Estimating Linear Quadratic Models With Integrated Processes

Gregory, Allan W., Adrian R. Pagan and Gregor W. Smith

Working Paper No. 247
September 1990

Estimating Linear Quadratic Models with Integrated Processes

Allan Gregory, Adrian Pagan and Gregor Smith

Rochester Center for Economic Research
Working Paper No. 247

ESTIMATING LINEAR QUADRATIC MODELS WITH INTEGRATED PROCESSES[1]

Allan W. Gregory,* Adrian R. Pagan** and Gregor W. Smith*

*Department of Economics, Queen's University
**University of Rochester and
Adjunct Professor, Australian National University

Working Paper No. 247

September 1990

Abstract

The paper studies estimation of parameters in the linear–quadratic optimmization model when the forcing variables are integrated. It is shown that it is difficult to estimate one of the parameters of this model, the discount rate, when integration pertains, and there can even be a complete failure of identification. A variety of estimators is proposed that would enable inferences with standard chi–square test statistics. These estimators differ in the amount of a priori knowledge about the linear quadratic model exploited in their construction. A small Monte Carlo study assesses the sampling properties of the estimators and reveals some difficulties in their use.

[1]Financial support from the Social Sciences and Humanities Research Council of Canada is gratefully acknowledged. The research was supported by the National Science Foundation under Grant No. NSF SES–8719520. We are grateful to Neil Ericsson, Bruce Hansen, Hashem Pesaran and anonymous referees for discussion on this material.

> But unlike the economic theoretician, who usually works with general classes of functions, [the builder of econometric models] must work with particular functional forms (e.g., linear, quadratic, exponential.)
>
> – Bergstrom (1967, p. 3)

1. Introduction

When Rex Bergstrom wrote his 1967 monograph, ad–hoc specification dominated in econometric research. Nowhere was this truer than in dynamics. The intervening twenty years has seen some progress in rectifying this situation: inter–temporal optimizing theory has been utilized to suggest suitable dynamic specifications and a literature has emerged describing classes of models that seem appropriate for many economic data series. The former development tends to be described as the "Euler equation approach," while the latter deals with the class of error correction models (ECM). At a very general level it is hard to reconcile the two traditions, but if one follows Bergstrom's pragmatism in adopting specific functional forms for agents' utility and production functions, it is possible to achieve a synthesis. Nickell (1985) did just this, showing that Euler equations derived from quadratic objective functions could be re–cast as error correction models, so that the approaches are isomorphic. Each has its advantages. The Euler equation approach enables interpretation of the estimated parameters as those associated with objective functions, whilst the ECM methodology allows the "data to speak for itself", and is therefore compatible with a number of possible objective functions.

What was most striking about the trends cited above was the implication that the nature of the forcing variables faced by agents determined the appropriate ***specification*** for estimating equations. Concurrently with the emergence of this principle, work by Engle and Granger (1987), Phillips (1987), Sims, Stock and Watson (1990) ***inter alia***, showed that ***estimation theory and procedures*** must be adapted for the type of data econometricians use. Consequently, it is an obvious step to integrate the themes. It is this quest which motivates the current paper. Surprisingly, there is little research on the interface, the most prominent exception being Dolado, Galbraith and Banerjee (1989), although one can find reference to some of the issues scattered through applied papers such as Ilmakunnas (1989) and the book by Pesaran (1987).

The paper is organized as follows. Section 2 sets out the models studied and describes different representations of them that are found in the literature. Section 3 looks at some general issues of estimability and identifiability of parameters in these models. It is shown that the ability to identify parameters is dependent upon the nature of the forcing variables, and that it is quite likely that one of the parameters, the discount factor, cannot be estimated with much accuracy. Hence, the common practice of pre–setting this parameter has theoretical support. Broadly, section 3 is concerned with the ability to find consistent estimators of the unknown parameters, and ignores questions of inference. Section 4 rectifies this by applying ideas in the literature concerned with the estimation of co–integrating vectors, for example Phillips and Hansen (1990), Johansen (1988) and Park (1988), that describe estimators which enable hypothesis tests to be conducted in a standard way on the co–integrating vector. One interesting difference between our situation and that in the literature on estimation with integrated variables is the way in which we arrive at an ECM representation. In Engle and Granger and Phillips and Hansen, the ECM is derived as a ***consequence*** of co–integration between variables, and it is the co–integrating vector which is of prime importance. All other parameters are regarded as nuisance parameters and are effectively ignored by the use of robust estimation techniques. But, when an ECM is a consequence of a theoretical specification, as in the models studied in this paper, the "nuisance" parameters become of interest, and some modifications to standard methods of estimation need to be made to allow for this fact. Finally, section 5 evaluates the proposed estimators by a small simulation study.

2. The Model and Its Representations

As mentioned earlier, inter–temporal optimizing models vary according to the choice of function to be optimized, but a popular variant has been the linear–quadratic adjustment cost model in which a decision maker solves the following problem in an attempt to track a target value y_s^* over time

$$\min_{\{y_s\}} E_t \sum_{s=t}^{\infty} \beta^{s-t} [\delta(y_s - y_s^*)^2 + (y_s - y_{s-1})^2] , \tag{1}$$

where the expectation is taken with respect to information available to an agent at time t ($\mathcal{F}_t$).[2] The target variable is $y_t^* = x_t'\theta + e_t$, where e_t is a white noise error known to the decision maker, while x_t is a qx1 vector of forcing variables.[3] It is presumed that the error e_t appears in $\mathcal{F}_t = \{e_t, y_{t-j}, x_{t-j+1}\}_{j=1}^{\infty}$, but is not known to an investigating econometrician, whose information set is therefore a subset of $\mathcal{F}_t$, $\mathcal{G}_t$. The popularity of this model is evident in the literature concerning choice of factor of production levels, for example Kennan (1979), Nickell (1987), Layard and Nickell (1986), and Wren–Lewis (1986), for which y_t would be the level of employment (or hours worked) and x_t variables such as output and the real wage.

The first order condition for the optimization problem is the Euler equation

$$\Delta y_t = \beta E_t \Delta y_{t+1} + c(y_t - y_t^*) , \tag{2}$$

where $c = -\delta$ and E_t is the expectation taken with respect to $\mathcal{F}_t$. As is well known, both roots of the characteristic equation of this second–order difference equation are positive and they lie on either side of one. Denote the stable root of the quadratic $\beta z^2 - (1 + \beta + \delta)z + 1 = 0$ by λ, leading to the forward solution to (2) of

$$y_t = \lambda y_{t-1} + (1 - \lambda)(1 - \beta\lambda) E_t \sum_{s=t}^{\infty} (\beta\lambda)^{s-t} y_s^* . \tag{3}$$

Equations (2) and (3) form the basis of estimation techniques advanced for the linear/quadratic model. As with all rational expectations models (see Wickens (1982)) there are two general approaches to estimation. In the first, $E_t\Delta y_{t+1}$ is replaced by $\Delta y_{t+1} + \eta_{t+1}$, where $E_t(\eta_{t+1}) = 0$, and equation (2) could be written as

$$\Delta y_t = \beta \Delta y_{t+1} + c(y_t - x_t'\theta) + v_t , \tag{4}$$

where $v_t = \beta\eta_{t+1} - ce_t$ has the property that $E(v_t | \mathcal{G}_t) = 0$. As suggested by McCallum (1976), instrumental variable (IV) estimators of the coefficients of equation (4) might be performed with instruments drawn from $\mathcal{G}_t$.

