, even for sample sizes as small as 100, seems odd considering the fact mentioned above that the asymptotic efficiency gain of the adaptive estimator over the Gaussian pseudo-MLE depends only on the error distribution. Using results in Mitchell (1989) and Hodgson (1998b), we can show that the ratio of asymptotic variances of the adaptive and Gaussian pseudo-ML slope estimates is 0·44 for bivariate t_3 errors, and is 0·75 for bivariate t_5 errors. We can see from Table V that for the particular parameter configuration and sample size we have used, and when the degrees of freedom parameter is three, the adaptive estimator does improve upon the reduced rank regression estimator, although by a more modest degree than suggested by the asymptotic theory, the truncated MSE falling from 0·0275 to 0·0264 when an intercept is included and from 0·00472 to 0·00455 under the maintained hypothesis of zero intercept (the efficiency ratio is 0·96 in both cases). There is a corresponding tightening of the interquartile and interdecile ranges, but this is also quite modest. Changing the bandwidth produces almost no change in the outcome. Whether there would be a larger gain in using the bootstrap bandwidth selection procedure of Hsieh and Manski (1987) is of interest but is currently too computationally time consuming, as noted above.

Copyright © 1999 John Wiley & Sons, Ltd.

Table VI reports results of an analysis similar to that reported in Table V, but now with bivariate Gaussian innovations. The two estimators are asymptotically equivalent in this case, and this fact is borne out by the simulation results. Table VI suggests that there is no finite-sample efficiency loss in using the adaptive estimator for this DGP, even if the data actually are Gaussian. We note that simulation results in Hodgson (1995, 1998a) for t_3 and Gaussian errors, but with smaller samples and different DGPs than considered here, typically find larger efficiency losses for the adaptive estimator in the presence of Gaussian data, but also find the aforementioned larger efficiency gains for non-Gaussian data than we have found in Table V.

We can see from Table VII that the improvements from using the adaptive estimator are somewhat greater when the errors are mixed normal than when they are t, but are still quite modest, and are again smaller than the improvements reported in Hodgson (1995, 1998a). The Monte Carlo efficiency ratios are 0·88, 0·89, and 0·93 for the parameter values considered here (note that as γ decreases, so does the kurtosis of the distribution in our setup). The asymptotic efficiency ratios for these three distributions are, respectively, 0·13, 0·20, and 0·34.[5]

To summarize these simulation results, it seems that for our DGP, the efficiency gains to be obtained through implementation of the adaptive estimator are modest, especially compared to asymptotic gains predicted by theory and to the gains obtained in different DGPs with the same error distributions. On the other hand, we find that there is essentially zero efficiency loss when the errors are Gaussian. The gains we do obtain vary according to the actual innovation density.

5. CONCLUSIONS

We have applied adaptive estimators of cointegrated models to the estimation of a forward exchange market unbiasedness regression in levels. We have compared the estimates obtained using Gaussian pseudo-ML techniques with those obtained using adaptive estimation, for various specifications of the model. For a triangular representation with ARMA errors, the unbiasedness hypothesis is fairly well supported by both FM-OLS and the adaptive estimator. That robust and efficient estimators are not much different from FM-OLS for this representation makes sense given our finding that the innovations to the cointegrating regression's ARMA error process, though somewhat thick-tailed, are not drastically so relative to a Gaussian. We have also considered estimation of this model by full information maximum likelihood within an error correction representation. We find stronger evidence of non-Gaussianity in the bivariate innovations to the error correction model than we did in the univariate ARMA innovation process referred to above. The support for the unbiasedness hypothesis remains fairly strong.

We also report the results of Monte Carlo simulations comparing the Johansen (1988) and Hodgson (1998b) error correction model estimators for a DGP similar to that of our sample, for a variety of innovation densities. We find that there are efficiency gains (in terms of a truncated MSE criterion) for the adaptive estimator, but that they are smaller than would be suggested by asymptotic theory and by other simulation results reported in Hodgson (1995, 1998a). On the other hand, we find negligible finite sample efficiency loss when the data actually are Gaussian.

