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

Hodgson, Douglas J.

Working Paper No. 409
September 1995

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

Douglas J. Hodgson

Rochester Center for Economic Research
Working Paper No. 409

September 1995

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

Douglas J. Hodgson

Department of Economics

University of Rochester

Harkness Hall

Rochester, NY

14627

* I am grateful to Peter Phillips and Oliver Linton for advice and assistance, to the Alfred P. Sloan Foundation and the Social Sciences and Humanities Research Council of Canada for financial support, 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 the Introduction and Chapters 3 and 4 of my Ph.D. thesis at Yale University.

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, and so provide an attractive alternative to commonly employed Gaussian pseudo-MLE's when the data are believed to be generated by some unspecified non-Gaussian distribution. We consider two representations of cointegrated systems - linear cointegrating regressions and error correction models. The motivation for and advantages of adaptive estimators in such systems are discussed and their construction is described. A Monte Carlo study compares the finite sample performance of the adaptive estimators with the Gaussian pseudo-MLE's for various thick-tailed densities for the innovations, and obtains very encouraging results. An empirical application to estimating a forward exchange market unbiasedness model in an error correction representation finds that the daily data are thick-tailed and that the adaptive estimator provides a much stronger inference in favour of the unbiasedness hypothesis than does the Gaussian pseudo-MLE.

1. INTRODUCTION: ADAPTIVE ESTIMATION OF COINTEGRATED MODELS

This paper reports the first attempts to gauge the finite-sample performance, through Monte Carlo simulations and an empirical application, of the new technique of adaptive estimation of cointegrated systems. We obtain highly encouraging simulation results and a very striking empirical result, but before presenting and discussing these results, we provide an expository discussion on the technique of adaptive estimation, its history, its motivation, and its application to cointegrated models.

The data used to estimate econometric models often possess characteristics that do not conform to the assumptions under which standard estimation techniques have desirable properties like consistency, asymptotic normality, or efficiency. In such cases, empirical inference can be improved through the use of methods that do perform well given certain nonstandard characteristics of the data. Two basic examples are feasible generalized least squares and instrumental variables. The former is designed to deal with autocorrelated or heteroskedastic residuals in a linear regression model, a situation in which the ordinary least squares (OLS) estimator is inefficient, while the latter deals with endogenous regressors, in which case OLS is inconsistent.

Hodgson (1995a) falls within this general sphere of research concerned with the improved estimation of models when the data are nonstandard. Specifically, it deals with efficient estimation in the presence of data that are both nonstationary and non-Gaussian (i.e. non-normal). We consider data whose nonstationarity is due to the presence of unit roots, and investigate the asymptotically efficient estimation of cointegrated models whose underlying innovations are drawn from unknown and possibly non-Gaussian distributions. Two popular representations of cointegrated systems are considered: (i) linear cointegrating regressions with possibly endogenous regressors and time dependent errors, and (ii) reduced rank vector error correction models.

The presence of a unit root (i.e. an integrated, or difference stationary) component is often an implication of economic theory and is a more or less debatably well-established empirical fact for many economic and financial time series. The work of Samuelson (1965) and Mandelbrot (1966) provides theoretical justification for the existence of unit roots in speculative price series, and there is a large empirical literature devoted to testing the martingale hypothesis in such series (see, for example, Cootner (1964) and Fama (1970)). The seminal study by Nelson and Plosser (1982) found that the unit root hypothesis could not be rejected in favour of the trend stationary alternative for several key U.S. macroeconomic time series. That real aggregates possess stochastic trends is an implication of neoclassical growth theory, as noted by King, Plosser, Stock, and Watson (1991).

There is also considerable empirical evidence of non-Gaussianity in many integrated economic and financial time series. That speculative price returns are non-Gaussian is by now a widely accepted stylized fact. Such series are prone to occasional large shocks and bouts of high volatility, which can be associated with excess kurtosis or thick tails in the density of the return process. Seminal works in this area are Mandelbrot (1963) and Fama (1963,1965). Modeling conditional heteroskedasticity in the returns series can only partially account for the high kurtosis - researchers commonly find the innovations in ARCH models to still be non-Gaussian, as noted by Bera and Higgins

(1993) and Bollerslev, Chou, and Kroner (1992). The presence of infrequent large shocks, and a corresponding thick-tailed innovation density, has also been found in several macroeconomic series by Blanchard and Watson (1986) and Balke and Fomby (1994).

In econometric models containing two or more integrated time series, we frequently need to estimate long-run, or cointegrating, relationships. A collection of integrated, I(1), time series are cointegrated if they share stochastic trend components, in which case there is common long-run co-movement among the series. Formally, they are cointegrated if there exist among them one or more linear relationships that are stationary, or I(0), and so have spectral density functions that are positive and finite at the origin. Such linear relationships are known as cointegrating vectors and it is their estimation that is the topic of Hodgson (1995a).

The cointegrating vectors characterize the long-run dynamics present in a system of I(1) time series. However, most systems generally also have short-run, transitory dynamics. These can be modeled in a variety of ways, giving rise to different representations of cointegrated systems. One approach is to employ a linear cointegrating regression, with the regression errors and regressor first differences following a jointly stationary process that that may or may not be explicitly parameterized. Phillips and Durlauf (1986) and Phillips and Hansen (1990) analyze cointegrating regressions with nonparametric transitory components. Hodgson (1995b) assumes a parametric model in which the regression errors follow an ARMA process and the regressors follow a random walk.

A second approach is to employ an error correction representation, in which the first differences of the variables are expressed in a vector autoregressive representation with a lagged levels term included among the regressors. The coefficient matrix on the latter term has reduced rank equal to the number of linearly independent cointegrating relationships in the system, and can be factored into two matrices, one consisting of the system's cointegrating vectors and the other consisting of error correction coefficients, which reflect the adjustment of the variables to transitory deviations from their cointegrating relationships. Any cointegrated system has an error correction representation (see, for example, Engle and Granger (1987)), with the short-run dynamics being fully parameterized through the error correction coefficients and the coefficient matrices on the lagged differences. Hodgson (1995c) considers the estimation of error correction models.

Engle and Granger (1987) recommend estimating cointegrating vectors by OLS, which possesses the property of superconsistency, i.e. consistency at the rate n, where n denotes the sample size. The OLS estimator is the subject of analyses by Phillips and Durlauf (1986) and Stock (1987). Although being superconsistent, OLS is asymptotically mixed normal and asymptotically efficient only under the restrictive assumptions of exogenous regressors and iid Gaussian errors. The exogeneity and independence assumptions are relaxed by Phillips and Hansen (1990), whose fully modified OLS estimator is asymptotically mixed normal under very general nonparametric specifications of the endogeneity and dependence. Similar properties hold for the estimators of Park (1992), Phillips and Loretan (1991), Saikkonen (1991), and Stock and

Watson (1993). These studies all obtain estimates in a linear regression representation that are asymptotically mixed normal and that, furthermore, are asymptotically efficient (i.e. have an asymptotic covariance matrix equal to the inverse of the asymptotic information matrix) under Gaussianity assumptions on the innovations to the Wold moving average representation of the regression errors and regressor first differences.

Little work has been carried out to date on efficiently estimating cointegrated models when the innovations are not Gaussian. Phillips (1993b) has derived robust fully modified LAD and M-estimators that account for endogeneity and dependence in the same manner as do Phillips and Hansen (1990) in the fully modified OLS case. Jeganathan (1994) has derived adaptive estimators for the case of iid errors whose joint density function with the innovations to the random walk regressors is elliptically symmetric. Hodgson (1995b) extends this analysis by using Kreiss' (1987) results on the adaptive estimation of ARMA processes to allow the errors to the cointegrating regression to have a stationary and invertible ARMA representation, although Jeganathan's (1994) assumptions of elliptical symmetry and random walk regressors are maintained. A further extension to allow for more general dependence and endogeneity along the lines of Phillips and Hansen (1990) would be desirable.

Estimation of the cointegrating parameters in error correction models is considered by Johansen (1988) and Ahn and Reinsel (1990). Both analyses derive superconsistent, asymptotically mixed normal, and asymptotically efficient Gaussian maximum likelihood estimators. If the data are non-Gaussian, then the first two properties hold but efficiency fails for these authors' estimators. Hodgson (1995c) considers the extension to non-Gaussian (but symmetric) innovation densities and derives asymptotically efficient adaptive estimates.

As outlined above, existing methods allow us to efficiently estimate cointegrated models containing transitory dynamics only if we are willing to assume that the innovations are drawn from a Gaussian distribution. Hodgson (1995a) is concerned with the efficient estimation of such models when the innovation density is unknown to the investigator and may be non-Gaussian. If the density is assumed to belong to some known parametric family, Gaussian or otherwise, then asymptotic efficiency can be achieved by computing the maximum likelihood estimator, as in Johansen (1988) for Gaussian error correction models, or by employing an asymptotically equivalent iterative estimator, as in Ahn and Reinsel (1990), also for Gaussian error correction models. However, if the parametric family to which the innovation density belongs is unknown, then the shape of the likelihood function is unknown, rendering infeasible the computation of the maximum likelihood estimator or asymptotically equivalent iterative estimator. Nevertheless, the question arises as to whether efficient estimation is somehow possible - in other words, is it possible to "adapt" for the fact that the density is unknown by using an estimate of it.

The concept of adaptive estimation, as employed here, was first formulated by Stein (1956), who asked if there were models whose parameters of interest could be as efficiently estimated not knowing some infinite dimensional nuisance parameter as they could be if the nuisance parameter *were* known. In our framework, the nuisance parameter is the density function of a model's iid innovation vector. Can we estimate the model as well, asymptotically, not knowing this density function as we could do if it were known? Over the years, this question has been answered affirmatively for several important models. The location parameter problem, in which a sequence of iid observations are used to estimate the location of the density from which they are drawn, was considered by Beran (1974) and Stone (1975). They showed that in the case of a symmetric univariate density of unknown shape, adaptive estimation is possible, and derived expressions for adaptive estimators. Linear regressions (among other models) were analyzed by Bickel (1982), who showed how to adaptively estimate the intercept and slope parameters assuming a symmetric error density and the slope parameters without assuming symmetry. Manski (1984) studied the adaptive estimation of nonlinear econometric models, and Kreiss (1987) brought the concept of adaptive estimation into the time series literature with his analysis of ARMA models. Steigerwald (1992a) adaptively estimates linear (non-cointegrating) regressions with ARMA errors, and Linton (1993) investigates the adaptive estimation of ARCH models, an important step in light of the aforementioned empirical findings of conditional non-Gaussianity in ARCH models. Finally, there is Jeganathan's (1994) work, referred to above, on the adaptive estimation of cointegrating regressions with iid errors.

Our understanding of how to efficiently estimate a model whose likelihood function is unknown is aided by first considering the case where the likelihood is known. In this situation, an alternative approach to calculating the maximum likelihood estimator is to begin with a preliminary estimator that is consistent (at some prespecified rate) and discretized (i.e. that belongs to one of the finite number of points resulting from drawing a fine grid, whose fineness increases with sample size, over the parameter space). We then adjust the preliminary estimator by a quantity consisting of the sample score function for the model premultiplied by the inverse of the sample information matrix for the model, both evaluated at the preliminary estimator. The resulting iterative estimator has an asymptotic mixed normal distribution with covariance matrix equal to the inverse of the asymptotic information matrix, and so is efficient. This construction requires knowledge of the score function and information matrix of the innovation density, and so of the density itself. This is where the problem arises in using such techniques to efficiently estimate a model whose innovation density is unknown.

The preceding iterative estimation strategy provides a clue as to how adaptive estimation can be achieved. Although we do not know the innovation density, and specifically its score function and information matrix, we may be able to efficiently estimate the model if we can consistently estimate these quantities. A natural possibility is to use the residuals of the model as evaluated from a consistent preliminary estimator to form nonparametric kernel estimates of the unknown density and of its score and information. These estimates can then be substituted into the formula for the iterative estimator in place of the unknown score function and information matrix.