Alternatively, if an assumption is made about how x_t is generated, the unknown expectation in (3) may be found and an estimating equation derived. There are two cases of particular interest to us about how x_t evolves, and the corresponding solved versions of (3) would be[4]

<u>Case I</u>: $(1 - \rho_1 L)x_t = \epsilon_t \qquad E_{t-1}(\epsilon_t) = 0$

$$\Delta y_t = (\lambda - 1)(y_{t-1} - x_{t-1}\theta) + (1 - \lambda)\theta[(1 - \rho_1\beta\lambda)^{-1}(1 - \beta\lambda)x_t - x_{t-1}] + (1 - \beta\lambda)(1 - \lambda)e_t \tag{5}$$

<u>Case II</u>: $(1 - \rho_2 L)(1 - L)x_t = \epsilon_t \qquad E_{t-1}(\epsilon_t) = 0$

$$\Delta y_t = (\lambda - 1)(y_{t-1} - x_{t-1}\theta) + (1 - \beta\lambda\rho_2)^{-1}(1 - \lambda)\Delta x_t'\theta + (1 - \beta\lambda)(1 - \lambda)e_t . \tag{6}$$

If x_t is integrated of order one (I(1)) we would get (5) with $\rho_1 = 1$, while if it is of order two (I(2)), we would get (6) with $\rho_2 = 1$. Because it is central to later work it is useful to re-write the special case of (5) when $\rho_1=1$ as (5′).

$$\Delta y_t = (\lambda-1)(y_{t-1}-x_{t-1}\theta) + (1-\lambda)\Delta x_t\theta + (1-\beta\lambda)(1-\lambda)e_t, \tag{5′}$$

or in the equivalent form of the partial adjustment model

$$y_t = \lambda y_{t-1} + (1 - \lambda)x_t\theta + (1 - \beta\lambda)(1 - \lambda)e_t \ . \qquad (7)$$

In turn (7) is usefully re–parameterized in the format of Bewley (1979)

$$y_t = -(\lambda/(1 - \lambda))\Delta y_t + x_t\theta + (1 - \beta\lambda)e_t \ . \qquad (8)$$

(8) is important since it demonstrates that θ is the co–integrating vector between y_t and x_t as all other variables in (8) are I(0). Depending on the context, any one of (5′), (7) or (8) will be adopted when working with Case I.

3. Some General Issues of Estimation and Identification

Estimation procedures for these models vary according to whether they are "single equation" or "systems of equations" oriented. In the former instance attention is focussed upon either the Euler equation (4) or the "solved" forms in (5) and (6). In the latter, if $|\rho_1| < 1$ it is apparent that ρ_1 enters both the equation for y_t and that for x_t, and this points to the likelihood that joint estimation of the two equations would be profitable. Of the "single equation" proponents, two proposals stand out : those by Kennan (1979) and Dolado, Banerjee and Galbraith (1989).

Kennan (1979) was essentially concerned with Case I ($|\rho_1| < 1$), but where x_t followed a p'th order autoregression and was strictly exogenous. With a higher order autoregression of order p, (5) would contain terms $x_t, \ldots, x_{t-p+1}$. He proposed regressing y_t against $y_{t-1}, x_t, \ldots, x_{t-p+1}$ to obtain an estimator of λ, solving for $\hat{\delta}$ from the polynomial connecting (λ,β,δ) after prespecifying β, and then regressing $\delta^{-1}(\Delta y_t - \beta\Delta y_{t+1}) + y_t$ against x_t (see (4)) to find $\hat{\theta}$. Of course if one knew that p = 1, both θ and λ could have been estimated directly from (5), replacing ρ_1 by $\hat{\rho}_1$, the OLS estimator of ρ_1 from $x_t = \rho_1 x_{t-1} + \epsilon_t$.

Kennan observes that some corrections need to be made to standard errors owing to the use of "generated regressors", see Newey (1984) and Pagan (1984), in finding $\hat{\theta}$. A further difficulty arises when $\rho_1 = 1$, caused by the fact that the regressors y_{t-1} and x_t are co-integrated. This feature means that the regressor cross product matrix in his first stage regression to find λ, namely that associated with (7), is asymptotically singular, and that fact precludes the estimators of both parameters from attaining the T–convergence rate normally associated with integrated regressors. Park and Phillips (1988) deal with this complication by a "co–ordinate rotation". In general, if z_{1t} and z_{2t} are cointegrated with co–integrating vector α, so that $z_{1t} = z_{2t}\alpha + \zeta_t$, the relation $z_{1t}\gamma_1 + z_{2t}\gamma_2$ can be re–expressed as $(\gamma_1\alpha + \gamma_2)z_{2t} + \zeta_t\gamma_1$, where the two "new" regressors, z_{2t} and ζ_t, are respectively I(1) and I(0). Then γ_1 would be estimated $T^{1/2}$ consistently whereas $(\gamma_1\alpha + \gamma_2)$ is estimated T–consistently. Equating $z_{1t} = y_{t-1}$ and $z_{2t} = x_t$, inspection of (7) and (8) shows that $\gamma_1 = \lambda$, $\alpha = \theta$ and $\gamma_2 = (1-\lambda)\theta$. "Co–ordinate rotation" therefore gives the correct asymptotics directly and means the estimation of (8) rather than (7), and, if one wanted to proceed with Kennan's approach, it would seem best to work with the latter equation to determine θ and λ.

Retaining the restriction that $\rho_1 = 1$, Dolado, Galbraith and Banerjee (1989) note that the ordering in Kennan's procedure might be reversed. Under these circumstances y_t is I(1), the long–run response of y_t to x_t in (5) is θ, and the regression of y_t against x_t consistently estimates this long–run response. Having found θ they then suggest that one might consistently estimate δ and β by applying instrumental variables (IV) to (4) with instruments selected from $\mathcal{G}_t$.

The sequential estimation strategy espoused by Dolado et al., which is rooted in the two–step approach in Engle and Granger (1987), and which seeks to exploit different convergence rates for estimators of parameters of variables exhibiting different degrees of integration, is an interesting one. However, it is clearly important to know what its limits would be. One concern that is immediate with all of these models is whether the parameters can be identified. Identification issues in the linear quadratic model are rarely treated explicitly, although a parameter that seems to have been difficult to estimate has been the discount factor β. Some authors set it to a pre–specified number, for example Kennan (1979). Others do this but suggest it is estimable, for example, Blanchard and Melino (1986, p. 389) who comment, "We could in principle estimate β. Recent papers indicate that obtaining accurate estimates of β in models such as ours is difficult. Our choice of $\beta = .99$ is arbitrary. Varying β between .95 and 1.00 has however very little effect on the estimated parameters." Only a few actually estimate it, recent examples being Ilmakunnas (1989) and Dolado et al. (1989). Such a diverse set of responses does raise the possibility that there may be a problem of identifiability with this parameter, and we therefore proceed to examine this question, firstly in the context of (5), and subsequently treating estimation in the more general environment of (6).

That there is a serious identification problem for some of the parameters of (5) should be immediately apparent from (5), when $\rho_1 = 1$ and θ is a scalar, as there are three unknown parameters β, δ and θ, but only two regressors (see (7)). This seems to explain why those authors who pre–set β do so, as they invariably work with the solved equation (5). Those who seem to believe that β is estimable e.g Dolado et al, proceed instead from (4), with the additional assumption that θ can be consistently estimated from some other source. Indeed, θ ***can*** be consistently estimated from the regression of y_t against x_t, provided x_t is I(1) (Stock (1987)). However, since equation (4) uses less information than (5), it is hard to see how it is therefore possible to estimate more parameters. Some reconciliation of these views is needed, and we turn to that task now.

Dolado et al. (1989) have as a maintained assumption that x_t is strictly exogenous in both Cases I and II, i.e., the correlation between ϵ_t and e_t is zero. ***Assuming that θ is known*** they would recommend estimation of δ and β in Case I by performing IV on (4) with instruments $z_{t-1} = y_{t-1} - x_{t-1}\theta$ and Δy_{t-1} for z_t and Δy_{t+1} respectively.[5] For this scenario we now show that only one of the two parameters β and δ in (4) is in fact identifiable.