In implementing the estimator, we have employed a Gaussian kernel with a bandwidth parameter selected according to the Silverman (1986) rule of thumb method for density estimation. In the empirical implementation, we also selected the bandwidth according to the bootstrap method advocated by Hsieh and Manski (1987) but found very little effect on our

[5] These figures were computed numerically using Monte Carlo integration.

Copyright © 1999 John Wiley & Sons, Ltd.

empirical findings. The computational cost of this bootstrap approach renders infeasible an investigation of its properties through Monte Carlo simulations. In our simulations, we used a bandwidth that differed from the rule of thumb value by the same amount as the bootstrap bandwidth did in our empirical investigation, but found very little effect on our results of such variation.

The fact that the finite sample efficiency gains in our Monte Carlo investigation fall well short of the asymptotic gains indicated by theory suggests that there may be value in further work on implementation of the adaptive estimators. One possible line of inquiry would be to continue to use a bivariate kernel method, with more systematic attention given to the choice of kernel and bandwidth. The use of a thicker tailed kernel than the Gaussian (such as the logistic kernel suggested by Schick (1987) in a univariate application) obviates the need for trimming, although simulation results reported by Hsieh and Manski (1987) and Hodgson (1995) indicate the presence of trimming is probably not of primary concern in the implementation of the estimator. The results reported in the present paper, with limited application of the bootstrap bandwidth selection procedure of Hsieh and Manski (1987), indicate little change in using this approach rather than the straightforward Silverman (1986) rule of thumb approach, although further Monte Carlo comparison of these approaches would be of interest.

Nevertheless, we conjecture that there are probably more fruitful directions for improving the performance of adaptive estimators than in the areas of kernel and bandwidth selection. For example, we refer to work in progress by Hodgson, Choo and Linton (1998, unpublished manuscript), and Yang (1997, unpublished manuscript). The former paper considers the implementation of adaptive estimators in multivariate models when the innovations are iid from an elliptically symmetric density. This approach requires the potentially restrictive assumption of elliptical symmetry, but has the corresponding advantage of reducing the bivariate nonparametric density estimation problem in the applications reported in the present paper to a univariate density estimation problem (see also Stute and Werner, 1991), which is desirable because the finite sample behaviour of a univariate non-parametric kernel density estimator is better than that of the corresponding bivariate estimator. Yang (1997, unpublished manuscript) is investigating the application of numerical methods to the maximization of a non-parametric likelihood function, an approach which has the advantage of not relying on the use of the potentially poorly behaved OPG information matrix estimator typically used in the one-step iterative approach to adaptive estimation.

ACKNOWLEDGEMENTS

I am grateful to Peter Phillips and Oliver Linton for advice and assistance, to the referees and the Editor for their comments and suggestions, to the Alfred P. Sloan Foundation and the Social Sciences and Humanities Research Council of Canada for financial support, to the NSF for assistance under CAREER grant SBR-9701959, and to the Fiscal Policy and Economic Analysis Branch of the Department of Finance, Ottawa, Canada, for providing me with office and computer facilities while part of this research was carried out. This paper is drawn from Chapters 3 and 4 of my PhD thesis at Yale University and is a substantially revised version of Hodgson (1995). GAUSS programs implementing the adaptive estimators considered in this paper are available at http://troi.cc.rochester.edu/ ~ dshn

Copyright © 1999 John Wiley & Sons, Ltd.

REFERENCES

Ahn, S. K. and G. C. Reinsel (1990), 'Estimation for partially nonstationary multivariate autoregressive models', *Journal of the American Statistical Association*, **85**, 813–823.

Baillie, R. T. and T. Bollerslev (1989), 'Common stochastic trends in a system of exchange rates', *Journal of Finance*, **44**, 167–181.

Baillie, R. T., R. E. Lippens, and P. C. McMahon (1983), 'Testing rational expectations and efficiency in the foreign exchange market', *Econometrica*, **51**, 553–563.