Since Stone (1975), the adaptive estimation literature has proceeded in the foregoing manner by employing trimmed Gaussian kernel estimates of the innovation score and information. The investigator selects a value for the bandwidth parameter and computes kernel estimates of the density and its vector of first partial derivatives. These estimates are then evaluated at each of the estimated residuals, with the ratio between the derivative estimate and the density estimate constituting the score estimate at each point, excepting certain points which are trimmed by setting the score estimate equal to zero. Trimming typically occurs at isolated tail observations, where the kernel density estimate is so close to zero that the score estimate is exploding. The more heavily smoothed is the density estimator, the less need there is for trimming. In the computation of adaptive estimators, the amount of trimming, as with smoothing, is left to the discretion of the investigator.

Section 2 briefly summarizes the details of the construction outlined above for the two respective representations of cointegrated models. We show how to use a trimmed Gaussian kernel estimator to compute adaptive estimates for each representation. Section 3 evaluates the finite sample performance of the adaptive estimators through a series of Monte Carlo simulations similar in structure to those performed by Hsieh and Manski (1987) for linear non-cointegrating regressions. We consider bivariate models with normal, *t*, variance contaminated mixture of normal, and bimodal mixture of normal innovation densities and report interdecile and interquartile ranges, bias, and truncated and untruncated mean squared errors for adaptive and Gaussian estimators and, for linear regression models, least absolute deviations. Sample sizes range from 100 to 500 and various settings of the smoothing and trimming parameters are considered. For non-Gaussian densities the adaptive estimators substantially improve upon the Gaussian estimators, a result insensitive to smoothing and trimming parameter selection, with the efficiency loss when the Gaussian assumption is correct being moderate to small. Section 4 contains an empirical implementation of adaptive estimation to the problem of estimating the cointegrating parameter in a forward exchange market unbiasedness model. In a model consisting of daily observations on the logarithms of a spot exchange rate and the associated lagged forward rate, several researchers have evaluated the evidence for the hypothesis that the slope parameter in the regression of the spot rate on the lagged forward rate is unity, i.e. that the forward rate is an unbiased predictor of the spot rate. It is not uncommon for investigators using a Gaussian pseudo-MLE to get estimates some distance from one. Since daily exchange rate data can be highly non-Gaussian, one would imagine that sharper inferences could be obtained through the use of more efficient estimators, which may also serve to reduce finite sample bias induced in the Gaussian estimator by outliers. We report estimates of a forward unbiasedness model for the Canada-U.S. exchange rate, specified as an error correction representation, using both Johansen's (1988) Gaussian pseudo-MLE and the adaptive estimator developed in Chapter 2. The Johansen estimate is 0.937, approximately two estimated asymptotic standard errors from unity, while the adaptive estimate ranges from 0.995 to 0.998, depending on the bandwidth setting, and clearly leads to a stronger inference in favour of the unbiasedness hypothesis.

2. THE MODELS AND ADAPTIVE ESTIMATORS

This section summarizes the methodology for computing adaptive estimators as developed in Hodgson (1995b and c) for cointegrating regressions with ARMA errors and error correction models, respectively. For each representation in turn, we present the model to be estimated and the details of its adaptive estimation.

(a) Linear Regression with ARMA Errors

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 system from this relationship follow a stationary and invertible ARMA(p,q) process, with p and q known. For every t=1,...,n, we have

(1) $$Y_t = BX_t + u_t$$

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

(3) $$X_t = X_{t-1} + v_t$$

where X_t and B are m-vectors and $(\varepsilon_t, v_t')'$ are iid from the unknown elliptically symmetric density $p(\varepsilon, v)$. We define

$$\psi(\varepsilon, v) = \left(\partial p(\varepsilon, v) / \partial\varepsilon\right) / p(\varepsilon, v)$$

and

$$\lambda^2 = \iint |\psi(\varepsilon, v)|^2 p(\varepsilon, v) dv d\varepsilon,$$

assuming $0 < \lambda^2 < \infty$, where $|\cdot|$ denotes the Euclidean norm. Since $p(\varepsilon, v)$ is unknown, so are $\psi(\varepsilon, v)$ and λ^2; in computing an adaptive estimator, we will require consistent nonparametric kernel estimates of these latter two quantities.

We denote the vector of ARMA coefficients by $\eta = (a_1, \ldots, a_p; b_1, \ldots, b_q)'$, the p+q+m-dimensional full parameter vector by $\theta = (\eta', B')'$, and a p+q+m-dimensional scaling matrix by $\delta_n = diag\left[n^{-1/2} I_{p+q}, n^{-1} I_m\right]$. We use the parameters of the MA process to define the infinite sequence of constants $\{\gamma_k(\theta)\}$ as follows:

$$\left(1 + b_1 z + \cdots + b_q z^q\right)^{-1} = \sum_{k=0}^{\infty} \gamma_k(\theta) z^k ,$$

such that

$$\gamma_s(\theta) + b_1 \gamma_{s-1}(\theta) + \cdots + b_q \gamma_{s-q}(\theta) = 0 \quad \forall s \geq 1,$$

with $\gamma_s(\theta) = 0 \;\; \forall s < 0$ and $\gamma_0(\theta) = 1$. We further define

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

$$\Gamma_{t-1}(\eta_n, B) = \sum_{j=0}^{t-1} \gamma_j(B, \eta_n)\left[X_{t-j} - \sum_{k=1}^{p} a_k^n X_{t-j-k} \right] - v_t,$$

where $\eta_n - \eta = O(n^{-1/2})$ and $a_k^n - a_k = O(n^{-1/2})$, and

$$H_{t-1}(\theta) = \left(Z_{t-1}(\eta, \eta, B)', \Gamma_{t-1}(\eta, B)'\right)'.$$

The $p+q+m$-vector $H_{t-1}(\theta)$ plays an important role in our derivation of an adaptive estimator because it enters the expressions for both the score function and the information matrix for the model, which we can write respectively as

$$(4) \qquad W_n(\theta) = -\sum_{t=1}^{n} \delta_n H_{t-1}(\theta) \psi(\varepsilon_t, v_t)$$

and

$$(5) \qquad S_n(\theta) = \lambda^2 \sum_{t=1}^{n} \delta_n H_{t-1}(\theta) H_{t-1}(\theta)' \, \delta_n .$$

If $\psi(\varepsilon, v)$ and λ^2 were known, we could use $W_n(\theta)$ and $S_n(\theta)$, evaluated at some consistent preliminary estimate, to form an asymptotically efficient iterative estimator. Instead, we must replace $\psi(\varepsilon, v)$ and λ^2 in (4) and (5) with kernel estimates.

To derive these nonparametric estimators, we follow Jeganathan (1994), using our assumption of elliptical symmetry to write:

$$p(\varepsilon, v) = |\det \Omega|^{-1/2} f^* \left(\left| \Omega^{-1/2} \begin{pmatrix} \varepsilon \\ v \end{pmatrix} \right| \right)$$

for some f^*, where $\Omega = \begin{pmatrix} \omega_{11} & \omega_{21}' \\ \omega_{21} & \Omega_{22} \end{pmatrix}$. We define $\ell_{11} = \left(\omega_{11} - \omega_{21}' \Omega_{22}^{-1} \omega_{21}\right)^{1/2}$, so that

$$\begin{aligned} p(\varepsilon, v) &= |\det \Omega|^{-1/2} f^* \left(\left| \begin{matrix} \ell_{11}^{-1}(\varepsilon - \omega_{21}' \Omega_{22}^{-1} v \\ \Omega_{22}^{-1/2} v \end{matrix} \right| \right) \\ &\equiv f(z, v), \end{aligned}$$

where $z = \varepsilon - \omega_{21}' \Omega_{22}^{-1} v$. Our elliptical symmetry assumption implies that the conditional density of ε given v is symmetric with mean $\omega_{21}' \Omega_{22}^{-1} v$, and that, given v, f is symmetric about zero in z.

In nonparametrically estimating $\psi(\varepsilon, v)$, we employ the following definitions:

$$\begin{aligned} &z_t(\theta) = \varepsilon_t(\theta) - \omega_{21}' \Omega_{22}^{-1} v_t, \\ &\pi(z, v, \sigma) = \left(1 / (\sigma\sqrt{2\pi})^{m+1}\right) \exp\left(-\left(|z|^2 + |v|^2\right) / 2\sigma^2\right), \\ &\hat{f}_{\sigma,t}(x, y, \theta) = (1/2(n-1)) \sum_{\substack{i=1 \\ i \neq t}}^{n} \left\{\pi\left(x + z_i(\theta), y + v_i, \sigma\right) + \pi\left(x - z_i(\theta), y + v_i, \sigma\right)\right\}. \end{aligned}$$

We additionally define $f_{\sigma,t}'(x, y, \theta)$ as the partial derivative of $\hat{f}_{\sigma,t}$ with respect to x. Note that $\hat{f}_{\sigma,t}$ is a Gaussian kernel density estimator, symmetric by construction, with bandwidth parameter σ. Given fixed y, we employ the above notation in our definition of a score function estimator, as follows:

$$\hat{\psi}_{n,t}(x,y,\theta)=\begin{cases}\dfrac{\hat{f}_{\sigma,t}'(x,y,\theta)}{\hat{f}_{\sigma,t}(x,y,\theta)} & if \begin{cases}\hat{f}_{\sigma,t}(x,y,\theta)\geq m_n\\ |(x,y')'|\leq \alpha_n\\ \left|\hat{f}_{\sigma,t}'(x,y,\theta)\right|\leq c_n\hat{f}_{\sigma,t}(x,y,\theta)\end{cases}\\ 0 & otherwise\end{cases}.$$

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

As indicated above, we may employ our kernel estimates to compute the following estimates of the score and information of the model:

(6) $$\hat{W}_n(\theta)=-\sum_{t=1}^{n}\delta_n H_{t-1}(\theta)\hat{\psi}_{n,t}(\varepsilon_t(\theta),v_t,\theta),$$

(7) $$\hat{S}_n(\theta)=\hat{\lambda}^2(\theta)\sum_{t=1}^{n}\delta_n H_{t-1}(\theta)H_{t-1}(\theta)'\delta_n,$$

where $\hat{\lambda}^2(\theta)=n^{-1}\sum_{t=1}^{n}\hat{\psi}_{n,t}(\varepsilon_t(\theta),v_t,\theta)^2$. An adaptive estimator is then obtained by taking a discretized[1], δ_n^{-1}-consistent preliminary estimator θ_n^{**}, and adjusting it by an iterative term involving the quantities in (6) and (7), evaluated at θ_n^{**}, as follows:

(8) $$\tilde{\theta}_n=\theta_n^{**}+\delta_n\hat{S}_n^{-1}(\theta_n^{**})\hat{W}_n(\theta_n^{**}).$$

The asymptotic distribution of $\tilde{\theta}_n$ is mixed normal with a covariance matrix equal to the inverse of the asymptotic information matrix for the model, so that $\tilde{\theta}_n$ is asymptotically efficient. Specifically, we have

$$\delta_n^{-1}(\tilde{\theta}_n-\theta)\Rightarrow MN\left(0,\lambda^{-2}J^{-1}(\theta)\right)$$

[1] For a definition of discretization, see, for example, Jeganathan (1994). It essentially involves rounding off an estimate to the nearest point on a fine grid whose fineness increases with the sample size.

where the information matrix is $\lambda^2 J(\theta)$, $J(\theta) = diag(J_{ZZ}(\theta), J_{\Gamma\Gamma}(\theta))$, $J_{ZZ}(\theta) = E\left[Z_t Z_t'\right]$,

and $J_{\Gamma\Gamma}(\theta) = \left[\sum_{j=0}^{\infty} \gamma_j(\theta)\right]^2 \left[1 - \sum_{k=1}^{p} a_k\right]^2 \int_0^1 M_2 M_2'$, where M_2 is a Brownian motion with covariance matrix $E\left[v_1 v_1'\right]$.

(b) Error Correction Model

The procedure involved in obtaining an adaptive estimator for the error correction representation is similar to that for the regression model with ARMA errors, the chief differences being in the form of the model's score function and information matrix, and in the fact that we must nonparametrically estimate the entire vector-valued score function for the multivariate innovation density, rather than just its first element.

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:

$$(9) \qquad \Delta X_t = ABX_{t-1} + \sum_{j=1}^{k-1} \Phi_j \Delta X_{t-j} + \varepsilon_t \,,$$

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 $\psi(\varepsilon) = \left(\partial p(\varepsilon) / \partial\varepsilon\right) / p(\varepsilon)$, and whose finite, positive definite $q \times q$ information matrix we denote by $\Omega = \int \psi(\varepsilon)\psi(\varepsilon)' \, p(\varepsilon) d\varepsilon$.