As all the variables in (4) are I(0) a requirement for asymptotic identification of the parameters of (4) is that the "relevance" condition be satisfied, namely that the covariance matrix between instruments and regressors is non-singular.[6] When $\rho_1 = 1$, $z_t = y_t - x_t\theta = (1 - \lambda L)^{-1}[(1 - \lambda)(1 - \beta\lambda)e_t - \theta\lambda\epsilon_t]$, and, under standard assumptions, the covariance matrix converges to

$$\begin{bmatrix} E(\Delta y_{t-1}\Delta y_{t+1}) & E(\Delta y_{t-1}z_t) \\ E(z_{t-1}\Delta y_{t+1}) & E(z_{t-1}z_t) \end{bmatrix}$$

$$= \begin{bmatrix} (\lambda-1)^2\gamma_2 - \theta(1-\lambda)^2 a + (\lambda-1)d & (\lambda-1)\gamma_2 + \theta(1-\lambda)a + d \\ (\lambda-1)\gamma_1 & \gamma_1 \end{bmatrix}, \qquad (9)$$

where $a = E(\Delta x_{t-1}z_t)$, $\gamma_k = E(z_t z_{t-k})$ and $d = (1-\beta L)^2(1-\lambda)^2\lambda\sigma_e^2$. Clearly this matrix is singular. Of course one can still estimate unidentified models, although estimators cannot be consistent and, as Phillips (1989) shows, the distribution of <u>both</u> of the IV estimators of β and δ will be rendered non-standard.[7]

It is interesting to speculate about the identifiability of β when $|\rho_1|<1$. ***If θ is known*** there are two regressors and two parameters in (5), suggesting that β might be identified. Pesaran (1987, p. 134–147) deals with identification issues in Euler equations such as (2), although some modification is needed to it to fit the variant that he conducts an identification investigation on. Specifically, he excludes e_t from the conditional expectation in (2). Consequently, replacing $y_t - y_t^*$ by $y_t - x_t'\theta - e_t = z_t - e_t$, his version of (2) would be $\Delta y_t = \beta E_t(\Delta y_{t+1}) + cz_t - ce_t$. After allowing

z_t to be a stationary AR(r) process, he gives a necessary condition for identification as $r \geq 2$, suggesting that when z_t is only a first order autoregressive process there will be identification difficulties. In fact, when $\rho_1 = 1$, $z_t = (1 - \lambda L)^{-1}[(1 - \lambda)(1 - \beta\lambda)e_t - \theta\lambda\epsilon_t]$, and therefore it is an AR(1), which agrees with our finding. Such an outcome should be contrasted with what happens when $\rho_1 \neq 1$. Then, $(1 - \lambda L)(1 - \rho_1 L)z_t = \{(1 - \lambda)(1 - \beta\lambda)(1 - \rho_1 L)e_t + [(\frac{(1 - \beta\lambda)(1 - \lambda)}{1 - \rho_1\lambda\beta} - 1) + \lambda L]\epsilon_t\theta\}$, making z_t an ARMA(2,1) process. Applying Pesaran's rule, the necessary condition for β and δ to be identified is therefore satisfied. The reduction from an ARMA(2,1), when $\rho_1 \neq 1$, to an AR(1), when $\rho_1 = 1$, occurs because the MA(1) in the latter case becomes $-\lambda\Delta\epsilon_t\theta + (1 - \lambda)(1 - \beta\lambda)\Delta e_t$, and the common unit root to both the AR(2) and the MA(1) cancels. In summary, once $\rho_1 = 1$ it is no longer possible to separately identify β and δ.[8]

Modification to these conclusions is needed if x_t is allowed to evolve as in Case II. Suppose that θ and ρ_2 are known (ρ_2 can be consistently estimated by regressing Δx_t against Δx_{t-1}), and focus on the simple IV estimator discussed previously. The forcing variable becomes $z_t = (1-\lambda L)^{-1}[(1-\lambda)(1-\beta\lambda)e_t - \theta\lambda(1-\rho_2\beta\lambda)^{-1}(1-\rho_2\beta)\Delta x_t]$, and the fact that Δx_t is an AR(1) means that z_t will be an ARMA(2,1), satisfying the necessary condition for identification. It seems likely that this will be sufficient as well, although it is hard to verify the conjecture. Perhaps a more relevent concern is not whether there is a complete failure of identifiability, but whether the estimation problem is sufficiently well determined to enable precise estimation of β and δ even in very large samples. One way to address this issue is to conceive of the demonstrated failure of identifiability when $\rho_2=0$ (Case II coincides with Case I) as arising from perfect collinearity between the instruments, thereby leading to a singular correlation matrix. Then it may be asked how far away from singularity the correlation matrix between instruments and regressors becomes as ρ_2 increases. There seems little to be said analytically about this feature, as the covariance matrix is a complex function of β, δ, ρ_2, σ_e^2 and σ_ϵ^2. For this reason we have simply computed the eigenvalues of the correlation matrix for many configurations of Case II, finding that the smallest eigenvalue was always less than .1. For example, in the Monte Carlo "basic experiment" described in equation (20), and with $(1-\rho_2 L)\Delta x_t = \epsilon_t$, the smallest eigenvalues are .01(ρ_2=.5), .05 (ρ_2=.8) and .07 (ρ_2=.9).[9] This finding points to the extreme difficulty of ***jointly*** estimating both β and δ. To see that the source of the collinearity is due to β, assume that β is known, form $\beta\Delta y_{t+1}$, and move it to the LHS of (4), after which δ may be estimated by IV with z_{t-1} as an instrument for z_t. Then the corresponding eigenvalues (now the correlation coefficients between the two variables) would be .5, .49 and .48 respectively.

A different view of the estimability of β is to be had from the ECM (6). Again fixing θ and ρ_2, estimation would proceed by linearizing (6) around some initial estimators of λ and β, performing a regression to get updated estimates of these parameters, and continuing to iterate this sequence until convergence. At the termination of iterations $\hat{\delta}$ is recoverable by solving the polynomial $\hat{\beta}\hat{\lambda}^2 - (1 + \hat{\beta} + \hat{\delta})\hat{\lambda} + 1 = 0$. The precision of estimation depends upon the covariance matrix of the derivatives of (6) with respect to λ and β, and these derivatives are $[-(1-\rho_2\beta\lambda)^{-1}(\rho_2\beta-1)\Delta x_t'\theta + z_{t-1}]$ and $-(1-\rho_2\beta\lambda)^{-2}\rho_2\lambda(1-\lambda)\Delta x_t'\theta$ respectively, where $z_t = (1 - \lambda L)^{-1}[(1 - \lambda)(1 - \beta\lambda)e_t - \theta\lambda(1 - \rho_2\beta\lambda)^{-1}(1 - \rho_2\beta)\Delta x_t]$. Correlation between these terms is high for values of ρ_2 near zero but weakens quickly as ρ_2 increases. For the example mentioned above, which led to a near singular matrix in the IV case, the correlation between the derivatives is −.8 when $\rho_2 = .1$ but −.28 when $\rho_2 = .5$. This feature shows that the ECM has performed a successful re–parameterization so as to ensure a low correlation between regressors, whereas the Euler equation variant (4) does not do this. Despite that fact, the ability to estimate β accurately depends not just upon the correlation between derivatives but also their variances; in particular we are interested in the magnitude of the variance of the derivative of (6) with respect to β. Setting ρ_2=0 (Case 1) makes this variance identically zero, and so β is not identified, but it is equally apparent that the variance will tend to be small unless ρ_2 is quite large. Moreover, this variance will tend to be much smaller than that of the derivative with respect to λ unless ρ_2 is large, illustrating the point that it is β which is generally hard to estimate. In the basic experiment being referred to the ratio of the variance of the β–derivative to that for the λ–derivative is forty when ρ_2=.5 and thirteen when ρ_2=.8.

As ρ_2 approaches unity, both derivatives become I(1) processes, creating the potential for sharp estimation. Some care has to be exercised with the argument however. If β=1, it can be seen from the formula for z_t that this variable would be I(0), and so both derivatives are dominated by terms in $\Delta x_t \theta$ i.e. they would be co–integrated, meaning that only one of the parameters could be estimated T–consistently. Interestingly enough, if one reverts to estimating from the Euler equation, it would be β that can be estimated T–consistently, as it attaches to an I(1) variable Δy_{t+1}, whereas δ is associated with the I(0) variable z_t.[10] It is hard to know what to make of this case. Theoretically β could not be unity, although the fact that it is generally in the range .96–.99 might indicate that the limiting case might be of interest as an approximation. Moreover, it also seems unlikely that most x_t variables would be I(2), at least after a log transformation. Nevertheless one should keep this extreme configuration in mind as a counterexample about the ability to estimate β , even if one's presumption is that the difficulties in estimating β documented above point in the direction of pre–setting it. It is hard to escape the feeling that this is exactly what happens in applied studies, with the discount rate being fixed after poor estimates are obtained in initial (or previous) investigation. Our analysis might therefore be seen as providing a pragmatic justification for fixing β in applied work with the linear/quadratic model.

