ON ERROR CORRECTION MODELS: SPECIFICATION, INTERPRETATION, ESTIMATION

George Alogoskoufis

Birkbeck College, University of London and CEPR

Ron Smith

Birkbeck College, University of London

Abstract. Error Correction Models (ECMs) have proved a popular organising principle in applied econometrics, despite the lack of consensus as to exactly what constitutes their defining characteristic, and the rather limited role that has been given to economic theory by their proponents. This paper uses a historical survey of the evolution of ECMs to explain the alternative specifications and interpretations and proceeds to examine their implications for estimation. The various approaches are illustrated for wage equations by application to UK labour market data 1855–1987. We demonstrate that error correction models impose strong and testable non-linear restrictions on dynamic econometric equations, and that they do not obviate the need for modelling the process of expectations formation. With the exception of a few special cases, both the non-linear restrictions and the modelling of expectations have been ignored by those who have treated ECMs as merely reparameterisations of dynamic linear regression models or vector autoregressions.

Keywords. Error correction models; wage equations.

1. Introduction

The notion of an ‘Error Correction Model’ (ECM) is considered to be a very powerful organising principle in applied econometrics and has been applied widely, especially since the appearance of the seminal paper by Davidson et al. (1978). It has also prompted a range of statistical developments, most notably the concept of cointegration (Engle and Granger 1987). In fact, in a recent paper Hylleberg and Mizon (1989) suggest that ‘when estimating structural models it is our experience from practical applications that the error correction formulation provides an excellent framework within which it is possible to apply both the data information and the information obtainable from economic theory’ (p. 124). Similar views have been expressed by others.

This quotation and others like it raise two sets of questions. The first is related to the fact that the success of the principle has occurred despite the lack of any consensus as to exactly what defines an error correction model. The ECM has been interpreted as: a method of adjusting a policy instrument to maintain a
target variable close to its desired value (Phillips 1957); A reparameterisation of
the dynamic linear regression models in terms of differences and levels (some
sections of Hendry, Pagan and Sargan 1984); a restricted form of a dynamic
linear regression model, which imposes long-run proportionality (linear
homogeneity) among some regressors (other sections of Hendry, Pagan and
Sargan 1984); a generalization of the simple ‘partial adjustment mechanism’
(Hendry and von Ungern Sternberg 1981); a quasi reduced form derived from
special cases of rational expectation models of intertemporal optimization with
costs of adjustment (Nickell 1985); and a particular representation of a vector
autoregression appropriate for cointegrated vectors (Engle and Granger 1987).

The second set of issues raised by the Hylleberg and Mizon quotation is that,
contrary to the impression given, most of the applications of ECM appear to be
characterized by their relative lack of concern with formal economic theory in
general and the modelling of expectations in particular. Using the distinction
suggested by Aldrich (1989), the practice of ECM proponents has been to model
the time-series relationships in the data and then try to interpret the results,
rather than to derive relations directly from economic theory, impose the error
correction mechanism as an auxiliary adjustment hypothesis, and then estimate
the resulting dynamic model. The methodological position of modelling the time
series properties of the data, and then trying to put interpretations on the
resulting model is expounded in various places, one of the early ones being
Mizon (1977). This methodological position has been refined in recent years. An
influential summing up can be found in Hendry and Wallis (1984).

The present survey has three aims, which are related to the two sets of
questions discussed above. The first is to clarify the differences between various
specifications of error correction models. We thus provide a brief historical
account of their development and evolution, and survey their similarities and
differences. Our second aim is to distinguish between different specifications and
interpretations of ECMs, and to discuss their economic theoretical status. This
enables us to survey the range of controversies that have been provoked. Our
third aim is to contribute to the discussion on the specification, estimation and
testing of error correction models. We make two points: The first is that the
practice of being cavalier about identification problems and the process of
expectations formation, which characterises some of the most influential
proponents of error correction models, can be very dangerous as it can lead to
very misleading conclusions. The second point, which is seldom recognized, is
that ECMs imply tight non-linear restrictions among the parameters of single
dynamic econometric equations. With the exception of special cases these
restrictions have so far been ignored in the literature. To illustrate our points we
go through an empirical example associated with the early development of the
concept, namely the determination of money wages in the United Kingdom. We
use annual historical data from 1855 to 1987.