We assume 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 (9) as

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

where $\Phi = [\Phi_1, \ldots, \Phi_{k-1}]$ and $Z_{t-1} = [\Delta X_{t-1}', \ldots, \Delta X_{t-k+1}']'$. We define $\alpha = vec(A), \varphi = vec(\Phi),$ and $\beta = vec(B_0)$, which we then gather into the m-dimensional full parameter vector $\theta = [\alpha', \varphi', \beta']'$, where $m = 2qr - r^2 + q^2(k-1)$. 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 $\delta_n = diag\left[n^{-1/2} I_s, n^{-1} I_{m-s}\right]$.

Analogously to the previous model, we define the matrix $H_{t-1}(\theta)$, which is central to our derivation of expressions for the model's score function and information matrix, and hence to the construction of an adaptive estimator, as follows:

$$H_{t-1}(\theta) = \left[\left(I_q \otimes W_{t-1}\right)', \left(I_q \otimes Z_{t-1}\right)', \left(-A' \otimes X_{2,t-1}\right)'\right]',$$

where $W_{t-1} = X_{1,t-1} - B_0 X_{2,t-1}$. We can the write the scaled score and information, respectively, as:

$$W_n(\theta) = -\sum_{t=1}^{n} \delta_n H_{t-1}(\theta) \psi(\varepsilon_t),$$

$$S_n(\theta) = \sum_{t=1}^{n} \delta_n H_{t-1}(\theta) H_{t-1}(\theta)' \delta_n.$$

Also analogously to the previous model, we must obtain nonparametric estimates of $\psi(\varepsilon)$ and Ω, and again we employ a Gaussian kernel technique. We write the kernel and symmetric density estimator respectively as follows:

$$\pi(x,\theta) = 1/\left(\sigma\sqrt{2\pi}\right)^q \exp\left(-|x|^2/2\sigma^2\right),$$

$$\hat{p}_{\sigma,t}(x,\theta) = (1/2(n-1))\sum_{\substack{i=1 \\ i\neq t}}^{n}\left\{\pi(x+\varepsilon_i(\theta),\sigma) + \pi(x-\varepsilon_i(\theta),\sigma)\right\},$$

where σ again denotes the bandwidth parameter, and the partial derivative of $\hat{p}_{\sigma,t}(x,\theta)$ with respect to the j^{th} element of x is denoted by $\hat{p}^j_{\sigma,t}(x,\theta)$. Our estimate of the j^{th} element of the score vector is given by

$$\hat{\psi}^j_{n,t}(x,\theta) = \begin{cases} \dfrac{\hat{p}^j_{\sigma,t}(x,\theta)}{\hat{p}_{\sigma,t}(x,\theta)} & if \begin{cases} \hat{p}_{\sigma,t}(x,\theta) \geq m^j_n \\ |x| \leq \alpha^j_n \\ \left|\hat{p}^j_{\sigma,t}(x,\theta)\right| \leq c^j_n \hat{p}_{\sigma,t}(x,\theta) \end{cases} \\ 0 & otherwise \end{cases}$$

where $c^j_n \to \infty, \alpha^j_n \to \infty, \sigma \to 0, m^j_n \to 0$, for every j=1,...,q. We then have $\hat{\psi}_{n,t}(x,\theta) = \left(\hat{\psi}^1_{n,t}(x,\theta),\ldots,\hat{\psi}^q_{n,t}(x,\theta)\right)'$.

With these nonparametric estimates in hand, we define our estimated model score and information as

$$\hat{W}_n(\theta) = -\sum_{t=1}^{n} \delta_n H_{t-1}(\theta)\hat{\psi}_{n,t}(\varepsilon_t(\theta),\theta)$$

and

$$\hat{S}_n(\theta) = \sum_{t=1}^{n} \delta_n H_{t-1}(\theta)\hat{\Omega}_n(\theta)H_{t-1}(\theta)'\delta_n,$$

where $\hat{\Omega}_n(\theta) = n^{-1}\sum_{t=1}^{n}\hat{\psi}_{n,t}(\varepsilon_t(\theta),\theta)\hat{\psi}_{n,t}(\varepsilon_t(\theta),\theta)'$. An adaptive estimator can then be computed, employing a δ_n^{-1}-consistent, discretized preliminary estimator θ_n^{**}, as follows:

$$\widetilde{\theta}_n = \theta_n^{**} + \delta_n \hat{S}_n(\theta_n^{**})^{-1}\hat{W}_n(\theta_n^{**}).$$

The asymptotic distribution of $\widetilde{\theta}_n$ is

$$\delta_n^{-1}(\widetilde{\theta}_n - \theta) \Rightarrow MN\left(0, S(\theta)^{-1}\right),$$

where the asymptotic matrix is

$$S(\theta) = diag\left[\Omega \otimes E\left[M_t M_t'\right], A'\Omega A \otimes \int_0^1 B_2 B_2'\right],$$

where $M_t = \left[W_t', Z_t'\right]'$ and B_2 is a Brownian motion with covariance matrix $E\left[\Delta X_{2t}\Delta X_{2t}'\right]$.

3. SOME SIMULATION EVIDENCE ON THE FINITE SAMPLE PERFORMANCE OF ADAPTIVE ESTIMATORS IN COINTEGRATED MODELS

That it is possible to asymptotically efficiently estimate the models presented above in the absence of parametric assumptions on the likelihood is a result that should be of considerable value and interest to empirical researchers. Nevertheless, the foregoing description of the technique of adaptive estimation leaves the empirical researcher with numerous questions and doubts regarding the implementation and performance of adaptive estimators in particular applications. Paramount among these concerns are the sensitivity of results to the selection of the smoothing and trimming parameters that typically appear in the expression for the kernel estimate of the score function of the density of the model's innovations, and the extent to which asymptotic theory accurately represents the finite sample efficiency gains to be obtained, a concern common in econometrics but heightened in the present case by the fact that kernel density estimates are being employed. Little published work has appeared to address these problems, with the Monte Carlo studies of Hsieh and Manski (1987) and Steigerwald (1992b) providing important exceptions. In the context of a linear non-cointegrating regression, Hsieh and Manski (1987) compare the finite sample performance of adaptive estimators with various alternative estimators, including ordinary least squares, for several different error distributions, using root mean squared error and interquartile range as the standards of comparison. They evaluate the sensitivity of the results to smoothing and trimming parameter selection by calculating adaptive estimators for a range of possible settings of these parameters.

In this section, we are concerned primarily with the extent to which theoretical asymptotic efficiency gains over Gaussian pseudo-MLE's, for the two representations of cointegrated systems described above, can be realized in a fairly small sample (100 observations) when no elaborate data-based methods are used to select the smoothing and trimming parameters (Hsieh and Manski (1987) found that the use of such methods can substantially improve estimator performance). Although we are not concerned with the question of *how* to optimally select smoothing and trimming parameters for given samples, we do conduct simulations for a variety of parameter settings in order to get some idea about how sensitive finite sample efficiency gains are to the settings selected.

Our primary criterion for evaluating the adaptive estimator's performance is the ratio of the mean squared error (MSE) of the estimator to the MSE of the Gaussian estimator. This statistic is used because we can compare it with the ratio of asymptotic variances to see how well the theoretical efficiency gains are realized. Recognizing the fact, discovered by Phillips (1994), that the Johansen estimator has a very heavy-tailed finite sample distribution, which lacks moments of any order, we calculate for the error correction representation a modified MSE ratio, using only the central 80 per cent of realizations of the estimators. We also report interdecile and interquartile ranges for the estimators. Since an estimator's MSE involves the square of its bias, we also report bias statistics in order to indicate the relative extents to which differences between estimators' MSE's are due to bias and variance. For the linear representation, we report all of these performance statistics for the least absolute deviations (LAD) estimator as well, in order to get some idea of the relative robustness properties of LAD and adaptive estimation. To provide some indication of the rate at which asymptotic efficiency gains can be approached with increasing sample size, we also report MSE statistics for samples of 250 and 500.

In subsection *(a)*, we describe our simulated models, which are special cases of the general frameworks described above. Subsection *(b)* discusses issues involved in the implementation of our study, including the selection of smoothing and trimming parameters and the finite sample distribution of the Johansen estimator. In subsection *(c)*, we report and discuss our simulation results.

(a) The Simulated Models

In this subsection we describe the data generating processes used for the two models. In both cases we consider bivariate models with sample sizes of 100, 250, and 500, and innovations $\varepsilon = (\varepsilon_1, \varepsilon_2)'$ drawn from the pdf $p(\varepsilon)$, which takes on each of the following four forms:

(I) variance-contaminated mixture of normals: $p(\varepsilon) = 0.9N(0,1/3\,I) + 0.1N(0,27I)$;

(ii) Student's *t* with 3 degrees of freedom: $p(\varepsilon) = t_3(0,3I)$;

(iii) bimodal mixture of normals: $p(\varepsilon) = 0.5N(\sqrt{2},I) + 0.5N(-\sqrt{2},I)$;

(iv) normal: $p(\varepsilon) = N(0,3I)$.

These densities are designed to each have the same covariance matrix, viz. $3I$. The Gaussian case is considered in order to gauge the extent of the finite sample efficiency loss we may obtain through the use of the adaptive estimator in place of the correctly specified MLE. In all cases, we run 1000 iterations. We now present the particular details of each simulation setup, in turn.

The error correction model has one cointegrating relationship and is derived from a VAR of order one, and so can be written as follows:

(10) $\Delta X_t = ab' X_{t-1} + \varepsilon_t$,

where $X_t = (X_{1t}, X_{2t})'$, $a = (a_1, a_2)'$, $b = (1, -b_0)'$, and ε_t is iid from the unknown density $p(\varepsilon)$. In estimating the model, the first element of b is restricted to equal unity in order to achieve identification, so that b_0 is the sole cointegrating coefficient. Our simulation study is concerned only with estimation of this parameter. In the simulations, we set $a_1 = a_2 = 1$ and $b_0 = 2$.

For the linear regression model, we assume iid errors and a univariate regressor whose first difference is iid and uncorrelated with the regression errors. The Gaussian MLE of this model is ordinary least squares, and adaptive estimation is analyzed by Jeganathan (1994)[2]. The data generating process is as follows:

(11) $X_{1t} = b_0 X_{2t} + \varepsilon_{1t}$

(12) $\Delta X_{2t} = \varepsilon_{2t}$

with $\varepsilon_t \equiv iid\ p(\varepsilon)$, with $p(\varepsilon)$ assumed to be elliptically symmetric. Again, we are concerned with estimation of the cointegrating parameter b_0, and in the simulations set $b_0 = 2$.

The asymptotic efficiency gains available through use of the adaptive estimators in these models can be measured by the ratios of their asymptotic variances to those of the relevant Gaussian pseudo-MLEs, whose asymptotic distributions are also mixed normal, but with larger variances. The ratios turn out to be the same for both models, as shown in Hodgson (1995a). In the Gaussian case, the ratio is obviously equal to unity. For the densities specified above, it equals approximately 0.47 for the *t*, 0.13 for the variance-contaminated mixture of normals, and 0.47 for the bimodal mixture of normals[3].

---

[2] If the independence assumption on the errors is dropped, then the fully modified OLS procedure of Phillips and Hansen (1990) is asymptotically equivalent to the Gaussian MLE, while adaptive estimation in the case where the errors admit an ARMA representation is described in Section 2.

[3] The first figure was obtained from Mitchell (1989), while the latter two were computed numerically.

(b) Implementation of the Study

The objective of this study is to evaluate the finite sample performance of the adaptive estimators of the models described above relative to alternative estimators, especially the Gaussian pseudo-MLEs but also LAD for the linear model. Two varieties of problems arise in implementing such a study. The first concerns the specification of the smoothing and trimming parameters called for in the nonparametric score estimator. The second deals with the standard adopted for comparing the performances of alternative estimators.