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

Barhnart, S. W. and A. C. Szakmary (1991), 'Testing the unbiased forward rate hypothesis: evidence on unit roots, co-integration, and stochastic coefficients', *Journal of Financial and Quantitative Analysis*, **26**, 245–267.

Bickel, P. J. (1982), 'On adaptive estimation', *Annals of Statistics*, **10**, 647–671.

Box, G. E. P. and D. A. Pierce (1970), 'Distribution of residual autocorrelations in autoregressive-integrated moving average time series models', *Journal of the American Statistical Association*, **65**, 1509–1526.

Corbae, D., K.-G. Lim, and S. Ouliaris (1992), 'On cointegration and tests of forward market unbiasedness', *Review of Economics and Statistics*, **74**, 728–732.

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.

Engle, R. F. (1982), 'Autoregressive conditional heteroskecasticity with estimates of the variance of United Kingdom inflation', *Econometrica*, **50**, 987–1007.

Fama, E. F. (1984), 'Forward and spot exchange rates', *Journal of Monetary Economics*, **14**, 319–338.

Froot, F. A. and J. A. Frankel (1989), 'Forward discount bias: is it an exchange rate premium?' *Quarterly Journal of Economics*, **104**, 139–161.

Hakkio, C. S. and M. Rush (1989), 'Market efficiency and cointegration: an application to the sterling and deutschemark exchange markets', *Journal of International Money and Finance*, **8**, 75–88.

Hamilton, J. D. (1994), *Time Series Analysis*, Princeton University Press, Princeton.

Hansen, L. P. and R. J. Hodrick (1980), 'Forward exchange rates as optimal predictors of future spot rates: an econometric analysis', *Journal of Political Economy*, **88**, 829–853.

Hodgson, D. J. (1995), Adaptive estimation of cointegrated models: Simulation evidence and an application to the forward exchange market. RCER Working Paper #409, University of Rochester.

Hodgson, D. J. (1998a), 'Adaptive estimation of cointegrating regressions with ARMA errors', *Journal of Econometrics*, **85**, 231–267.

Hodgson, D. J. (1998b), 'Adaptive estimation of error correction models', *Econometric Theory*, **14**, 44–69.

Hsieh, D. A. and C. F. Manski (1987), 'Monte Carlo evidence on adaptive maximum likelihood estimation of a regression', *Annals of Statistics*, **15**, 541–551.

Jarque, C. M. and A. K. Bera (1980), 'Efficient tests for normality, homoskedasticity and serial independence of regression residuals', *Economics Letters*, **6**, 255–259.

Jeganathan, P. (1995), 'Some aspects of time series theory with applications to time series models', *Econometric Theory*, **11**, 818–887.

Jeganathan, P. (1997), 'On asymptotic inference in linear cointegrated time series systems', *Econometric Theory*, **13**, 692–745.

Johansen, S. (1988), 'Stochastic analysis of cointegration vectors', *Journal of Economic Dynamics and Control*, **12**, 231–254.

Kreiss, J.-P. (1987a), 'On adaptive estimation in autoregressive models when there are nuisance functions', *Statistics & Decisions*, **5**, 59–76.

Kreiss, J.-P. (1987b), 'On adaptive estimation in stationary ARMA processes', *Annals of Statistics*, **15**, 112–133.

Linton, O. (1993), 'Adaptive estimation in ARCH models', *Econometric Theory*, **9**, 539–569.

McCallum, B. T. (1994), 'A reconsideration of the uncovered interest parity relationship', *Journal of Monetary Economics*, **33**, 105–132.

Mitchell, A. F. S. (1989), 'The information matrix, skewness tensor and α-connections for the general multivariate elliptic distribution', *Annals of the Institute of Mathematical Statistics* (Tokyo), **41**, 289–304.

Copyright © 1999 John Wiley & Sons, Ltd.