The rest of the paper is organised as follows: In section 2, we examine the
evolution of the concept. We go through the approaches of Phillips, Sargan,
Hendry and Granger. Section 3 is devoted to theoretical considerations. We
briefly survey some of the problems in interpreting long run parameters and the literature on error correction models as optimal adjustment rules. In section 4 we move to the application of the different concepts, stressing the need for modelling the process of expectations formation, and the fact that the predominant specification of ECMs implies a number of non-linear overidentifying restrictions among the parameters of single dynamic econometric equations. The last section sums up our conclusions.

2. The evolution of the concept

The early development of ECM is very much an LSE story — Phillips, Sargan and Hendry in particular. Thus, we shall examine the evolution of the concept mainly in terms of their work. Since we are re-reading their work primarily in terms of later developments, rather than in terms of their own contemporary intentions, we are running the risk that the discussion may in places seem anachronistic and we try to point out where our interpretation differs from the concerns of the original authors.

2.1. Phillips

Although there are plausible precursors, Phillips (1954, 1957) introduced the terminology of error correction to economics, in his analysis of feedback control mechanisms for stabilisation policy.

Consider a state variable $x(t)$ influenced by a control variable $y(t)$ and exogenous shocks. There is a desired level for $x(t)$, $x^*(t)$, and there is an error associated with it, $e(t) = x^*(t) - x(t)$. What Phillips called ‘error correction type stabilisation policy’ then adjusts the control variable according to proportional, integral and derivative (PID) feedbacks from the errors:

$$y(t) = y^*(t) + f_p e(t) + f_i \int e(t) \, dt + f_d \frac{\partial e(t)}{\partial t}$$  (1)

Phillips’ formulation of the control variable was slightly more complicated since he allowed for lags in the policy implementation. In addition he did not provide for the intercept denoted above by $y^*$. This is the equilibrium level of the control variable, at which $x = x^*$, and is a natural addition to the model (e.g. Turnovsky 1977, p. 322). The control aspect of the problem with which Phillips was concerned are discussed further in Salmon (1982, 1988).

In discrete time, an equivalent form to (1) is:

$$y_t = y_t^* + k_p e_{t-1} + k_i \sum_{i=1}^{\infty} e_{t-i} + k_d \Delta e_t$$  (2)

Applying the first difference operator to (2) gives:

$$\Delta y_t = \Delta y_t^* + k_p \Delta e_{t-1} + k_i e_{t-1} + k_d \Delta^2 e_t$$  (3)

This PID equation is the first way of defining the ECM that we might consider.
One source of confusion is that subsequent literature, which has started from a model in first differences like (3), has tended to refer to the $e_{t-1}$ term as proportional feedback, whereas in the Phillips formulation it is integral feedback. In addition, later literature has often been specified directly in terms of a target value for $y$, which may be a function of exogenous variables, dispensing completely with the idea of controlling $x$ around its target.

Although Phillips did not use it explicitly for this purpose, this basic approach can very easily be applied to the concerns of the famous Phillips (1958) article on wage inflation. There he says that the rate of growth of wages depends on the level and change in unemployment and on the rate of inflation. Suppose unemployment is treated as the error (the deviation between actual employment and the target full employment level). Then, the wage can be treated as the control variable (of unions say) which feeds back on the error. If we use this interpretation, which is not explicitly given by Phillips himself (see Desai 1984 for a discussion of the derivation of the Phillips curve), then we can write the PID version of his curve in log-linear form as:

$$
\Delta w_t = A \Delta w^*_t + k_p \Delta u_{t-1} + k_d \Delta^2 u_t
$$

where $\Delta w^*_t$ is the equilibrium or ‘target’ wage inflation rate, $w$ is the log of wages and $u$ the unemployment rate.

2.2. Sargan

The next milestone in the evolution of ECMs for econometric purposes is Sargan’s (1964) Colston paper. This is primarily a study in econometric methodology dealing with different methods of estimating structural equations with autocorrelated errors. The wage and price equations are used to illustrate the techniques. Sargan estimates an equation of the form,

$$
\Delta w_t = a_0 + a_1 \Delta p_{t-1} - a_2 u_{t-1} - a_3 (w - p)_{t-1}
$$

where $p$ is the log of the price level. This is an annual data version of Sargan’s equation (15) p. 291, where two additional terms, an incomes policy dummy and a time trend are being ignored.