It is quite common for theoretical analyses of nonparametric kernel estimators to specify the asymptotic behaviour of bandwidth parameters without providing much direction to the applied investigator regarding the selection of such parameters given a particular sample. For the estimators described above, the problem of choosing the bandwidth σ is compounded by the fact that the trimming parameters m_n, c_n, and α_n must simultaneously be chosen[4]. Our approach is to compute the adaptive estimator for each sample at each of nine combinations of these parameters (for samples of 100; for larger samples we only use one combination because our main concern in such cases is with the effect of sample size on estimator performance). Following Hsieh and Manski (1987), we reduce the trimming parameter problem from three dimensions to one by selecting *h*, where $\alpha_n = h, c_n = h/3$, and $m_n = (6\pi)^{-1}\exp(-h^2/3)$. For *n*=100, we allow *h* to take the values 10, 15, and 20, while σ is set at 0.67, 0.77 (Silverman's (1986) rule-

[4] We delete the superscript *j* on the trimming parameters in the error correction model because the same trimming values will be used for estimating both components of the score vector in our simulations.

of-thumb bandwidth for bivariate density estimation problems), and 0.87. For n=250, we set (σ,h)=(.75,15) (Silverman's rule-of thumb is 0.65), while for n=500, the setting is (σ,h)=(.70,20) (rule-of thumb=.58). Hsieh and Manski (1987), analyzing a univariate model, reduced their trimming problem to one dimension in such a way that if the innovation density were normal, then all three parameters would be effective at the same point. Our settings roughly follow Hsieh and Manski's, although in our case they won't be effective at exactly the same point in the case of a Gaussian density.

As one of the objectives of Hsieh and Manski's (1987) study was to shed light on the question of optimal trimming and smoothing parameter selection, they calculated adaptive estimates for a much wider range of settings than do we, and also considered data-based bootstrap methods. Our goals here are less ambitious in that we vary the parameters not in order to acquire evidence on their optimal selection but in order to get a rough idea of the sensitivity of efficiency gains to their variation. We ask whether efficiency gains can be obtained though the crude choice of the rule-of thumb bandwidth with a moderate degree of trimming (h=15 corresponds to trimming observations that lie more than nine standard deviations from the origin; for samples of size 25 and 50, Hsieh and Manski (1987) advised trimming at about eight standard deviations), and if so, we seek to determine the sensitivity of these gains to slightly different settings of the parameters.

In speaking of efficiency gains, we assume the existence of some yardstick for their measurement. In the preceding section, we compared asymptotic efficiency in terms of asymptotic covariance. Using variance or MSE as a criterion in finite samples can be undesirable, however, since the estimators' finite sample distributions need not be normal, symmetric, or unbiased. They may not even have moments, as is the case for the Johansen estimator of the error correction model, as shown by Phillips (1994). Thus, although a comparison of MSE's is a potentially attractive finite sample analogue of the asymptotic variance ratio derived in the previous section, it also has clear shortcomings. In the next section, we report the ratios of the simulation MSE's between the adaptive and Gaussian estimators for linear regressions, but for the error correction model report truncated MSE's obtained by throwing out the upper and lower ten per cent of realized parameter estimates. We also report interdecile and interquartile ranges, which complement the latter two MSE statistics in providing evidence on the concentration of the estimators' distributions around the true parameter value.

(c) The Simulation Results

Linear Regression Model

Tables 1-4 report results on MSE, bias, and interquartile and interdecile ranges for the adaptive, OLS, and LAD estimators. The results are probably biased in disfavour of the adaptive estimates because, in their computation, the restriction that the regression errors and the regressor first differences are uncorrelated was not used. This means that in computing our density estimate, rather than using the estimated residuals $\hat{\varepsilon}_{1t}$, we use the residuals of an OLS regression of $\hat{\varepsilon}_{1t}$ on ε_{2t} (see Jeganathan (1994) and Subsection 2(a) above). This presumably adds some extra noise to the adaptive estimates.

The results for Gaussian innovations, given in Table 1, indicate that the losses that one should expect to obtain from using the adaptive estimator are moderate, ranging from

18% to 30%, and are fairly insensitive to smoothing and trimming parameter specification. The loss from using the robust LAD estimator is considerably larger, at 71%. The interquartile and interdecile range statistics support these conclusions, as they vary only slightly with the (σ,h) setting and indicate considerably larger dispersion for LAD. The bias statistics indicate slightly larger bias in the adaptive and LAD estimators than in OLS, but with differences small enough to affect the differences in MSE only negligibly.

The Student's t and variance contaminated mixed normal are both unimodal thick-tailed densities, and the results obtained for them, in Tables 2 and 3, respectively, are quite similar. Broadly speaking, the adaptive estimator provides moderate efficiency gains but is outperformed by LAD. The adaptive estimator provides efficiency gains for the t distribution ranging from 9 to 24 per cent, while the gain from using the LAD is 40 per cent. For the mixture of normals, the adaptive range is 1 to 43 per cent, while the gain for LAD is a very impressive 82 per cent. The interquartile and interdecile ranges tell a similar story, while the bias statistics are again too small to account for much of the MSE differences.

Finally, we present results for the bimodal mixture of normals in Table 4. The results are quite drastic. The OLS and LAD estimates are heavily biased, as is clearly illustrated by a perusal of the interquartile and interdecile range statistics. The efficiency gains to the adaptive estimator are large, ranging from 72 to 74 per cent. In contrast, the LAD estimator has very poor properties, with an MSE of more than two and a half times that of OLS.

Overall, the adaptive estimator performs fairly well in comparison with OLS and LAD. For none of the error distributions considered does it perform the worst of the three, whereas the other two estimators are both worst in two cases. For unimodal thick-tailed distributions, LAD performs extremely well, but the adaptive estimator still improves significantly upon OLS. In addition, the adaptive estimator does not share the extremely bad performance of LAD for the Gaussian and bimodal mixed normal distributions. As noted above, the adaptive estimator is at a disadvantage in these simulations because we have not used the restriction that the errors are uncorrelated with the regressors. Also mitigating against the adaptive estimator is the fact that we have not used the fact that the distributions considered are elliptically symmetric, which would allow us to reduce the dimensionality of our density estimation problem from two to one, which would definitely improve the estimator's efficiency.

The Error Correction Model

The simulation results for the error correction model are presented in Tables 5-8. The patterns here are similar to those for the linear regression model. The losses from employing the adaptive estimator in the case of Gaussian innovations are small to moderate, ranging from 7 to 21 percent. The gains for non-Gaussian innovations are moderate to moderately large, and go a significant proportion of the way to the asymptotic efficiency gains. The results are fairly robust to smoothing and trimming parameter selection and are negligibly affected by finite sample bias. For the Student's t model, the gains are in the 9 to 22 per cent range, for the variance contaminated mixture of normals they range from 23 to 65 per cent, and for the bimodal mixture of normals the range is 48 to 59 per cent.

Larger Sample Sizes

Tables 9 and 10 present MSE and truncated MSE results for the linear regression model and error correction model, respectively, for sample sizes of 250 and 500. We only consider one setting of the smoothing and trimming parameters in computing the adaptive estimates. For the linear model, the qualitative nature of the results changes little as the sample size is increased. There are moderate improvements in the adaptive estimator relative to both OLS and LAD, with little change in the relative performances of the latter two estimators, except that the losses from using LAD in the Gaussian case fall from 71 per cent when n=100 to 46 per cent when and n=500, and increase from two and a half times to over three times as large for the bimodal mixture of normals model. There also appear to be improvements in the adaptive estimator relative to the Johansen estimator with increasing sample size. For the normal model and both mixed normal models, the truncated MSE ratio between the estimators falls as the sample is increased from 250 to 500 and in the latter case is approximately equal to the best ratio that was obtained by varying the smoothing and trimming parameters for the sample of 100. The improvements are more significant for the Student's t model, where the best truncated MSE ratio for n=100 is 0.78 while the ratio is 0.62 and 0.61 for respective sample sizes of 250 and 500.

TABLE 1: LINEAR MODEL, NORMAL ERRORS (n=100)

(a) MSE of Adaptive Estimator and Ratio with MSE of OLS

$\sigma \backslash h$	10	15	20
.67	5.929×10^{-4} 1.26	5.971×10^{-4} 1.27	6.107×10^{-4} 1.30
.77	5.729×10^{-4} 1.22	5.865×10^{-4} 1.25	5.867×10^{-4} 1.25
.87	5.542×10^{-4} 1.18	5.605×10^{-4} 1.20	5.605×10^{-4} 1.20

$MSE_{OLS} = 4.690 \times 10^{-4}$

$MSE_{LAD} = 8.012 \times 10^{-4}$

$MSE_{LAD} / MSE_{OLS} = 1.71$

(b) Interdecile and Interquartile Ranges of Adaptive Estimator

$\sigma \backslash h$	10	15	20
.67	1.974-2.028 1.988-2.013	1.974-2.028 1.989-2.013	1.973-2.029 1.989-2.013
.77	1.974-2.028 1.989-2.013	1.974-2.028 1.989-2.013	1.974-2.028 1.989-2.013
.87	1.974-2.028 1.990-2.013	1.974-2.028 1.990-2.013	1.974-2.028 1.990-2.013

OLS: 1.977-2.025
1.989-2.011
LAD: 1.969-2.033
1.987-2.014

(c) Bias

Adaptive (σ=.87,h=15)	1.292×10^{-3}
OLS	8.435×10^{-4}
LAD	1.047×10^{-3}

TABLE 2: LINEAR MODEL, STUDENT'S t ERRORS (n=100)

(a) MSE of Adaptive Estimator and Ratio with MSE of OLS

$\sigma \setminus h$	10	15	20
.67	5.151×10^{-4} 0.78	5.663×10^{-4} 0.85	5.755×10^{-4} 0.87
.77	5.042×10^{-4} 0.76	5.494×10^{-4} 0.83	5.980×10^{-4} 0.90
.87	5.149×10^{-4} 0.78	5.623×10^{-4} 0.85	6.019×10^{-4} 0.91

$MSE_{OLS} = 6.625 \times 10^{-4}$

$MSE_{LAD} = 3.985 \times 10^{-4}$

$MSE_{LAD} / MSE_{OLS} = 0.60$

(b) Interdecile and Interquartile Ranges of Adaptive Estimator

$\sigma \setminus h$	10	15	20
.67	1.976-2.024 1.990-2.011	1.975-2.027 1.990-2.011	1.973-2.025 1.989-2.011
.77	1.975-2.024 1.990-2.011	1.975-2.026 1.989-2.011	1.974-2.026 1.989-2.012
.87	1.975-2.026 1.990-2.012	1.975-2.026 1.989-2.012	1.974-2.026 1.988-2.012

OLS: 1.973-2.027
1.989-2.012
LAD: 1.978-2.020
1.991-2.010

(c) Bias

Adaptive (σ=.87,h=15)	3.688×10^{-4}
OLS	-1.369×10^{-4}
LAD	-1.847×10^{-4}

TABLE 3: LINEAR MODEL, VAR. CONTAM. MN ERRORS (n=100)

(a) MSE of Adaptive Estimator and Ratio with MSE of OLS

$\sigma \setminus h$	10	15	20
.67	4.750×10^{-4} 0.59	4.564×10^{-4} 0.57	4.749×10^{-4} 0.59
.77	6.007×10^{-4} 0.75	5.778×10^{-4} 0.72	5.762×10^{-4} 0.72
.87	7.940×10^{-4} 0.99	6.887×10^{-4} 0.86	6.869×10^{-4} 0.85

$MSE_{OLS} = 8.040 \times 10^{-4}$

$MSE_{LAD} = 1.468 \times 10^{-4}$

$MSE_{LAD} / MSE_{OLS} = 0.18$

(b) Interdecile and Interquartile Ranges of Adaptive Estimator

$\sigma \setminus h$	10	15	20
.67	1.980-2.020 1.992-2.009	1.981-2.021 1.992-2.008	1.980-2.023 1.992-2.008
.77	1.979-2.024 1.991-2.011	1.978-2.023 1.991-2.010	1.977-2.023 1.990-2.010
.87	1.974-2.028 1.988-2.013	1.974-2.026 1.990-2.011	1.974-2.026 1.989-2.011

OLS: 1.971-2.028
1.986-2.012
LAD: 1.987-2.012
1.994-2.005

(c) Bias

Adaptive (σ=.87,h=15)	5.480×10^{-4}
OLS	-8.004×10^{-5}
LAD	-1.275×10^{-4}

TABLE 4: LINEAR MODEL, BIMODAL MN ERRORS (n=100)

(a) MSE of Adaptive Estimator and Ratio with MSE of OLS

$\sigma \setminus h$	10	15	20
.67	3.773×10^{-4} 0.26	3.786×10^{-4} 0.26	3.765×10^{-4} 0.26
.77	3.787×10^{-4} 0.26	3.775×10^{-4} 0.26	3.775×10^{-4} 0.26
.87	4.077×10^{-4} 0.28	4.038×10^{-4} 0.28	4.038×10^{-4} 0.28