Ouliaris, S. and P. C. B. Phillips (1994), *COINT 2.0—GAUSS Procedures for Cointegrated Regressions*, Predicta Software Inc., Madison, CT.
Park, J. Y. (1992), 'Canonical cointegrating regressions', *Econometrica*, **60**, 119–143.
Phillips, P. C. B. (1991), 'Optimal inference in cointegrated systems', *Econometrica*, **59**, 283–306.
Phillips, P. C. B. (1994), 'Some exact distribution theory for maximum likelihood estimators of cointegrating coefficients in error correction models', *Econometrica*, **62**, 73–93.
Phillips, P. C. B. (1995), 'Robust nonstationary regression', *Econometric Theory*, **11**, 912–951.
Phillips, P. C. B. and B. E. Hansen (1990), 'Statistical inference in instrumental variables regression with I(1) processes', *Review of Economic Studies*, **57**, 99–125.
Phillips, P. C. B. and M. Loretan (1991), 'Estimating long-run economic equilibria', *Review of Economic Studies*, **58**, 407–436.
Phillips, P. C. B., J. W. McFarland, and P. C. McMahon (1996), 'Robust tests of forward exchange market efficiency with empirical evidence from the 1920's', *Journal of Applied Econometrics*, **11**, 1–22.
Saikkonen, P. (1991), 'Asymptotically efficient estimation of cointegration regressions', *Econometric Theory*, **7**, 1–21.
Schick, A. (1987), 'A note on the construction of asymptotically linear estimators', *Journal of Statistical Planning and Inference*, **16**, 89–105.
Silverman, B. (1986), *Density Estimation for Statistics and Data Analysis*, Chapman and Hall, London.
Steigerwald, D. (1992a), 'Adaptive estimation in time series regression models', *Journal of Econometrics*, **54**, 251–276.
Steigerwald, D. (1992b), 'On the finite sample behaviour of adaptive estimators', *Journal of Econometrics*, **54**, 371–400.
Stein, C. (1956), 'Efficient nonparametric estimation and testing', *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, 1, University of California Press, Berkeley, 187–196.
Stock, J. H. and M. W. Watson (1993), 'A simple estimator of cointegrating vectors in higher order integrated systems', *Econometrica*, **61**, 783–820.
Stone, C. (1975), 'Adaptive maximum likelihood estimation of a location parameter', *Annals of Statistics*, **3**, 267–284.
Stute, W. and U. Werner (1991), 'Nonparametric estimation of elliptically contoured densities', in G. Roussas (ed.), *Nonparametric Functional Estimation and Related Topics*, Kluwer, Amsterdam.
Watson, M. W. (1994), 'Vector autoregressions and cointegration', in R. Engle and D. McFadden (eds), *Handbook of Econometrics*, Vol. 4, North-Holland, Amsterdam.
White, H. (1982), 'Maximum likelihood estimation of misspecified models', *Econometrica*, **50**, 1–25.

Copyright © 1999 John Wiley & Sons, Ltd.

http://www.jstor.org

LINKED CITATIONS

You have printed the following article:

Adaptive Estimation of Cointegrated Models: Simulation Evidence and an Application to the Forward Exchange Market

Douglas J. Hodgson

Journal of Applied Econometrics, Vol. 14, No. 6. (Nov. - Dec., 1999), pp. 627-650.

Stable URL:

http://links.jstor.org/sici?sici=0883-7252%28199911%2F12%2914%3A6%3C627%3AAEOCMS%3E2.0.CO%3B2-Y

This article references the following linked citations. If you are trying to access articles from an off-campus location, you may be required to first logon via your library web site to access JSTOR. Please visit your library's website or contact a librarian to learn about options for remote access to JSTOR.

[Footnotes]

[1] **Optimal Inference in Cointegrated Systems**

P. C. B. Phillips

Econometrica, Vol. 59, No. 2. (Mar., 1991), pp. 283-306.

Stable URL:

http://links.jstor.org/sici?sici=0012-9682%28199103%2959%3A2%3C283%3AOIICS%3E2.0.CO%3B2-5

References

Estimation for Partially Nonstationary Multivariate Autoregressive Models

Sung K. Ahn; Gregory C. Reinsel

Journal of the American Statistical Association, Vol. 85, No. 411. (Sep., 1990), pp. 813-823.