Sargan gives two interpretations of this equation. Firstly, he interprets it as the combination of a lagged partial adjustment process, whereby nominal wages adjust to remove part of the gap between actual and ‘target’ real wages in the previous period, and an equation for target real wages $(w - p)^*$. This interpretation can be written as follows:

$$
\Delta w_t = \gamma [(w - p)^*_{t-1} - (w - p)_{t-1}]
$$

$$
(w - p)^*_t = \omega_t + \theta \Delta p_t - \eta u_t
$$

where $\gamma$ is the proportion of the gap between past equilibrium and actual real wages that is adjusted in the current period, $\omega_t$ represents some exogenous factors (productivity and incomes policies in Sargan’s case), $\eta$ is the semi-elasticity of equilibrium real wages with respect to the unemployment rate and
θ is the responsiveness of equilibrium real wages to inflation (sic). This formulation is a special case of Phillips' ‘error correction’, with integral control only, where the error is defined in terms of the deviation of real wages from target. In fact, Phillips’ control papers are not referenced, though his (1958) paper is. Gilbert (1989) discusses the influence of Phillips on Sargan’s formulation. The Sargan formulation has been widely adopted in subsequent work, and Dawson (1981) extends it to a full PID form.

One strange feature of the interpretation in (6a) is that equilibrium or target real wages depend on price inflation, a clear violation of the homogeneity properties of general equilibrium models. However, this feature can be dispensed with, if one assumes that (6a) is an adjustment equation in expected real wages, and that inflationary expectations are a function of lagged inflation. Thus (6a) can be modified to,

$$
\Delta w_t - \Delta \rho_t = \gamma [(w - p)_t - (w - p)_{t-1}]
$$

$$
(w - p)_t = \omega - \eta u_t
$$

$$
\Delta \rho_t = \pi (1 - \rho) + \rho \Delta p_{t-1}
$$

The superscript $e$ denotes the expectation of the relevant variable, assumed currently unobserved by wage setters. $\pi$ and $\rho$ are parameters of an AR(1) expectations process, which under rational expectations could be related to the actual process governing inflation.

Sargan also interprets equation (5) as a Phillips curve, where $E$ is a refined measure of the excess supply of labour.

$$
\Delta w_t = \beta_0 + \beta_1 \Delta \rho_t - \beta_2 E_{t-1}
$$

$$
E_t = u_t + \gamma_2 (w - p)_t
$$

This interpretation suggests that the difference between $E$ and $u$ is positively related to the real wage because of part-time work and temporary labour hoarding.

The point to note at this stage, which is quite crucial for what follows, is that Sargan is quite happy to allow for the same estimates to be given a number of observationally equivalent interpretations. Thus we see that the same econometric equation can be given three quite different theoretical interpretations, with possibly different policy implications, by simply being reparameterised in terms of different unobservables: the target real wage, expected inflation or excess labour supply.

2.3. Hendry

The current popularity of ECMs is largely due to David Hendry, whose work was influenced by both Phillips and Sargan. Gilbert (1986) surveys Hendry’s methodological position, which is distributed over many empirical papers. For his contributions on error correction models one of the most influential of these is Davidson, Hendry, Srba and Yeo (1978). This introduced the ECM form for
the aggregate time-series relationship between consumers expenditure and income. Davidson et al. reference Sargan (1964) but not Phillips, and although the estimated relationship is given a 'feedback' interpretation, among many others, the term 'error correction' is not used in the paper. The term is introduced in Hendry (1980), and the links to the PID model are attempted in Hendry and von Ungern Sternberg (1981). Within Hendry's work there are, at least two ways that an ECM can be characterised. Both of these appear in Hendry, Pagan and Sargan (1984).

For a start, Hendry emphasised the importance of 'general to specific modelling' (Mizon 1977), and in this context the ECM can be interpreted as a reparameterisation of the general 'auto-regressive distributed lag' (ADL) or 'dynamic linear regression' (DLR) models. For example, for two variables $y$ and $x$, the first order DLR is:

$$y_t = a_0 + b_0 x_t + b_1 x_{t-1} + a_1 y_{t-1} + \epsilon_t$$

where $\epsilon_t$ is a white noise residual. (7) can be re-written as:

$$\Delta y_t = \alpha_0 + \beta_0 \Delta x_t + \beta_1 x_{t-1} + \alpha_1 y_{t-1} + \epsilon_t$$

where $\alpha_0 = a_0$, $b_0 = \beta_0$, $\alpha 1 = a_1 - 1$, $\beta_1 = (b_0 + b_1)$, or

$$\Delta y_t = \beta_0 \Delta x_t - \lambda(y_{t-1} - \psi x_{t-1} - \psi_0) + \epsilon_t$$

where $\lambda = -\alpha_1$, $\psi_1 = \beta_1/\alpha_1 = (b_0 + b_1)/(1 - a_1)$, $\psi_0 = \alpha_0/\alpha_1$.