Systems estimation has been outlined in Hansen and Sargent (1980) (when $|\rho_1|<1$), and has been extensively used in the literature e.g. Sargent (1978). If $\rho_1=1$, θ would be the co-integrating vector, and in principle any estimators of such a vector could be utilized, e.g., Johansen (1988), Phillips (1988b) and Park (1988), although there are other parameters to be estimated here, namely δ, and that may demand a modified response. For the objectives of this section however, it is sufficient to note that, as all the discussion given previously concerning identification of δ and β proceeded under the assumption that ρ_2 was known, it must apply equally to systems estimators.

4. Estimation of the Linear/Quadratic Model

In this section we consider the estimation of the unknown parameters in (4) and (5′), i.e., it is assumed that a unit root appears in the equation describing the evolution of x_t and that this fact is known to an investigator. As the analysis closely follows that for Case I mention of Case II will be made only when necessary. Because the discussion in the previous section pointed to severe difficulties in estimating β, it will be treated as known throughout this section.

4.1 Single Equation Estimation from the Euler Equation

With β assumed known (4) could be re-defined as

$$\psi_t = c(y_t - x_t'\theta) + v_t , \tag{10}$$

where $\psi_t = \Delta y_t - \beta\Delta y_{t+1}$ and $c = -\delta$. A simple estimator of the two unknown parameters θ and c is available from the logic of Engle and Granger's (1987) two-step method. First, θ is estimated by regressing y_t against x_t to produce $\tilde{\theta}$, and then c is estimated by doing instrumental variables of ψ_t against $y_t - x_t'\tilde{\theta}$ to produce $\tilde{c}$. The regression to get $\tilde{\theta}$ is essentially from (8), where the error term is $-(1 - \lambda)^{-1}\lambda\Delta y_t + (1 - \beta\lambda)e_t$. As $\Delta y_t = (\lambda - 1)z_{t-1} + (1 - \lambda)\Delta x_t\theta + (1 - \beta\lambda)(1 - \lambda)e_t$ from (5′), and z_t was shown in section 3 to be a stationary AR(1), it follows that $T(\tilde{\theta} - \theta_o)$ has a limiting distribution, provided that ϵ_t and e_t are restricted as in Phillips and Hansen (1990). It is then an easy matter to demonstrate that $T^{1/2}(\tilde{c} - c_o)$ will be asymptotically normally distributed, provided that this is so for the IV estimator using instruments based on the true θ.

Now it is well known that $\tilde{\theta}$ will not be asymptotically normally distributed when x_t is I(1) unless x_t is strictly exogenous, necessitating some adjustments to it in order that inferences about θ can be made from the t–statistics associated with the regression. There are a number of suggested adjustments, e.g., Park's (1988) "Canonical Cointegrating Regressions" method or the "fully modified estimator" of Phillips and Hansen (1990). As it is the latter which is adopted in this paper, a brief description of its *modus operandi* seems in order.

Consider equation (8) and the equation describing Δx_t, collected below as (11a) and (11b),

$$y_t = x_t'\theta + \xi_t \tag{11a}$$

$$\Delta x_t = \epsilon_t , \tag{11b}$$

where $\xi_t = -(1 - \lambda)^{-1}\lambda\Delta y_t + (1 - \beta\lambda)e_t$. Two corrections to $\tilde{\theta}$ are made by Phillips and Hansen. First, they correct a "bias" in $T(\tilde{\theta}-\theta)$ due to endogeneity of Δx_t. This involves replacing y_t by $y_t^+ = y_t - \hat{\Omega}_{12}\hat{\Omega}_{22}^{-1}\Delta x_t$, where $\hat{\Omega}$ is an estimator of Ω, the "long run covariance matrix of $\zeta_t' = [\xi_t \ \Delta x_t]$ ",[11] and regressing y_t^+ against x_t to produce $\theta^* = (\Sigma\, x_t x_t')^{-1}\, \Sigma\, x_t y_t^+$. A further correction is then made for "autocorrelation bias" by modifying θ^* to $\hat{\theta}_{PH} = \theta^* - (\Sigma\, x_t x_t')^{-1}\, Tg$, where $g = \hat{\Delta}$

$\begin{bmatrix} 1 \\ -\hat{\Omega}_{22}^{-1} & \hat{\Omega}_{21} \end{bmatrix}$ and $\hat{\Delta}$ is a consistent estimator of $\Delta = \sum_{k=0}^{\infty} E\,(\Delta x_0\ \zeta_k')$. After these modifications t–statistics formed with $\hat{\theta}_{PH}$ are asymptotically normally distributed.

Some comments on this estimator are in order, particularly since it forms the backbone of many of the other estimators that we study. First, it applies immediately to Case II, since (8) would just be augmented by an extra I(0) term $[(1 - \lambda\beta\rho_2)^{-1}(1 - \lambda) - 1]\Delta x_t\theta$, and by its very nature the estimator adapts to serial correlation in Δx_t. Second, the procedure works ***only*** if it is known that there is a unit root in the x_t equation and that root is **prescribed** rather than estimated — see Philips and Hansen (p. 103).

Now, if x_t had been an I(0) process, it would have been desirable to exploit the non–linearity in parameters that characterizes (10). To do so involves the linearization of (10) around some initial consistent estimators, say $\tilde{\theta}$ and $\tilde{c}$, to give

$$w_t \;=\; \tilde{z}_t c + (\tilde{\delta}x_t')\theta + \nu_t\ , \tag{12}$$

where $\tilde{z}_t = y_t - x_t'\tilde{\theta}$; $w_t \;=\; \psi_t - \tilde{c}x_t'\tilde{\theta}$ and $\nu_t \;=\; v_t - x_t'(\theta - \tilde{\theta})(c - \tilde{c})$. This expansion is exact as the second derivatives of the function with respect to θ and c are zero, and it is only the cross derivative which is non–zero. (12) suggests the possibility of producing another estimator of θ by regressing w_t against $\tilde{\delta}x_t$. A complication however is the presence of the I(0) term $\tilde{z}_t c$ on the RHS of (12), although an obvious solution, employed by Hansen (1989), is to purge (12) of that term by subtracting $\tilde{z}_t\tilde{c}$ from both sides, making the new dependent variable $w_t - \tilde{z}_t\tilde{c}$, and then to run a regression of $(w_t - \tilde{z}_t\tilde{c})$ against $(\tilde{\delta}x_t)$ to find an estimator of θ. Conceivably one might iterate this process.

Although the estimation strategy described above is a straightforward application of classical non–linear regression methods, its properties are not immediately obvious. For this reason an alternative derivation of (12), which aids understanding and is very useful, is to return to (4) and invert it, obtaining

$$y_t = x_t'\theta + \delta^{-1} (\beta\Delta y_{t+1} - \Delta y_t + v_t) . \tag{13}$$

If δ was known, θ might be estimated from the regression

$$y_t + \delta^{-1}\psi_t = x_t'\theta + \delta^{-1}v_t , \tag{14}$$

or

$$\delta y_t + \psi_t = \delta x_t'\theta + v_t . \tag{15}$$

Replacing δ by $\tilde{\delta}$ in (15) makes the dependent variable $\tilde{\delta}y_t + \psi_t = \psi_t + \tilde{\delta}x_t'\tilde{\theta} + \tilde{\delta}(y_t - x_t'\tilde{\theta}) = w_t - \tilde{z}_t\tilde{c}$, and (15) is therefore equivalent to (12), where the error term in the latter absorbs the terms created by the shift from δ to $\tilde{\delta}$. Hence, the new estimator of θ arising from linearization can be computed from the regression of $y_t + \tilde{\delta}^{-1}(\Delta y_t - \beta\Delta y_{t+1})$ against x_t.[12] Of course, just like $\tilde{\theta}$, some adjustments need to be made to enable standard inferential procedures to pertain, but that can be accommodated by applying the Phillips–Hansen technology to

$$y_t + \tilde{\delta}^{-1} (\Delta y_t - \beta\Delta y_{t+1}) = x_t'\theta + \nu_t \tag{16b}$$

$$\Delta x_t = \epsilon_t, \tag{16b}$$

rather than to (11a) and (11b).[13] The resulting estimator will be designated $\hat{\theta}_E$, with the E to represent the fact that it derives from the Euler equation. Notice that in Case I $y_t - x_t'\theta$ in (8) and (13) must be identical, so the difference between $\hat{\theta}_{PH}$ and $\hat{\theta}_E$ resides solely in the fact that $\delta\psi_t$ is purged from the RHS of (8) when constructing $\hat{\theta}_E$. ***Prima facie***, it might be expected that the resulting estimator of θ, $\hat{\theta}_E$, would have better small sample behavior than $\hat{\theta}_{PH}$, as it uses prior knowledge about the structure of the error term in (13), rather than requiring that the long–run covariance adjustments described earlier do all the work.