Stable URL:

http://links.jstor.org/sici?sici=0162-1459%28199009%2985%3A411%3C813%3AEFPNMA%3E2.0.CO%3B2-X

Common Stochastic Trends in a System of Exchange Rates

Richard T. Baillie; Tim Bollerslev

The Journal of Finance, Vol. 44, No. 1. (Mar., 1989), pp. 167-181.

Stable URL:

http://links.jstor.org/sici?sici=0022-1082%28198903%2944%3A1%3C167%3ACSTIAS%3E2.0.CO%3B2-B

NOTE: *The reference numbering from the original has been maintained in this citation list.*

http://www.jstor.org

LINKED CITATIONS

Testing Rational Expectations and Efficiency in the Foreign Exchange Market
Richard T. Baillie; Robert E. Lippens; Patrick C. McMahon
Econometrica, Vol. 51, No. 3. (May, 1983), pp. 553-563.
Stable URL:
http://links.jstor.org/sici?sici=0012-9682%28198305%2951%3A3%3C553%3ATREAEI%3E2.0.CO%3B2-9

On Adaptive Estimation
P. J. Bickel
The Annals of Statistics, Vol. 10, No. 3. (Sep., 1982), pp. 647-671.
Stable URL:
http://links.jstor.org/sici?sici=0090-5364%28198209%2910%3A3%3C647%3AOAE%3E2.0.CO%3B2-1

Distribution of Residual Autocorrelations in Autoregressive-Integrated Moving Average Time Series Models
G. E. P. Box; David A. Pierce
Journal of the American Statistical Association, Vol. 65, No. 332. (Dec., 1970), pp. 1509-1526.
Stable URL:
http://links.jstor.org/sici?sici=0162-1459%28197012%2965%3A332%3C1509%3ADORAIA%3E2.0.CO%3B2-A

On Cointegration and Tests of Forward Market Unbiasedness
Dean Corbae; Kian-Guan Lim; Sam Ouliaris
The Review of Economics and Statistics, Vol. 74, No. 4. (Nov., 1992), pp. 728-732.
Stable URL:
http://links.jstor.org/sici?sici=0034-6535%28199211%2974%3A4%3C728%3AOCATOF%3E2.0.CO%3B2-O

Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation
Robert F. Engle
Econometrica, Vol. 50, No. 4. (Jul., 1982), pp. 987-1007.
Stable URL:
http://links.jstor.org/sici?sici=0012-9682%28198207%2950%3A4%3C987%3AACHWEO%3E2.0.CO%3B2-3

Forward Discount Bias: Is it an Exchange Risk Premium?
Kenneth A. Froot; Jeffrey A. Frankel
The Quarterly Journal of Economics, Vol. 104, No. 1. (Feb., 1989), pp. 139-161.
Stable URL:
http://links.jstor.org/sici?sici=0033-5533%28198902%29104%3A1%3C139%3AFDBIIA%3E2.0.CO%3B2-7

NOTE: *The reference numbering from the original has been maintained in this citation list.*

http://www.jstor.org

LINKED CITATIONS

Forward Exchange Rates as Optimal Predictors of Future Spot Rates: An Econometric Analysis
Lars Peter Hansen; Robert J. Hodrick
The Journal of Political Economy, Vol. 88, No. 5. (Oct., 1980), pp. 829-853.
Stable URL:
http://links.jstor.org/sici?sici=0022-3808%28198010%2988%3A5%3C829%3AFERAOP%3E2.0.CO%3B2-J

Monte Carlo Evidence on Adaptive Maximum Likelihood Estimation of a Regression
David A. Hsieh; Charles F. Manski
The Annals of Statistics, Vol. 15, No. 2. (Jun., 1987), pp. 541-551.
Stable URL:
http://links.jstor.org/sici?sici=0090-5364%28198706%2915%3A2%3C541%3AMCEOAM%3E2.0.CO%3B2-R