$MSE_{OLS} = 1.453 \times 10^{-3}$

$MSE_{LAD} = 3.715 \times 10^{-3}$

$MSE_{LAD} / MSE_{OLS} = 2.56$

(b) Interdecile and Interquartile Ranges of Adaptive Estimator

$\sigma \setminus h$	10	15	20
.67	1.978-2.021 1.989-2.008	1.978-2.021 1.989-2.008	1.978-2.021 1.989-2.008
.77	1.974-2.016 1.986-2.005	1.974-2.016 1.986-2.005	1.974-2.016 1.986-2.005
.87	1.971-2.013 1.983-2.003	1.971-2.013 1.983-2.003	1.971-2.013 1.983-2.003

OLS: 2.001-2.061
2.008-2.037
LAD: 2.002-2.099
2.016-2.060

(c) Bias

Adaptive (σ=.87,h=15)	-7.170×10^{-3}
OLS	2.633×10^{-2}
LAD	4.371×10^{-2}

TABLE 5: ERROR CORRECTION MODEL, NORMAL ERRORS (n=100)
(TMSE denotes the MSE of the central 80% of the empirical distribution)

(a) TMSE of Adaptive Estimator and Ratio with TMSE of Johansen

$\sigma \setminus h$	10	15	20
.67	4.937×10^{-5} 1.19	4.851×10^{-5} 1.17	5.015×10^{-5} 1.21
.77	4.662×10^{-5} 1.12	4.769×10^{-5} 1.15	4.768×10^{-5} 1.15
.87	4.439×10^{-5} 1.07	4.529×10^{-5} 1.09	4.529×10^{-5} 1.09

$TMSE_{JOH} = 4.156 \times 10^{-5}$

(b) Interdecile and Interquartile Ranges of Adaptive Estimator

$\sigma \setminus h$	10	15	20
.67	1.986-2.015 1.993-2.007	1.986-2.015 1.993-2.007	1.985-2.015 1.993-2.007
.77	1.986-2.014 1.993-2.007	1.986-2.014 1.993-2.007	1.986-2.014 1.993-2.007
.87	1.986-2.014 1.994-2.007	1.986-2.014 1.994-2.007	1.986-2.014 1.994-2.007

JOH: 1.987-2.014
1.994-2.007

(c) Bias (over central 80% of distribution)

Adaptive (σ=.87,h=15)	2.091×10^{-4}
Johansen	2.818×10^{-4}

TABLE 6: ERROR CORRECTION MODEL, STUDENT'S *t* ERRORS (n=100)
(TMSE denotes the MSE of the central 80% of the empirical distribution)

(a) TMSE of Adaptive Estimator and Ratio with TMSE of Johansen

$\sigma \setminus h$	10	15	20
.67	3.085×10^{-5} 0.79	3.245×10^{-5} 0.83	3.521×10^{-5} 0.90
.77	3.064×10^{-5} 0.78	3.274×10^{-5} 0.83	3.570×10^{-5} 0.91
.87	3.097×10^{-5} 0.79	3.312×10^{-5} 0.84	3.470×10^{-5} 0.88

$TMSE_{JOH} = 3.926 \times 10^{-5}$

(b) Interdecile and Interquartile Ranges of Adaptive Estimator

$\sigma \setminus h$	10	15	20
.67	1.988-2.012 1.995-2.005	1.988-2.013 1.995-2.005	1.988-2.013 1.994-2.006
.77	1.988-2.012 1.995-2.005	1.988-2.012 1.995-2.006	1.988-2.013 1.994-2.006
.87	1.988-2.012 1.995-2.006	1.988-2.012 1.995-2.006	1.988-2.012 1.995-2.006

JOH: 1.987-2.013
1.994-2.006

(c) Bias (over central 80% of distribution)

Adaptive (σ=.87,h=15)	2.509×10^{-4}
Johansen	3.731×10^{-4}

TABLE 7: ERROR CORRECTION MODEL, VAR. CONTAM. MN ERRORS (n=100)
(TMSE denotes the MSE of the central 80% of the empirical distribution)

(a) TMSE of Adaptive Estimator and Ratio with TMSE of Johansen

$\sigma \setminus h$	10	15	20
.67	2.108×10^{-5} 0.49	1.487×10^{-5} 0.35	1.630×10^{-5} 0.38
.77	2.722×10^{-5} 0.63	1.778×10^{-5} 0.41	1.906×10^{-5} 0.44
.87	3.323×10^{-5} 0.77	2.272×10^{-5} 0.53	2.279×10^{-5} 0.53

$TMSE_{JOH} = 4.306 \times 10^{-5}$

(b) Interdecile and Interquartile Ranges of Adaptive Estimator

$\sigma \setminus h$	10	15	20
.67	1.989-2.009 1.995-2.004	1.992-2.009 1.996-2.003	1.991-2.009 1.996-2.004
.77	1.988-2.011 1.995-2.005	1.990-2.009 1.996-2.003	1.991-2.010 1.996-2.004
.87	1.987-2.012 1.994-2.005	1.990-2.010 1.996-2.004	1.990-2.010 1.996-2.005

JOH: 1.987-2.016
1.995-2.007

(c) Bias (over central 80% of distribution)

Adaptive (σ=.87,h=15)	4.941×10^{-6}
Johansen	7.799×10^{-4}

TABLE 8: ERROR CORRECTION MODEL, BIMODAL MN ERRORS (n=100)
(TMSE denotes the MSE of the central 80% of the empirical distribution)

(a) TMSE of Adaptive Estimator and Ratio with TMSE of Johansen

$\sigma \setminus h$	10	15	20
.67	9.246×10^{-5} 0.41	9.191×10^{-5} 0.41	9.234×10^{-5} 0.41
.77	1.019×10^{-4} 0.46	1.014×10^{-5} 0.45	1.014×10^{-4} 0.45
.87	1.159×10^{-4} 0.52	1.165×10^{-4} 0.52	1.165×10^{-4} 0.52

$TMSE_{JOH} = 2.231\times10^{-4}$

(b) Interdecile and Interquartile Ranges of Adaptive Estimator

$\sigma \setminus h$	10	15	20
.67	1.979-2.022 1.991-2.009	1.979-2.022 1.992-2.009	1.979-2.022 1.992-2.009
.77	1.977-2.023 1.990-2.009	1.977-2.023 1.990-2.009	1.977-2.023 1.990-2.009
.87	1.976-2.024 1.990-2.010	1.976-2.024 1.990-2.010	1.976-2.024 1.990-2.010

JOH: 1.973-2.035
1.988-2.015

(c) Bias (over central 80% of distribution)

Adaptive (σ=.87,h=15)	-4.333×10^{-4}
Johansen	2.161×10^{-3}

TABLE 9: MSE RESULTS FOR LINEAR MODEL AND LARGER SAMPLES
(adaptive estimator settings of (σ ,h) are (.75,15) for n=250 and (.7,20) for n=500)

(a) Normal Errors

	n=250	n=500
OLS	7.775×10^{-5}	2.233×10^{-5}
Adaptive	8.809×10^{-5}	2.435×10^{-5}
Adaptive/OLS	1.13	1.09
LAD	1.343×10^{-4}	3.267×10^{-5}
LAD/OLS	1.73	1.46

(B) Student's t Errors

	n=250	n=500
OLS	8.714×10^{-5}	2.231×10^{-5}
Adaptive	8.215×10^{-5}	1.673×10^{-5}
Adaptive/OLS	0.94	0.75
LAD	5.707×10^{-5}	1.469×10^{-5}
LAD/OLS	0.65	0.66

(c) Var. Contam. MN Errors

	n=250	n=500
OLS	1.019×10^{-4}	2.454×10^{-5}
Adaptive	6.123×10^{-5}	1.337×10^{-5}
Adaptive/OLS	0.60	0.54
LAD	2.010×10^{-5}	4.696×10^{-6}
LAD/OLS	0.20	0.19

(d) Bimodal MN Errors

	n=250	n=500
OLS	2.122×10^{-4}	5.846×10^{-5}
Adaptive	7.227×10^{-5}	2.007×10^{-5}
Adaptive/OLS	0.34	0.34
LAD	6.893×10^{-4}	1.970×10^{-4}
LAD/OLS	3.25	3.37

TABLE 10: TMSE RESULTS FOR ECM AND LARGER SAMPLES
(adaptive estimator settings of (σ,h) are (.75,15) for n=250 and (.7,20) for n=500)
(TMSE is MSE for central 80% of distribution of estimator)

(a) Normal Errors

	n=250	n=500
Johansen	6.458×10^{-6}	1.628×10^{-6}
Adaptive	7.047×10^{-6}	1.772×10^{-6}
Adaptive/Joh	1.09	1.09

(b) Student's t Errors

	n=250	n=500
Johansen	6.613×10^{-6}	1.610×10^{-6}
Adaptive	4.122×10^{-6}	9.790×10^{-7}
Adaptive/Joh	0.62	0.61

(c) Var. Contam. MN Errors

	n=250	n=500
Johansen	6.524×10^{-6}	1.556×10^{-6}
Adaptive	2.895×10^{-6}	5.967×10^{-7}
Adaptive/Joh	0.44	0.38

(d) Bimodal MN Errors

	n=250	n=500
OLS	3.357×10^{-5}	7.882×10^{-6}
Adaptive	1.577×10^{-5}	3.241×10^{-6}
Adaptive/OLS	0.47	0.41

4. FORWARD EXCHANGE MARKET UNBIASEDNESS, COINTEGRATION, AND ADAPTIVE ESTIMATION

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 both in terms of its bearing on the question of the efficiency of speculative markets, and in terms of 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 latter aspect is emphasized by McCallum (1994); both aspects are discussed by Baillie and McMahon (1989)).

There are two alternative approaches to modeling 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 analyzed by Hakkio and Rush (1989) and Barnhart and Szakmary (1991).

Working with an error correction representation has the advantage of providing a unified framework for modeling and estimating parameters associated with both the long-run and short-run aspects of the forward unbiasedness problem. In this section, we estimate an error correction representation with emphasis placed upon the cointegrating parameter (although we hope to devote attention to the error correction coefficients in future work). In empirical applications, this parameter is typically estimated using ordinary least squares or some variant (see, for example, McCallum (1994), Hakkio and Rush (1989), Corbae, Lim, and Ouliaris (1992), Baillie and Bollerslev (1989), Baillie, Lippens, and McMahon (1983) and 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-a-vis the normal. The lack of robustness of least squares procedures in the presence of significant deviations from Gaussianity motivated Phillips (1993a), in an application to the Australian dollar - U.S. dollar exchange rate, to employ a robust least absolute deviations estimator for the forward market unbiasedness cointegrating regression. Phillips found that inference in such models can be quite sensitive to the estimator used and advised against the use of least squares and Gaussian reduced rank regression procedures.

We report the results of forward market unbiasedness cointegrating parameter estimation for the Canadian dollar's exchange rate with respect to the U.S. dollar using daily data from the early 1990's. Noting Phillips' objections to the use of Gaussian procedures such as least squares and the Johansen approach, the estimation strategy followed here employs the adaptive estimation techniques developed in Hodgson (1995c) for error correction representations and described in Section 2. The illustration of adaptive estimation in action should be of considerable interest in its own right. Adaptive estimators have appeared mainly in the theoretical statistics and econometrics literature.

Several questions arise regarding the actual empirical implementation of adaptive estimators. These questions are addressed, discussed, and illustrated in the analysis reported below.

(a) The Empirical Model

We consider the cointegrating relationship between the logarithm of the 90-day forward exchange rate of the Canadian dollar in terms of the U.S. dollar, denoted by f_t, and the logarithm of the spot exchange rate on the date of delivery of the forward contract sold at *t*, denoted by s_{t+90}. We use daily noon rates throughout, with *t* running from November 29, 1990 to June 30, 1993, for a sample of 650 observations. Since we are estimating an error correction representation, eight additional days of data immediately preceding this sample period are used to fit an eighth-order VAR. The data were obtained from the Bank of Canada and we matched the appropriate spot rate with each day's forward rate according to the procedure described in Cornell (1989).