In the analysis above attention has centered upon how to estimate θ in such a way as to be able to utilize standard test statistics. Corresponding to each way of estimating θ would be alternative estimators of $z_t = y_t - x_t'\theta$, and these would generate different IV estimators of δ. In all cases however, the choice of instruments could be made according to the principles of GMM estimation set out in Hansen (1982), although because of the one–dependent nature of the error v_t, the optimal instruments involve all past lags of Δy_t and z_t — see Hansen and Singleton (1988). Thus, there is more than one way of setting up the IV estimator. In all instances however, the fact that θ can be estimated T–consistently means that $T(\hat{\delta}-\delta_0)$ will have a limiting normal distribution whose exact covariance matrix will depend on the type of IV estimator selected. Notice that the one–dependent nature of v_t makes it necessary to do a robust computation of the standard errors of $\hat{c}$ in order to make proper inferences.

4.2 Single Equation Estimation from the ECM

If the expectation is solved for, estimation can be done with the ECM's (5) or (6). Here, the information that x_t is known to possess a unit root has been exploited in settling on a specification, and so prior knowledge of a unit root is extremely important to this approach. Just as for the Euler equation discussion, β will be taken to be known so that the task is to estimate θ and λ in Case I ($\rho_1=1$); once λ has been quantified δ can be recovered by factorizing the polynomial $(1-\lambda z)(1-\lambda z^{-1}) = \beta z^2-(1+\beta+\delta)z + 1$. An initial consistent estimator of λ, $\tilde{\lambda}$, could be found by Kennan's method, i.e., regress y_t against y_{t-1} and x_t and use the estimated coefficient on y_{t-1}, but there would be other alternatives such as 2SLS on (8), with instruments y_{t-1} and x_t for Δy_t and x_t – see Bewley (1979)– or OLS on (5) with θ replaced by $\tilde{\theta}$. It is the last of these options that we select.

Linearizing (5) around $\tilde{\theta}$ and $\tilde{\lambda}$ gives

$$y_t - x_t'\tilde{\theta}\tilde{\lambda} = [\tilde{z}_{t-1} - \Delta x_t'\tilde{\theta}]\lambda + [(1-\tilde{\lambda})x_t']\theta + u_t \tag{17}$$

where $u_t = x_t'(\theta - \tilde{\theta})(\lambda - \tilde{\lambda}) + (1 - \beta\lambda)(1 - \lambda)e_t$. If x_t was I(0), one would perform the regression in (17) to get updated estimates of θ and λ and, if the u_t were normally distributed, this two-step estimator would be asymptotically efficient relative to the MLE. Because $z_{t-1} - \Delta x_t'\theta$ is I(0) in Case I($\rho_1=1$), rather than perform the full regression in (17), one might proceed with the following sequential estimator.[14] First, $[\tilde{z}_{t-1} - \Delta x_t'\tilde{\theta}]\lambda$ is moved to the LHS of (17) and θ is estimated. Second, with that estimate of θ replacing θ, move $(1-\tilde{\lambda})x_t'\theta$ to the LHS of (17) and determine λ by a regression.

To relate this way of estimating θ to those presented previously, observe that the regression equation it is based upon is

$$(1 - \tilde{\lambda})^{-1}[y_t - x_t'\tilde{\theta}\tilde{\lambda} - (\tilde{z}_{t-1} - \Delta x_t'\tilde{\theta})\tilde{\lambda}] = x_t'\theta + \eta_t. \qquad (18).$$

In (18) the LHS is $(1 - \tilde{\lambda})^{-1}[y_t - \tilde{\lambda}y_{t-1}] = y_t + (1 - \tilde{\lambda})^{-1}\tilde{\lambda}\Delta y_t$, and this would be identical to the LHS of (8) except that λ has been replaced by $\tilde{\lambda}$. Comparing (18) with (11a) reveals that the estimator of θ differs from $\tilde{\theta}$ in the fact that the disturbance term in (11a) has been purged of a known component. The factor extracted differs from that for $\hat{\theta}_E$ (see (16b)), because of the different informational assumptions. If the process generating x_t is exactly known, the error term in any (y_t, x_t) regression can be reduced to a function of the "econometrician error" alone. Without this knowledge, there is a remaining component that depends on the "expectation error" as well.

Now this connection means that θ can be estimated by applying the Phillips–Hansen estimator to (8) and (11b), after adding $(1-\tilde{\lambda})^{-1}\tilde{\lambda}\Delta y_t$ to both sides of (8). The resulting system is

$$y_t + (1-\tilde{\lambda})^{-1}\tilde{\lambda}\Delta y_t = x_t'\theta + \eta_t \qquad (19a)$$

$$\Delta x_t = \epsilon_t, \qquad (19b)$$

and the estimator of θ will be designated $\hat{\theta}_{ECM}$. To get $\hat{\lambda}_{ECM}$, the second part of the sequential procedure is implemented with $\tilde{z}_t$ being regressed against $[\tilde{z}_{t-1}-\Delta x_t'\tilde{\theta}]$ to produce $\hat{\lambda}_{ECM}$. Because of the T–consistency of $\hat{\theta}_{ECM}$ the asymptotic distribution of $\hat{\lambda}_{ECM}$ is the same as if θ was known. If Case II is entertained (8) would have an extra term $[(1-\lambda\beta\rho_2)^{-1}(1-\lambda)-1]\Delta x_t'\theta$ in it, which would change the linearized equation. Nevertheless, by re–arranging the terms into those which are I(1) and those which are I(0), it is simple to apply the same sequential approach to the estimation of θ and λ (although different formulae would emerge). There is an additional complication though, caused by the use of an initial consistent estimator of ρ_2. This induces a "generated regressor" effect into the regression determining $\hat{\lambda}_{ECM}$, and a correction to the covariance matrix needs to be done following the formulae in Newey (1984). Actually, because ρ_2 appears in both (8) and in the equation generating Δx_t, it might be sensible to adopt a systems estimator to exploit the efficiency gains coming from such cross equation restrictions.

4.3 Systems Estimation

The estimators described in sections 4.1 and 4.2 are single equation estimators in the sense that they estimate the parameters of (4) or (8), ignoring the equation for Δx_t, except insofar as information on Δx_t is used to perform "bias adjustments". As the parameters to be determined are essentially the co–integrating vector θ and the "adjustment parameter" δ, there may be some gains to jointly estimating the system. For this to be so Δx_t would have to possess an autoregressive structure, such as in Case II. Johansen (1988) describes a LIML–type joint estimator of θ and λ when $(y_t\ x_t)$ follow a finite order VAR, Phillips (1988b) discusses FIML estimation of the parameters θ, λ and ρ_2, and Phillips (1990) outlines a spectral estimator of θ and λ. Conceptually, it is hard to apply the first two of these procedures to the estimation of parameters of the Euler equation, as the vector $(y_t\ x_t)$ does not follow a finite order

VAR owing to the MA(1) in the composite error term of (4). Hence, the spectral estimator looks most appealing, as it does not need to explicitly estimate the parameters of the autocorrelation process of Δx_t.

A similar problem arises with the ECM version. Take Case II and replace Δx_t in (6) by $(1-\rho_2 L)^{-1}\epsilon_t$. Multiplying (6) by $(1-\rho_2 L)$ gives a VAR in $(y_t\ x_t)$ but with a disturbance term of the form $\epsilon_t + (1-\rho_2 L)(1-\beta\lambda)(1-\lambda)e_t$ for the y_t equation, i.e., an MA(1) again. Consequently, the assumption of Johansen's method that these errors are white noise will be invalid. Moreover, it is not possible to replace this composite error with a single error term and then to proceed to maximum likelihood estimation, see Phillips (1988b). These facts raise interesting problems about joint estimation of λ and θ by systems methods that deserve fuller exploration. Most work with systems approaches has begun with the presumption that there is a finite order VAR for $(y_t\ x_t)$ and then derived an ECM representation that is compatible with such a VAR. Utilizing theoretical analysis to derive the ECM reverses this sequence and it highlights the fact that there may be no finite order VAR which encapsulates the theoretical model. Hence, estimation methods that emphasise a VAR structure are not necessarily appropriate to estimation of the linear quadratic model.