On Adaptive Estimation in Stationary ARMA Processes
Jens-Peter Kreiss
The Annals of Statistics, Vol. 15, No. 1. (Mar., 1987), pp. 112-133.
Stable URL:
http://links.jstor.org/sici?sici=0090-5364%28198703%2915%3A1%3C112%3AOAEISA%3E2.0.CO%3B2-F

Canonical Cointegrating Regressions
Joon Y. Park
Econometrica, Vol. 60, No. 1. (Jan., 1992), pp. 119-143.
Stable URL:
http://links.jstor.org/sici?sici=0012-9682%28199201%2960%3A1%3C119%3ACCR%3E2.0.CO%3B2-R

Optimal Inference in Cointegrated Systems
P. C. B. Phillips
Econometrica, Vol. 59, No. 2. (Mar., 1991), pp. 283-306.
Stable URL:
http://links.jstor.org/sici?sici=0012-9682%28199103%2959%3A2%3C283%3AOIICS%3E2.0.CO%3B2-5

Some Exact Distribution Theory for Maximum Likelihood Estimators of Cointegrating Coefficients in Error Correction Models
Peter C. B. Phillips
Econometrica, Vol. 62, No. 1. (Jan., 1994), pp. 73-93.
Stable URL:
http://links.jstor.org/sici?sici=0012-9682%28199401%2962%3A1%3C73%3ASEDTFM%3E2.0.CO%3B2-H

NOTE: *The reference numbering from the original has been maintained in this citation list.*

http://www.jstor.org

LINKED CITATIONS

Statistical Inference in Instrumental Variables Regression with I(1) Processes
Peter C. B. Phillips; Bruce E. Hansen
The Review of Economic Studies, Vol. 57, No. 1. (Jan., 1990), pp. 99-125.
Stable URL:
http://links.jstor.org/sici?sici=0034-6527%28199001%2957%3A1%3C99%3ASIIIVR%3E2.0.CO%3B2-6

Estimating Long-Run Economic Equilibria
Peter C. B. Phillips; Mico Loretan
The Review of Economic Studies, Vol. 58, No. 3, Special Issue: The Econometrics of Financial Markets. (May, 1991), pp. 407-436.
Stable URL:
http://links.jstor.org/sici?sici=0034-6527%28199105%2958%3A3%3C407%3AELEE%3E2.0.CO%3B2-C

Robust Tests of Forward Exchange Market Efficiency with Empirical Evidence From the 1920s
Peter C. B. Phillips; James W. McFarland; Patrick C. McMahon
Journal of Applied Econometrics, Vol. 11, No. 1. (Jan. - Feb., 1996), pp. 1-22.
Stable URL:
http://links.jstor.org/sici?sici=0883-7252%28199601%2F02%2911%3A1%3C1%3ARTOFEM%3E2.0.CO%3B2-M

A Simple Estimator of Cointegrating Vectors in Higher Order Integrated Systems
James H. Stock; Mark W. Watson
Econometrica, Vol. 61, No. 4. (Jul., 1993), pp. 783-820.
Stable URL:
http://links.jstor.org/sici?sici=0012-9682%28199307%2961%3A4%3C783%3AASEOCV%3E2.0.CO%3B2-Z

Adaptive Maximum Likelihood Estimators of a Location Parameter
Charles J. Stone
The Annals of Statistics, Vol. 3, No. 2. (Mar., 1975), pp. 267-284.
Stable URL:
http://links.jstor.org/sici?sici=0090-5364%28197503%293%3A2%3C267%3AAMLEOA%3E2.0.CO%3B2-2

Maximum Likelihood Estimation of Misspecified Models
Halbert White
Econometrica, Vol. 50, No. 1. (Jan., 1982), pp. 1-25.
Stable URL:
http://links.jstor.org/sici?sici=0012-9682%28198201%2950%3A1%3C1%3AMLEOMM%3E2.0.CO%3B2-C

NOTE: *The reference numbering from the original has been maintained in this citation list.*