We stack the forward rate and the future spot rate, as of period *t*, into the data vector $X_t = (s_{t+90}, f_t)$, which we model as being generated by an error correction model as follows:

$$(13) \qquad \Delta X_t = AB' X_{t-1} + \sum_{j=1}^{k-1} \Phi_j \Delta X_{t-j} + \varepsilon_t \,,$$

where A is a vector of error correction coefficients, *B* is a cointegrating vector, *k* is the number of lags in the VAR from which (13) is obtained, *t* runs from 1 to *n*=650, and $\{\varepsilon_t\}$ is a sequence of iid innovations drawn from the unknown bivariate density $p(\varepsilon)$, which we assume to be symmetric. We assume an identified model by imposing the normalization $B' = [1, -b]$. Our concern in this paper is with efficiently estimating the cointegrating parameter b.

Some comments on this model are in order. First, we proceed under the maintained hypothesis that the cointegrating relationship has a zero intercept. Most studies also estimate an intercept; we choose not to because our focus is on the efficient estimation of the slope parameter and because the equality to zero of the intercept is generally less debatable than the equality to unity of the slope. Second, the symmetry of $p(\varepsilon)$ is a strong assumption (although considerably more general than the standard assumption of Gaussianity) necessitated by the fact that the analysis in Hodgson (1995c) relies on this assumption. Extension of the model to allow for a more general class of densities for $p(\varepsilon)$ is a topic for further research. Third, the selection of the lag order k is problematic. The series $\{s_{t+90} - bf_t\}$ follows an MA(89) process. Since (13) essentially involves fitting an autoregression to this series, k should be infinite. Our approach to this problem is rather crude. Rather than employing any advanced order selection procedures or information criteria, we set k=8 and test the estimated residuals for autocorrelation, a procedure also adopted by Johansen and Juselius (1990).

Several empirical studies have attempted to estimate b, for different currencies, using different frequencies of data observation and time periods, and using different estimation procedures. Most studies have estimated a linear cointegrating regression using some variant of ordinary least squares. We shall argue below that because exchange rate data, especially those sampled at short intervals, tend to be driven by innovations whose distributions are more thick-tailed than the Gaussian, least squares estimators are inefficient and can be improved upon by adaptive estimators. First, we present in Table 11 some of the results obtained in the literature for the estimation of *b*. This is not an exhaustive list of all work estimating the forward unbiasedness parameter, nor even of all the results obtained in the studies quoted.

TABLE 11: SELECTED ESTIMATES OF *b*

Study and data	Estimator	Exchange rate *	Estimate (S.E.)
McCallum (1994) monthly, 1/78-7/90, 30-day	OLS	US/GM US/BP	.9898 (.016) .9770 (.016)
Barnhart, Szakmary (1991) monthly, 1/74-11/88, 30-day	iterative SUR	US/CD US/GM	.987 (.009) .975 (.011)
Baillie, Bollerslev (1989) daily, 3/80-1/85, 30-day	OLS	CD/US JY/US SF/US	.9599 (NR) .8476 (NR) .9756 (NR)
Corbae, Lim, Ouliaris (1992) weekly, 1/76-1/85, 90-day	CCR	CD/US SF/US FF/US JY/US	.9777 (.071) .9386 (.043) .9417 (.037) .8670 (.049)
Baillie, Lippens, McMahon (1983) four-weekly, 6/73-4/80 or 12/77-5/80, 30-day	OLS	BP/US CD/US FF/US SF/US	.9560 (.0219) .6418 (.1143) .8840 (.0397) .7974 (.1204)
Phillips (1993a) daily, 1/84-4/91, 90-day	RRR	AD/US	.934 (NR)

* x/y denotes currency x in terms of currency y; US=US dollar, GM=German mark, BP= British pound, CD=Canadian dollar, JY=Japanese yen, SF=Swiss franc, FF=French franc, AD=Australian dollar

As mentioned above, the estimates presented in Table 11 are just a selection of all the estimates of the forward unbiasedness cointegrating parameter reported in the literature[5]. The estimates quoted are generally somewhat farther from unity than those not quoted. Nevertheless, the former share with the latter two characteristics: first, they

[5] In all the regressions reported, an intercept is estimated in the cointegrating relationship.

are obtained from estimation techniques that are only asymptotically efficient under Gaussianity assumptions on a model's underlying innovations; and second, they are almost always less than unity. We have chosen to report some of the more extreme deviations from unity, but the result that b is usually estimated to be less than one is quite general. As we shall report below, the Johansen (1988) reduced rank regression estimate of b for our data set is also less than one. These findings could be due to a failure of the forward unbiasedness hypothesis, but they may also be a statistical phenomenon arising from the fact that inefficient estimators are being used for highly non-Gaussian data. That the estimates are always away from unity in the same direction suggests that bias, rather than dispersion, is the main problem if the problem is indeed statistical in nature. However, this suggestion is mitigated by the fact that the various estimates reported in the literature are not independent of one another, as they all involve exchange rates between some currency and the U.S. dollar and are usually for very similar sample periods.

(b) Empirical Results

The logarithms of the spot and forward rate series are plotted in Figure 1 and the results of the Phillips (1987) Z_α and Z_t unit root tests and the Johansen and Juselius (1990) trace and maximum eigenvalue cointegration tests, with k in (13) set at 8, are presented in Tables 12 and 13, respectively. For both series, the null hypothesis of a unit root is easily accepted by both statistics, while the null hypothesis that the variables are not cointegrated is strongly rejected by both statistics reported.

TABLE 12: RESULTS OF UNIT ROOT TESTS

Statistic	Spot rate	Forward rate	90% Crit. value
Phillips (1987) Z_α	0.96	-0.16	-11.3
Phillips (1987) Z_t	0.69	-0.10	-2.57

Note: The autoregression contained a constant and the spectral density matrix was estimated using a Parzen window and a lag truncation of 10.

TABLE 13: RESULTS OF COINTEGRATION TESTS

Statistic	Value	95% C.V.	97.5% C.V.
Trace	21.56	20.17	22.20
Max. Eigenvalue	18.20	15.75	17.62

Note: The critical values are for the null hypothesis of no cointegration and are obtained from Table A3 in Johansen and Juselius (1990).

The Johansen (1988) reduced rank regression Gaussian pseudo-ML estimate of b, with k=8, is $\hat{b}_{JOH}$=0.937 with an estimated asymptotic standard error of 0.033. We then obtain estimates $\{\hat{\varepsilon}_t\}$ of the residuals by running an OLS regression of ΔX_t on seven own lags and a lagged error correction term. The estimated residuals are plotted in Figures 2 and 3, and Gaussian kernel estimates of the marginal density of each residual series, calculated using Silverman's (1986) rule-of-thumb bandwidth, are plotted in Figures 4 and 5, along with plots of normal densities whose variances equal the sample variances of the respective residual series. Table 14 reports the results of Box-Pierce (1970) Q statistics for autocorrelation and Jarque-Bera (1980) tests for non-Gaussianity and excess kurtosis as applied to both residual series. The results strongly suggest that the residuals are uncorrelated and non-Gaussian, with excessively thick tails. The latter conclusion is supported by a glance at the density estimates in Figures 4 and 5. However, our assumptions that the innovations to the system are iid from a symmetric density are placed into question by Figures 2-5. The time series of the estimated residuals in Figures 2 and 3 seem to indicate the presence of conditional heteroskedasticity, while the density estimates in Figures 4 and 5 appear to be significantly asymmetric.

TABLE 14: AUTOCORRELATION AND NORMALITY TESTS ON OLS RESIDUALS

Statistic and asymp. distn.	Value for $\{\hat{\varepsilon}_{1t}\}$	Value for $\{\hat{\varepsilon}_{2t}\}$
Box-Pierce Q(5) $(\sim \chi_5^2)$	0.03	0.15
Box-Pierce Q(10) $(\sim \chi_{10}^2)$	2.76	5.33
Box-Pierce Q(20) $(\sim \chi_{20}^2)$	14.01	16.80
Jarque-Bera skewness-kurtosis $(\sim \chi_2^2)$	147.99	169.50
Jarque-Bera kurtosis (~N(0,24))	53.58	62.29

Note: $\{\hat{\varepsilon}_{1t}\}$ denote the estimated residuals to the spot rate equation, and $\{\hat{\varepsilon}_{2t}\}$ are the estimated residuals to the forward rate equation.

We use the Johansen (1988) estimate of the cointegrating parameter and the OLS estimates of the short-run coefficients as our preliminary estimator in arriving at an adaptive estimate. We compute adaptive estimates for a grid of smoothing and trimming parameter estimates and report the estimates and estimated asymptotic standard errors in Table 15. We reduce the trimming parameter selection problem to a univariate one in a manner similar to that of Hsieh and Manski (1987), but with modifications to account for the scale of our data. Our estimated residual series both have standard deviations in the neighbourhood of .0026, so we reduce the trimming parameter problem to the selection of r, where $\alpha_n = r$, $c_n = r/(.002)^2$, and $m_n = \exp(-r^2/(.002)^2)$. We allow r to take the values 0.04, 0.05, and 0.06, meaning that we trim at roughly 8, 10, and 12 standard deviations from the origin. Regarding the values of the smoothing parameter σ that we employ, the Silverman (1986) rule-of-thumb value for estimating the bivariate density of our estimated residual series is 0.00084, so we allow σ to assume the values 0.00084, 0.00093, 0.00103, and 0.00113.

TABLE 15: ADAPTIVE ESTIMATES AND STANDARD ERRORS FOR VARIOUS SMOOTHING AND TRIMMING PARAMETER SETTINGS

$\sigma \setminus r$	0.04	0.05	0.06
0.00084	.995 (.026)	.995 (.026)	.995 (.026)
0.00093	.997 (.028)	.997 (.028)	.997 (.028)
0.00103	.998 (.030)	.998 (.030)	.998 (.030)
0.00113	.997 (.032)	.997 (.032)	.997 (.032)

As can be seen from Table 15, the adaptive estimator adjusts the Gaussian pseudo-MLE much closer to unity, a result that is quite robust to selection of smoothing and trimming parameters. In fact, the results are entirely insensitive to variation in the trimming parameter, suggesting that over the range of variation considered the amount of trimming is not changing. Since this range is fairly substantial relative to the scale of the data, it is very probable that little or no trimming is taking place at the settings considered. The estimated asymptotic standard errors are more sensitive to variation in the bandwidth than are the parameter estimates. They increase with the bandwidth σ, an expected result because the inverse of the estimated asymptotic information matrix of the innovation density appears in the expression for the parameter estimator's covariance matrix, and our kernel estimate of this information matrix decreases with σ, the reason being that with more smoothing the tails of the kernel density estimate descend to zero more slowly, so that the estimated score function at tail observations is smaller, causing the sum of the squared score estimates, i.e. the information matrix estimate, to be smaller.

5. CONCLUDING REMARKS

This study has investigated both the finite sample efficiency gains to be obtained by using adaptive estimators for two popular representations of cointegrated models and the effects on inference regarding the forward exchange market unbiasedness hypothesis of employing adaptive rather than Gaussian pseudo-ML estimators. In the simulation study, we found that for non-Gaussian data the gains over the Gaussian pseudo-MLE were significant for samples ranging in size from 100 to 500 while the losses when the data generating process actually is Gaussian were fairly modest. The results were not very sensitive to the user-specified trimming and smoothing parameter values selected, at least for bandwidths in a rough neighbourhood of Silverman's (1986) rule-of-thumb value. For the linear cointegrating regression model, we also computed robust LAD estimates in order to evaluate the robustness properties of the adaptive estimator. It stood up to comparison quite well, being outperformed by LAD for thick-tailed unimodal densities but performing much better for Gaussian and bimodal mixed normal innovations. We conjecture that the adaptive estimator's performance could be improved for the distributions considered by making use of elliptical symmetry restrictions.

The empirical study applies the methodology of adaptive estimation in error correction models, developed in Hodgson (1995c), to the problem of estimating a forward exchange market unbiasedness model. A survey of the literature finds that such models are commonly estimated by Gaussian pseudo-MLE techniques which often obtain estimates of the cointegrating parameter between spot exchange rates and lagged forward rates that are less than the value of one implied by the forward unbiasedness hypothesis. We estimate a model of the Canada-U.S. exchange rate with daily data using Johansen's (1988) Gaussian pseudo-MLE for error correction models and obtain a result considerably less than unity. We find that the estimated residuals have thicker tails than the normal, a common result for daily exchange rate data. We then adaptively estimate the model and obtain estimates much closer to unity, a result robust to smoothing and trimming parameter selection. The adaptive estimator produces a much sharper inference in favour of the forward market unbiasedness hypothesis than does the Gaussian pseudo-MLE.