5. A Monte Carlo Study of the Estimators

A brief investigation was made of the sampling properties of the proposed estimators, with special emphasis being placed upon the correspondence of the size of test statistics with the theoretical predictions of the asymptotic theory. Specifically, the prediction is that the re–centered estimators of θ should have t–values that behave like a standard normal deviate in large samples. To assess this prediction, one thousand replications of each experiment were performed, and the proportions of times the t–values exceeded the critical values associated with the 10%, 5% and 1% levels of significance of a standard normal are reported. These are taken to indicate the true sizes of the test. All computation was done with programs written in the GAUSS language.

The basic experiment consists of the two equations

$$\Delta y_t = (\lambda - 1)(y_{t-1} - x_{t-1}\theta) + (1 - \lambda)\theta\Delta x_t + (1 - \beta\lambda)(1 - \lambda)e_t \quad (20a)$$

$$\Delta x_t = \epsilon_t , \quad (20b)$$

where $[e_t \ \epsilon_t]$ are drawn from a bivariate normal distribution with covariance matrix $\Sigma = \begin{bmatrix} \sigma_{ee} & \sigma_{e\epsilon} \\ \sigma_{\epsilon e} & \sigma_{\epsilon\epsilon} \end{bmatrix}$ and the parameters θ, δ and Σ vary across experiments.[15] West (1986) has a similar experimental design. Our basic experiment puts $\theta = 1$, $\delta = .5$ and $\Sigma = \begin{bmatrix} 1 & .5 \\ .5 & 1 \end{bmatrix}$, meaning that the forcing variable x_t is actually an endogenous variable with zero drift. For ease of reference estimators are given a designation that refers to the procedure that generated them. Hence $\tilde{\theta}_{OLS}$ is the OLS estimator $\tilde{\theta}$ described from the regression in (11a); $\hat{\theta}_{PH}$ is based on this OLS estimator but now centered properly as suggested in Phillips and Hansen(1990); $\hat{\theta}_E$ is the Phillips–Hansen estimator of θ from (16) while $\hat{\delta}_E$ is the IV estimator from (4) with θ replaced by $\hat{\theta}_E$; and $\hat{\theta}_{ECM}$ and $\hat{\delta}_{ECM}$ are the Error Correction Model variants arising out of (19). For the IV estimator of δ the instruments employed were Δy_{t-1}, Δy_{t-2} and $\tilde{z}_{t-1}$. The estimator $\tilde{\theta}_{OLS}$ is included in order to illustrate the effects of ignoring the integration in x_t upon the sizes of test statistics.

Table 1 records the outcomes for the initial experiment and is intended to show how the sizes of the test statistics are related to variation in the cost of adjustment. Based on this initial experiment the ECM estimators work well with samples of 200 observations, and are acceptable even with T=100. As expected, the OLS estimator, being uncentered, has a large size distortion that can be reduced considerably by the Phillips–Hansen modifications. However, $\hat{\theta}_{PH}$ still has bias in its size relative to the

ECM–based estimator. What is striking, and surprising, is the poor performance of $\hat{\theta}_E$. To explain this phenomenon, recall that the equation defining this estimator is (16a), and in our experiments the error term of (16a) would be

$$v_t = \delta(e_t - \lambda e_{t+1}) - \theta(1-\lambda)\epsilon_{t+1} \; . \tag{21}$$

Equation (21) and (16b) comprise the "system" for the Phillips–Hansen procedure. A crucial part of the adjustment is to compute the ***long run*** covariance matrix of the disturbance terms in the two equations. It is easy to show that as δ becomes small this matrix becomes singular as the two errors will be approximately $-\theta(1-\lambda)\epsilon_{t+1}$ and ϵ_t. Accordingly, one of the maintained assumptions in the Phillips–Hansen theory, that the long run covariance matrix is non–singular, is violated.[16] This fact leads to a failure of the correction to "work" and in this instance means a bias towards the null as δ shrinks. By comparison the ECM estimator of θ is insensitive to δ as the error term in (19a) involves only e_t, and so the long run covariance matrix of (19a) and (19b) disturbances does not depend on δ. For $\hat{\theta}_{PH}$ the relevant errors are in (11); that in (11a) is quite complex being $(-\theta\lambda\epsilon_t + \delta\lambda e_t)(1 - \lambda L)^{-1}$. Since the long run covariance matrix is proportional to the spectral density at the origin, its determinant can be shown to be proportional to $(1-\lambda)^{-2}\delta^2\lambda^2$. Obviously, this also becomes singular as δ shrinks to zero, but the presence of $(1-\lambda)$ in the denominator rather than the numerator (as is true for $\hat{\theta}_E$) provides an offsetting factor, in that λ changes inversely with δ. Thus the determinant of the long run covariance matrix of the disturbances underlying each estimator is (.29, .28), (.19, .05), (.10, .004) and (.06, .0005) for δ=1,.5,.2,.1, where the first value in each bracket relates to $\hat{\theta}_{PH}$ and the second to $\hat{\theta}_E$. Clearly the singularity problem is much less marked for $\hat{\theta}_{PH}$. Even so, there is a tendency for it to over–reject. For small δ both $\hat{\delta}_E$ and $\hat{\delta}_{ECM}$ have accurate test sizes. However, as δ rises (the cost of adjustment falls), the rejection frequency rises dramatically, especially for $\hat{\delta}_E$, the reason being that the instruments become very poor as both Δy_t and z_t resemble white noise processes.

Table 2 reveals what happens as the degree of endogeneity of x_t increases. To achieve such an effect the covariance matrix Σ is altered to $\Sigma = \begin{bmatrix} 1 & .7 \\ .7 & 1 \end{bmatrix}$. All results appear to be insensitive to this change. Our final experiment in Table 3 examines the influence of the value of the long–run response, θ. Since the error term in the estimating equation for $\hat{\theta}_E$ is essentially $-\theta(1 - \lambda)\epsilon_{t+1}$, and that in the equation for Δx_t is ϵ_t, the long–run covariance matrix of these two quantities is singular. A similar problem occurs for $\hat{\theta}_{PH}$. For large θ the error term in the regression yielding it is effectively $(1 - \lambda L)^{-1}\theta(1 - \lambda)\epsilon_t$. Only the ECM estimator is immune to this singularity problem as the error term in its regression is the last term in (8), namely $(1 - \beta\lambda)e_t$. This explains the outcomes observed in Table 3 for the estimators of θ. What is odd in this table is the failure of the ECM estimator of δ. Given that θ is very accurately estimated by $\hat{\theta}_{ECM}$, $\hat{\lambda}$ is effectively the OLS estimate from (17) after fixing θ, and we would therefore expect good performance for $\hat{\delta}_{ECM}$. Exactly why this outcome is observed is a subject for future study.

6. Conclusion

The paper has examined the estimation of equations coming from linear quadratic optimisation models when the forcing variables are integrated. After demonstrating that there is frequently a lack of identification for some of the parameters in such models, we fixed one of them — the discount rate — and proceeded to look at how estimators of the remaining parameters could be derived. By the nature of Euler equation solutions there are two sets of parameters to be estimated; one associated with I(1) variables and the other with I(0) variables. Previous literature has largely concentrated upon the estimation of the parameters attached to I(1) variables, the co–integrating vector. Unfortunately, standard procedures for this have undesirable side effects upon the estimation of the I(0) variable parameters. Nevertheless, simple adjustments are implemented to overcome this problem.

We devise two general classes of estimators for the unknown parameters, corresponding to the two procedures normally used for estimating models with rational expectations. The first of these works with the Euler equation (E) and is found by replacing the unknown expectation with the observed value of the variables expectations are being formed about. The second solves for the expectation; in our context this leads to an error correction model, and the derived estimator is termed an ECM estimator. In all instances some corrections must be made to the estimators of the long–run response in order to allow for standard inferences. Within the category of E–estimators we investigated two different approaches aimed at effecting such adjustments, but differing by the amounts of information exploited about the structure of the error term.

A final section puts all these estimators to the test in a simple simulation study. It is found that the ECM estimator of the long run response generally works very well, while E estimators are much less reliable. Within the E–estimator class it was found that attempting to exploit knowledge about the structure of the error term in the estimating equation was generally harmful, as it led to certain matrices tending to be singular. When estimating adjustment cost terms the situation is not so simple. In fact, it sometimes seems as if there is an inverse relation between the ability to estimate long run responses and adjustment parameters.

Some interesting questions emerged from the study. Normally an Euler equation estimator is preferred in many rational expectation estimation contexts since it does not require the specification of an expectation generating process, and hence it exhibits a degree of robustness. Its poor performance in many of the experiments here might therefore be viewed as rather disturbing for applied research. But it has to be realized that the robustness property cited above is much less advantageous for determining the long–run response when there are integrated regressors. Then, to consistently estimate the long run response it is only necessary to determine which ***integrated*** variables appear in the expectation generating process; all others may be subsumed within an error term. This fact is not to deny that E–estimators are appealing; the adjustment cost parameters attach to I(0) variables and an incorrect specification of expectations means the ECM method will be working with a mis–specified model, and so inconsistent estimators of those parameters will be found. In these circumstances an E estimator is very attractive. Since the best estimator of the adjustment cost parameters frequently seems to come from the Euler equation, while the best estimator for the long run response is that from the ECM, the possibility is raised of using each estimator for the task at which it has a comparative advantage. Whether this conclusion holds up to further analysis will be dealt with in future work. Another extension is to models with expectations that do not derive from Euler equations. The estimation methods advanced here would continue to apply if conditions can be found that would enable the solution of models with expectations and integrated variables; an important advantage of the linear quadratic model is that a substantial literature can be drawn upon for that purpose. Wickens (1990) takes up that problem elsewhere in this volume.

TABLE 1

Size of Test Statistics as the Cost of Adjustment Term Varies

	T = 100			T = 200		
	.1	.05	.01	.1	.05	.01
1. $\theta = 1,\ \delta = .5,\ \Sigma = \begin{bmatrix} 1 & .5 \\ .5 & 1 \end{bmatrix}$						
$\tilde{\theta}_{OLS}$	.24	.16	.06	.27	.18	.07
$\hat{\theta}_{PH}$	.16	.10	.03	.14	.08	.02
$\hat{\theta}_{E}$	.04	.02	.01	.03	.01	.00
$\hat{\theta}_{ECM}$	.14	.08	.02	.11	.05	.02
$\hat{\delta}_{E}$	.09	.04	.01	.08	.05	.02
$\hat{\delta}_{ECM}$	.10	.04	.01	.10	.04	.01
2. $\theta = 1,\ \delta = .2,\ \Sigma = \begin{bmatrix} 1 & .5 \\ .5 & 1 \end{bmatrix}$						
$\tilde{\theta}_{OLS}$	.31	.22	.10	.34	.25	.11
$\hat{\theta}_{PH}$	.16	.09	.03	.14	.08	.02
$\hat{\theta}_{E}$	.03	.01	.00	.01	.00	.00
$\hat{\theta}_{ECM}$	.15	.09	.02	.12	.06	.02
$\hat{\delta}_{E}$	.11	.07	.03	.10	.06	.02
$\hat{\delta}_{ECM}$	.07	.03	.01	.07	.04	.01
3. $\theta = 1,\ \delta = 10,\ \Sigma = \begin{bmatrix} 1 & .5 \\ .5 & 1 \end{bmatrix}$						
$\tilde{\theta}_{OLS}$	.14	.08	.00	.13	.07	.02
$\hat{\theta}_{PH}$	.15	.08	.03	.10	.06	.02
$\hat{\theta}_{E}$	.09	.05	.01	.07	.04	.01
$\hat{\theta}_{ECM}$	.10	.06	.01	.08	.05	.02
$\hat{\delta}_{E}$	.82	.78	.70	.69	.65	.56
$\hat{\delta}_{ECM}$	.23	.20	.15	.17	.14	.08

TABLE 2

Size of Test Statistics as Endogeneity Increases

	T = 100			T = 200		
	.10	.05	.01	.10	.05	.01
1. $\theta = 1,\ \delta = .5,\ \Sigma = \begin{bmatrix} 1 & .5 \\ .5 & 1 \end{bmatrix}$						
$\tilde{\theta}_{OLS}$	.24	.16	.06	.27	.18	.07
$\hat{\theta}_{PH}$	.16	.10	.03	.14	.08	.02
$\hat{\theta}_{E}$	.04	.02	.01	.03	.01	.00
$\hat{\theta}_{ECM}$	.14	.08	.02	.11	.05	.02
$\hat{\delta}_{E}$	.09	.04	.01	.08	.05	.02
$\hat{\delta}_{ECM}$	.10	.04	.01	.10	.04	.01
2. $\theta = 1,\ \delta = .5,\ \Sigma = \begin{bmatrix} 1 & .7 \\ .7 & 1 \end{bmatrix}$						
$\tilde{\theta}_{OLS}$	.24	.16	.06	.27	.19	.07
$\hat{\theta}_{PH}$	.15	.10	.03	.14	.08	.02
$\hat{\theta}_{IV}$	.03	.01	.01	.02	.00	.00
$\hat{\theta}_{ECM}$	.14	.08	.02	.10	.05	.02
$\hat{\delta}_{IV}$	.10	.04	.01	.09	.04	.01
$\hat{\delta}_{ECM}$	.08	.05	.02	.10	.04	.01

TABLE 3

Size of Test Statistics as the Long Run Response Changes

	T = 100			T = 200		
	.10	.05	.01	.10	.05	.01
1. $\theta = 1,\ \delta = .5,\ \Sigma = \begin{bmatrix} 1 & .5 \\ .5 & 1 \end{bmatrix}$						
$\tilde{\theta}_{OLS}$	.24	.16	.06	.27	.18	.07
$\hat{\theta}_{PH}$	.16	.10	.03	.14	.08	.02
$\hat{\theta}_{E}$	.04	.02	.01	.03	.01	.00
$\hat{\theta}_{ECM}$	.14	.08	.02	.11	.05	.02
$\hat{\delta}_{E}$	.09	.04	.01	.08	.05	.02
$\hat{\delta}_{ECM}$	.10	.04	.01	.10	.04	.01
2. $\theta = 10,\ \delta = .5,\ \Sigma = \begin{bmatrix} 1 & .5 \\ .5 & 1 \end{bmatrix}$						
$\tilde{\theta}_{OLS}$	.26	.18	.07	.29	.20	.06
$\hat{\theta}_{PH}$	.07	.03	.01	.03	.02	.01
$\hat{\theta}_{E}$	.03	.02	.00	.01	.00	.00
$\hat{\theta}_{ECM}$	.10	.05	.01	.11	.05	.01
$\hat{\delta}_{E}$	.11	.07	.03	.10	.05	.02
$\hat{\delta}_{ECM}$	.02	.01	.00	.02	.00	.00

Footnotes

[2]Pesaran (1989) derives results when the cost function is augmented by higher order adjustment terms such as $(\Delta^2 y_s)^2$.

[3]It would be possible to allow e_t to be serially correlated provided it remained an integrated process of order zero. This last restriction cannot be relaxed as it would mean that the error term in "regressions" would be the same order of integration as the regressors.

[4]x_t is taken to be a scalar here for expositional purposes.

[5]Actually, in their simulation study they have a third instrument. It is not hard to show that the result we give is invariant to the number of instruments. Notice that the assumption of a fixed θ is for convenience only. If the model cannot be identified even when θ is known, the situation will be worse when θ needs to be estimated.

[6]This point is emphasized in the context of models such as these by Pesaran (1987, p. 193).

[7]He mentions that, by conditioning upon $\hat{\beta}$, i.e., forming $\Delta y_t - \hat{\beta}\Delta y_{t+1}$ as a dependent variable, it would be possible to get asymptotic normality for an IV estimator of δ.