FIGURE 1: LOGARITHMS OF SPOT AND FORWARD RATES

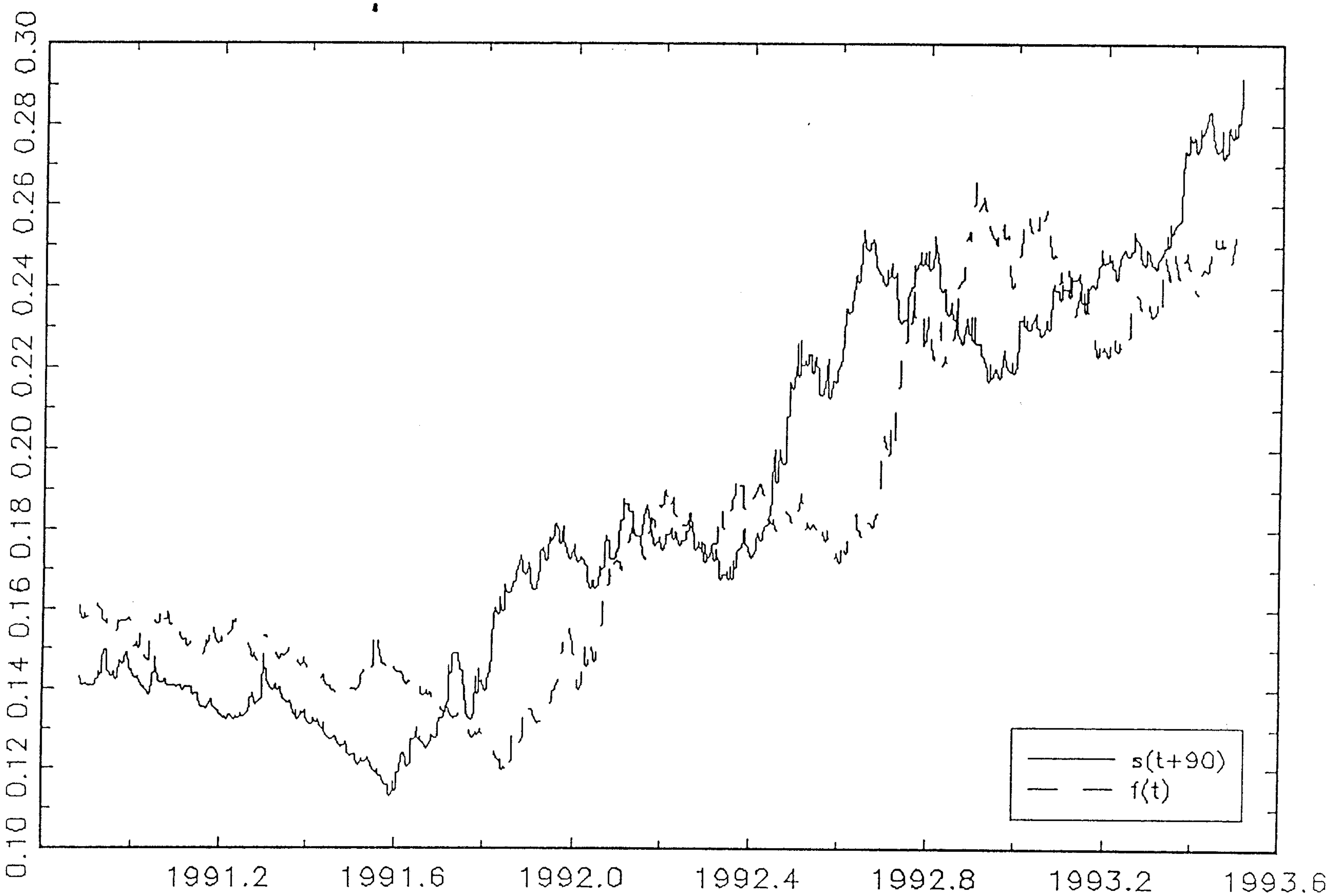


FIGURE 2: SPOT EQUATION ESTIMATED RESIDUALS

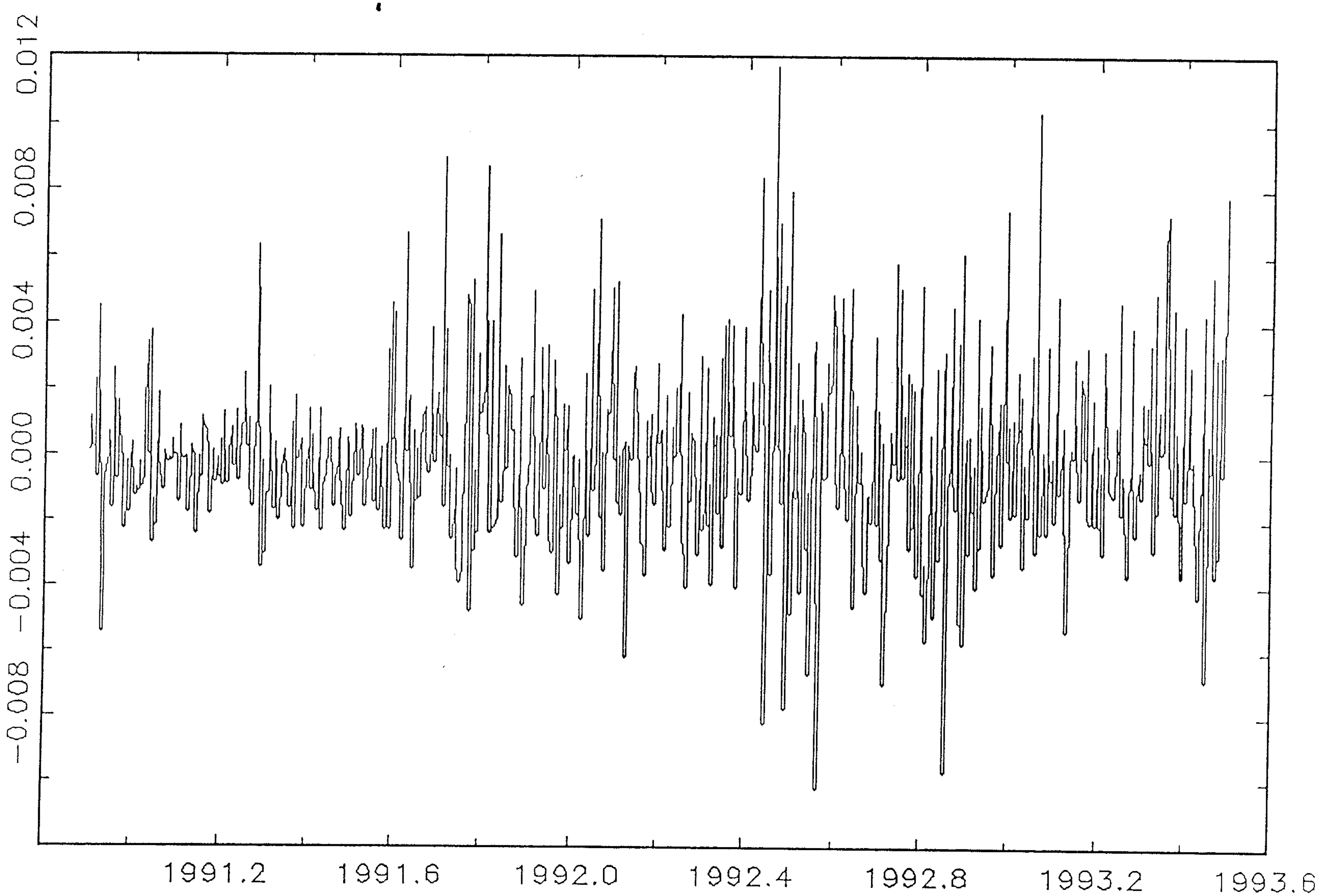


FIGURE 3: FORWARD EQUATION ESTIMATED RESIDUALS

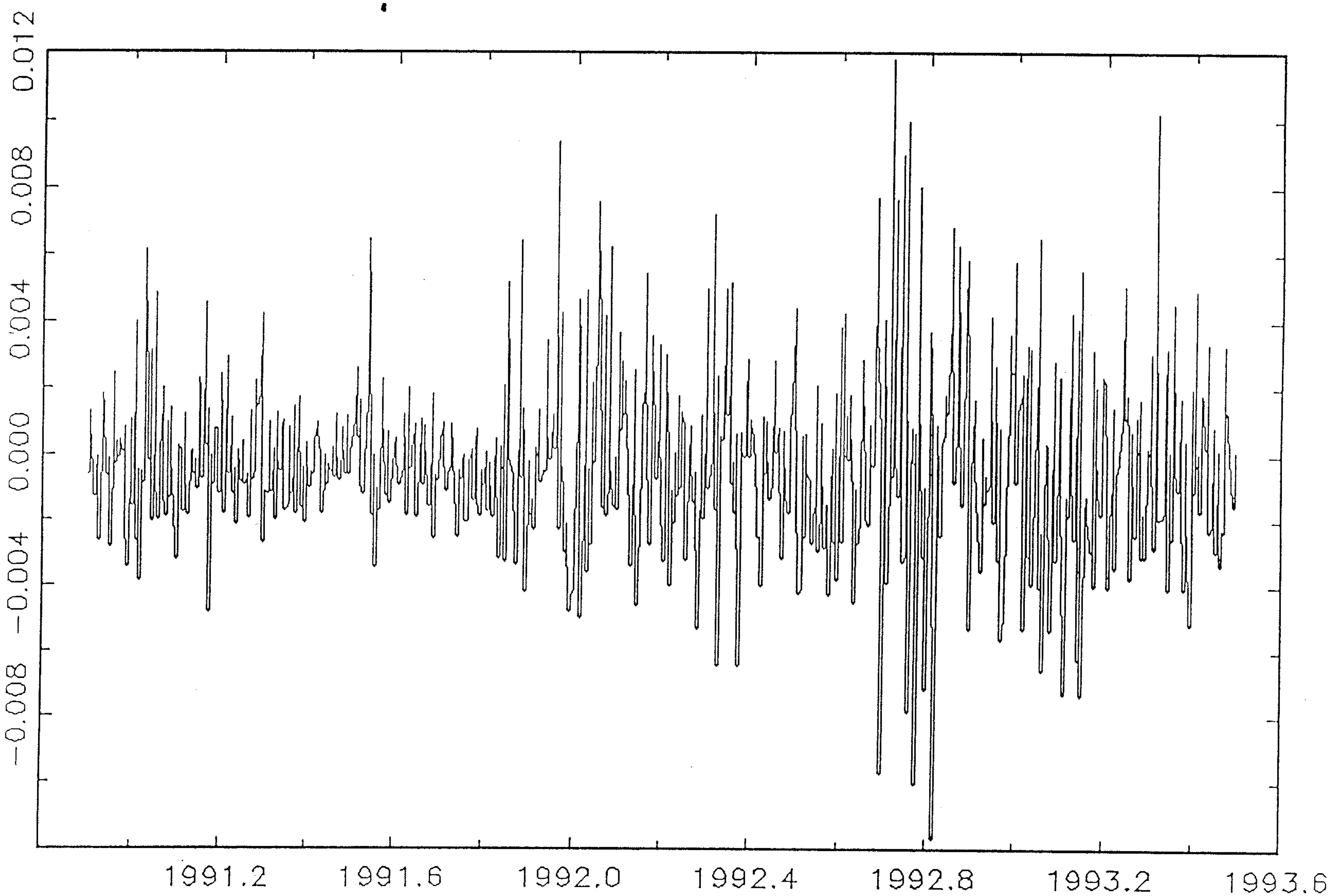


FIGURE 4: DENSITY ESTIMATE: SPOT EQUATION RESIDUALS

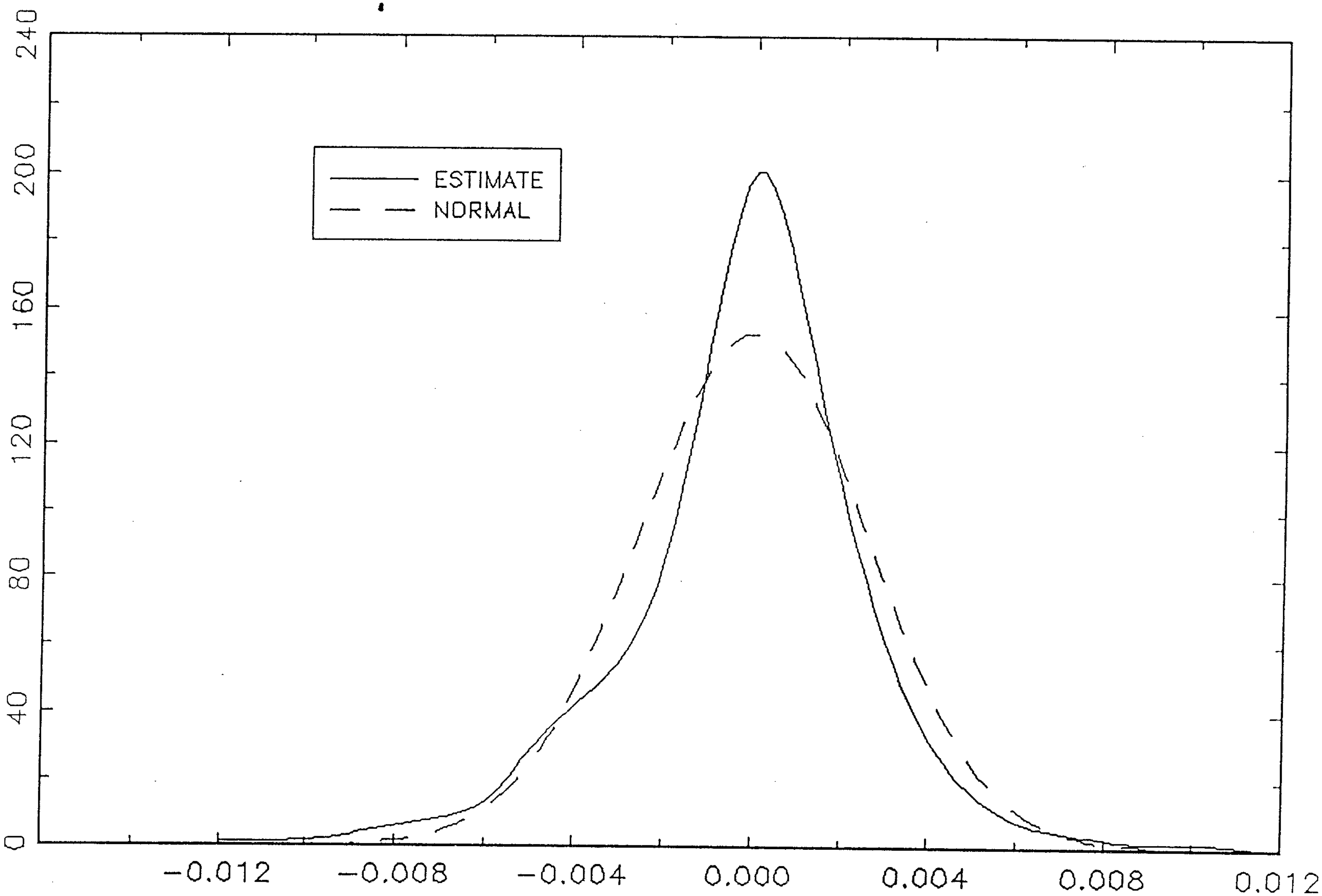
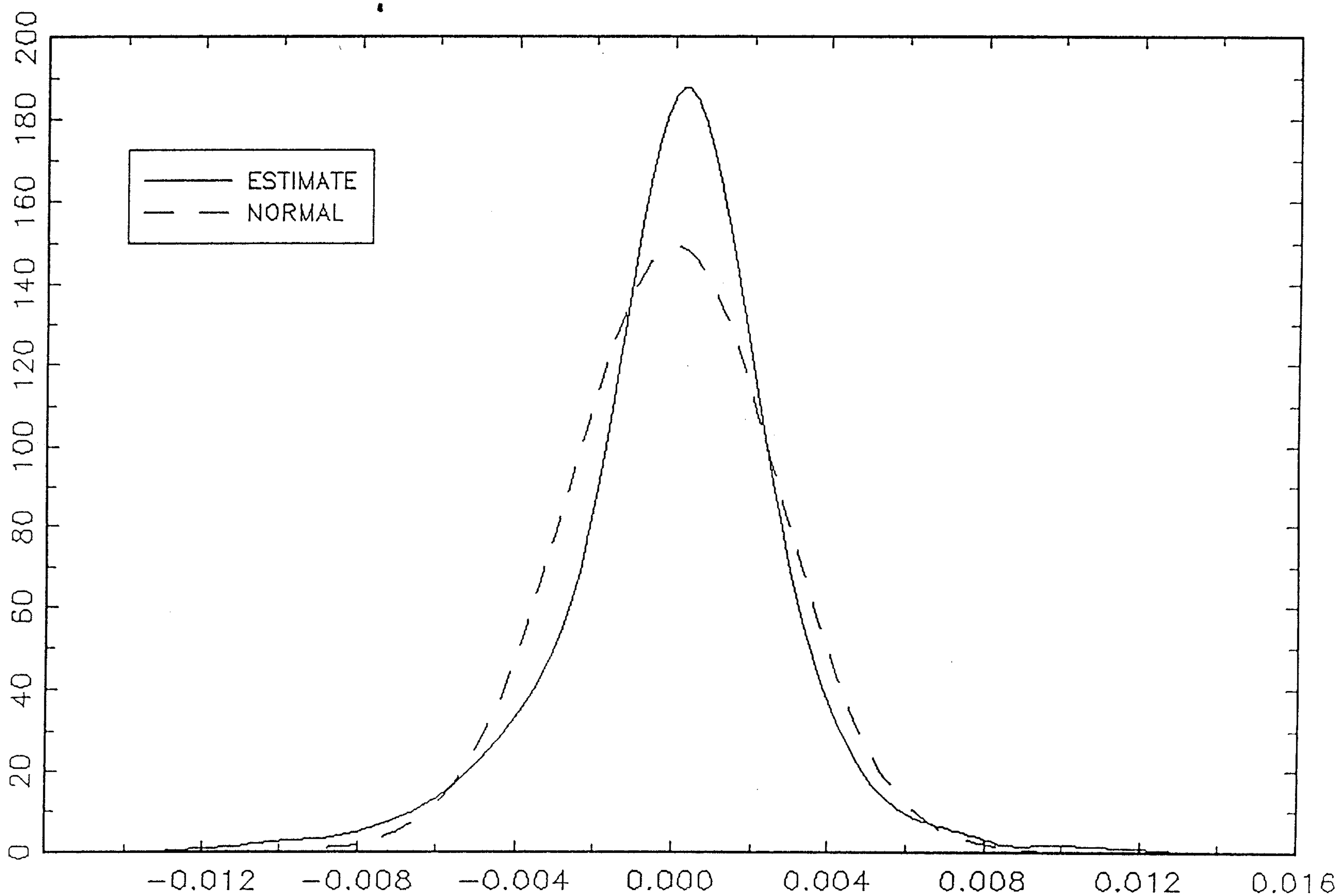


FIGURE 5: DENSITY ESTIMATE: FORWARD EQUATION RESIDUALS



REFERENCES

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

Baillie, R.T. and Bollerslev, T. 1989. Common stochastic trends in a system of exchange rates. *Journal of Finance* 44:167-181.

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

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

Balke, N.S. and Fomby, T.B. 1994. Large shocks, small shocks, and economic fluctuations: outliers in macroeconomic time series. *Journal of Applied Econometrics* 9:181-200.

Barnhart, S.W. and Szakmary, A.C. 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.

Bera, A.K. and Higgins, M.L. 1993. ARCH models: properties, estimation, and testing. *Journal of Economic Surveys* 7:305-362.

Beran, R. 1974. Asymptotically efficient rank estimates in location models. *Annals of Statistics* 2:63-74.

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

Blanchard, O.J. and Watson, M.W. 1986. Are business cycles all alike? In Gordon, R.J. (ed.), *The American Business Cycle: Continuity and Change*, Chicago, University of Chicago Press.

Bollerslev, T., Chou, R.Y., and Kroner, K.F. 1992. ARCH modeling in finance: a review of the theory and empirical evidence. *Journal of Econometrics* 52:5-59.

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

Cootner, P. (ed.) 1964. *The Random Character of Stock Market Prices*. Cambridge, Mass.; MIT Press.

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

Engle, R.F. and Granger, C.W.J. 1987. Co-integration and error correction: representation, estimation, and testing. *Econometrica* 55:251-276.

Fama, E.F. 1963. Mandelbrot and the stable Paretian hypothesis. *Journal of Business* 36:420-429.

Fama, E.F. 1965. The behaviour of stock market prices. *Journal of Business* 38:34-105.

Fama, E.F. 1970. Efficient capital markets: a review of theory and empirical work. *Journal of Finance* 25:383-417.

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

Froot, K.A. and Frankel, J.A. 1989. Forward discount bias: is it an exchange rate premium? *Quarterly Journal of Economics* 104:139-161.

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

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

Hodgson, D.J. 1995a. Adaptive estimation of cointegrated models. Ph.D. Thesis, Dept. of Economics, Yale University.

Hodgson, D.J. 1995b. Adaptive estimation of cointegrating regressions with ARMA errors. Chapter 1 of Hodgson (1995a).

Hodgson, D.J. 1995c. Adaptive estimation of error correction models. Chapter 2 of Hodgson (1995a).