[8]Even though β and δ may be identified when θ is known, the conclusion does not extend to the situation when θ is unknown, as there are then three parameters in (5) but only two regressors, y_{t-1} and x_t, as x_{t-1} cancels. Consequently, one of the three parameters cannot be identified even when $\rho_1 < 1$. In many contexts θ will in fact be known. For example, if y_t and x_t are logs of variables, and the ratio of the levels is a constant, θ would be unity. Notice that the "counting rule" just used is only a necessary condition, as evidenced by the case $\rho_1 = 1$ discussed in the text, where there were two parameters and two regressors, but only a single parameter was identified.

[9]The eigenvalues were found by estimating the correlation matrix from 2500 observations generated according to the theoretical model. This introduces a slight error. For example when $\rho_2=0$ the smallest eigenvalue is truly zero, but computed from the simulation it was 2.5×10^{-5}. However, this discrepency is clearly not important for the point we wish to make. Programs to perform the simulations and eigenvalue computations were done in GAUSS.

[10]Dolado et al. (1989) were the first to notice the curious result that, when $\beta = 1$, z_t is I(0), whereas for $\beta \neq 1$ it is I(1).

[11]If ζ_t is a covariance stationary process the long run covariance matrix is proportional to the spectral density at the origin.

[12]It is interesting to observe that this is exactly the regression employed by Kennan (1979) for estimating θ.

[13]It is easy to see that $T^{-1}\Sigma x_t\eta_t - T^{-1}\Sigma x_t(1-\beta\lambda)e_t$ is $o_p(1)$ (Case I, $\rho_1=1$) owing to the $T^{1/2}$ consistency of $\tilde{\lambda}$ and the fact that $T^{-1}\Sigma x_t\Delta y_t$ converges in law to a random variable. Consequently, the disturbance term in (19a) can be regarded as $(1-\beta\lambda)e_t$.

[14]The fact that $z_{t-1}-\Delta x_t'\theta$ is I(0) and x_t is I(1) means that the regressors are asymptotically uncorrelated, thereby enabling a sequential approach.

[15]λ is a function of δ once β is prescribed, and the latter is set to .97 in these experiments. All long run covariance matrices are estimated as in Newey and West (1987) using $T^{1/3}$ lags.

[16]Singularity in the long–run covariance matrix has been the Achilles heel of many "non–parametric adjustments" such as those being made here. An early example arose in testing for unit roots in the presence of an MA(1) error with a negative coefficient close to unity—see Schwert (1987). If the coefficient had been –1 the matrix would be singular.

References

Bergstrom, A. R. (1967) **The Construction and Use of Economic Models** (London: English Universities Press).

Bewley, R. A. (1979), "The Direct Estimation of the Equilibrium Response in a Linear Dynamic Model", **Economics Letters**, 3, 357–362.

Blanchard, O. J. and A. Melino (1986), "The Cyclical Behaviour of Prices and Quantities: The Case of the Automobile Market", **Journal of Monetary Economics**, 17, 379–407.

Dolado, J. J., J. W. Galbraith and A. Banerjee (1989), "Estimating Euler Equations with Integrated Series", **Oxford Applied Economics Discussion Paper No. 81**.

Engle, R. F. and C. W. Granger (1987), "Co–Integration and Error Correction: Representation, Estimation and Testing", **Econometrica**, 49, 1057–1072.

Hansen, L.P. (1982), "Large Sample Properties of Generalized Method of Moments Estimators", **Econometrica**, 50, 1029–1054.

Hansen,L.P and T.J.Sargent (1980), " Fomulating and Estimating Dynamic Linear Rational Expectations Models", **Journal of Economic Dynamics and Control**, 2,7–46.

Hansen, L. P. and K. J. Singleton (1988), "Efficient Estimation of Asset Pricing Models with Moving Average Errors" (mimeo)

Hansen, B."Testing for Structural Change of Unknown Form in Models with Non–Stationary Regressors", (mimeo, University of Rochester).

Ilmakunnas, P. (1989), "Survey Expectations vs. Rational Expectations in the Estimation of a Dynamic Model: Demand for Labour in Finnish Manufacturing," **Oxford Bulletin of Economics and Statistics**, 51, 297–314.

Johansen,S. (1988), " Statistical Analysis of Cointegration Vectors", **Journal of Economic Dynamics and Control**, 12, 231–254.

Kennan, J. (1979), "The Estimation of Partial Adjustment Models with Rational Expectations", **Econometrica**, 47, 1441–1455.

Layard, R. and S. Nickell (1986), "Unemployment in Britain", **Economica**, 53, S121–S170.

McCallum, B. (1976), "Rational Expectations and the Natural Rate Hypothesis: Some Consistent Estimates", **Econometrica**, 44, 43–52.

Newey, W. K. (1984), "A Method of Moments Interpretation of Sequential Estimators", **Economics Letters**, 14, 201–206.

Newey, W.K. and K.D. West (1987), "A Simple Positive Semi–Definite Heteroskedasticity and Autocorrelation Consistent Covariance Matrix", **Econometrica**, 703–708.

Nickell, S. (1985), "Error Correction, Partial Adjustment and All That: An Expository Note", **Oxford Bulletin of Economics and Statistics**, 47, 119–129.

Nickell, S. (1987), "Dynamic Models of Labour Demand," in O. Ashenfelter and R. Layard (eds.), **Handbook of Labour Economics** (North Holland, Amsterdam).

Pagan, A. R. (1984), "Econometric Issues in the Analysis of Regressions with Generated Regressors", **International Economic Review**, 25, 221–247.

Park,J.Y. (1988), "Canonical Co–integrating Regressions", **Center for Analytic Economics, Working Paper** No. 88–29, Cornell University.

Park, J. Y. and P.C.B. Phillips (1988), "Statistical Inference in Regressions with Integrated Processes: Part 1", **Econometric Theory**, 4, 468–497.

Park, J. Y. and P.C.B. Phillips (1989), "Statistical Inference in Regressions with Integrated Processes: Part 2", **Econometric Theory**, 5, 95–131.

Pesaran, H., " Costly Adjustment under Rational Expectations: A Generalization", (mimeo, U.C.L.A.)

Pesaran, H. (1987), **The Limits to Rational Expectations** (Blackwell, New York).

Phillips, P.C.B. (1987), "Time Series Regression with a Unit Root", **Econometrica**, 55, 277–301.

Phillips,P.C.B. (1988a), "Multiple Regression with Integrated Time Series", **Contemporary Mathematics**, 80, 79–104.

Phillips,P.C.B. (1988b), " Optimal Inference in Cointegrated Systems", **Cowles Foundation Discussion Paper** No. 866

Phillips, P.C.B. (1989), "Partially Identified Econometric Models", **Econometric Theory**, 5, 181–240.

Phillips, P.C.B. (1990),"Spectral Regression for Co–integrated Time Series", in W. Barnett, J. Powell and G. Tauchen (eds), **Nonparametric and Semiparametric Methods in Economics and Statistics** (Cambridge University Press, forthcoming)

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.

Sargent, T.J. (1978), "Estimation of Dynamic Labour Demand Schedules Under Rational Expectations", **Journal of Political Economy**, 86, 1009–1044.

Schwert, G.W. (1987), "Effects of Model Misspecification on Tests for Unit Roots in Macroeconomic Data", **Journal of Monetary Economics**, 20, 73–103.

Sims, C. A., J. Stock and M. W. Watson (1987), "Inference in Linear Time Series Models with Some Unit Roots", **Econometrica**, 58. 113–144.

West, K.D. (1986), "Full versus Limited Information Estimation of a Rational Expectations Model: Some Numerical Comparisons", **Journal of Econometrics**, 33, 367–385.

West, K. D. (1988), "Asymptotic Normality When Regressors Have a Unit Root", **Econometrica**, 56, 1397–1417.

Wren–Lewis, S. (1986), "An Econometric Model of U.K. Manufacturing Employment Using Survey Data on Expected Output", **Journal of Applied Econometrics**, 1, 297–316.

Wickens, M. R. (1982), "The Efficient Estimation of Econometric Models with Rational Expectations", **Review of Economic Studies**, 49, 817–838.

Wickens, M.R. (1990), "Rational Expectations and Integrated Variables" (this book)