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

Jarque, C.M. and Bera, A.K. 1980. Efficient tests for normality, homoscedasticity and serial independence of regression residuals. *Economics Letters* 6:255-259.

Jeganathan, P. 1994. Some aspects of asymptotic theory with applications to time series models. Forthcoming, *Econometric Theory*.

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

Johansen, S. and Juselius, K. 1990. Maximum likelihood estimation and inference on cointegration - with applications to the demand for money. *Oxford Bulletin of Economics and Statistics* 52:169-210.

King, R.G., Plosser, C.I., Stock, J.H., and Watson, M.W. 1991. Stochastic trends and economic fluctuations. *American Economic Review* 81:819-840.

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

Linton, O. 1993. Adaptive estimation in ARCH models. *Econometric Theory* 9:539-569.

Mandelbrot, B. 1963. The variation of certain speculative prices. *Journal of Business* 36:394-419.

Mandelbrot, B. 1966. Forecasts of future prices, unbiased markets, and martingale models. *Journal of Business* 39:242-255.

Manski, C.F. 1984. Adaptive estimation of non-linear regression models. *Econometric Reviews* 3:145-210 (including commentary).

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.

Nelson, C.R. and Plosser, C.I. 1982. Trends and random walks in macroeconomic time series: some evidence and implications. *Journal of Monetary Economics* 10:139-162.

Park, J.Y. 1992. Canonical cointegrating regressions. *Econometrica* 60:119-143.

Phillips, P.C.B. 1987. Time series regression with a unit root. *Econometrica* 55:277-301.

Phillips, P.C.B. 1993a. Forward exchange market unbiasedness: the case of the Australian dollar since 1984. Cowles Foundation Discussion Paper No. 1055, Yale University, August.

Phillips, P.C.B. 1993b. Robust nonstationary regression. Mimeo, Yale University, October.

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. and Durlauf, S.N. 1986. Multiple time series regression with integrated processes. *Review of Economic Studies* 53:473-495.

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

Phillips, P.C.B and Loretan, M. 1991. Estimating long-run economic equilibria. *Review of Economic Studies* 58:407-436.

Saikkonen, P. 1991. Asymptotically efficient estimation of cointegrating regressions. *Econometric Theory* 7:1-21.

Samuelson, P. 1965. Proof that properly anticipated prices fluctuate randomly. *Industrial Management Review* 6:41-49.

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

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. *Proc. Third Berkeley Symp. Math. Statist. Prob.* 1:187-196.

Stock, J.H. 1987. Asymptotic properties of least squares estimators of cointegrating vectors. *Econometrica* 55:1035-1056.

Stock, J.H. and Watson, M.W. 1993. A simple estimator of cointegrating vectors in higher order cointegrated systems. *Econometrica* 61:783-820.

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