This definition of the ECM imposes no restrictions on the DLR, but is in terms of different parameters, which can be given an economic interpretation as impact effects $\beta_0$, a scalar adjustment coefficient, $\lambda$, and long-run effects $\psi_1$. The latter are interpreted as the parameters of an equilibrium relationship about which economic theory is informative. Once again this raises the point that there are a large variety of representations which can be achieved through reparameterisations. They are all observationally equivalent, thus there are no statistical criteria that we can use to choose between them. The questions that arise then, relate to the parameters of interest from the point of view of economic theory. These questions can only be answered by an explicit theory. For certain applications, the restriction $\psi_1 = 1$ may be appropriate. For example, for variables in logarithmic form this ensures that the relevant ratios are constant in the long run. This unit restriction is imposed in Davidson et al. through the use of the logarithm of the lagged average propensity to consume. Unit coefficients in logarithmic specifications have a wide range of other applications like the velocity of money (Hendry 1980), the wage-gap from Cobb–Douglas marginal productivity equations (Bruno and Sachs 1985), or purchasing power parity (Edison 1987). They also have applications in levels, like in the term structure of interest rates. In Hendry's own work the unit coefficient hypothesis seems to have been the most common defining characteristic of the ECM, and this gives us its best known representation, which is an equation of the form:

$$\Delta y_t = \alpha + \beta \Delta x_t - \gamma(y_{t-1} - x_{t-1}) + \epsilon_t$$

(8)
The static long-run solution of this equation (when \( y_t = y_{t-1} = y \) and \( x_t = x_{t-1} = x \)) is:

\[
y = \alpha / \gamma + x
\]  

(9)

Davidson et al. interpret \( \exp(\alpha / \gamma) \) as an estimate of the long run average propensity to consume in a consumption function which postulates proportionality between consumers’ expenditure and income. This takes the form, \( Y = KX \).

However, proportionality is a statement about the form of the eventual equilibrium and has little to do with the nature of adjustment processes. Nevertheless, while it seems rather strange to treat a restriction on the character of the equilibrium as the defining characteristic of a model of dynamic adjustment, it is precisely the unit elasticity that makes this popular model operational and more than merely a parameterisation of the dynamic linear regression model. We shall return to the question of the restrictions implied by more general version of these ECMs in section 4 below.

2.4. Granger

The latest twist in the ECM saga has come from Granger and associates. As time series statisticians they noted that most economic series are highly trended with stationary growth rates. This implies that they are integrated of order one, \( I(1) \). This means that they become stationary after being differenced once. Then they go on to ask how a stationary variable \( \Delta y_t \) (integrated of order zero) in (7a) could be explained by two non stationary variables \( y_{t-1} \) and \( x_{t-1} \), which are \( I(1) \). The two sides of the equation are of different orders of integration, unless the linear combination \( y_t - \psi x_t \) is also stationary.

In general, linear combinations of \( I(1) \) variables will also be \( I(1) \), but if they happen to be \( I(0) \), the variables are said to be cointegrated (Engle and Granger 1987). If they are cointegrated then there exists an error correction representation. Conversely, if there is an error correction representation for the series, they are cointegrated. However, the definition of an error correction representation differs from the earlier ECMs we have used.

Engle and Granger define a vector stochastic process \( x_t \), which is \( I(1) \), as having an error correction representation if it can be expressed as

\[
A(L)(1 - L)x_t = -\gamma e_{t-1} + e_t
\]  

(10)

\( L \) is the backward shift or lag operator, such that \((1 - L)x_t = x_t - x_{t-1}\). \( A(L) \) is a polynomial in \( L \) of the form \((a_0 + a_1L + a_2L^2 + \cdots)\). \( e_t \) is a stationary multivariate disturbance. It is assumed that \( A(0) = I \), \( A(1) \) has all elements finite and \( \gamma \neq 0 \). The cointegrating vector is \( \alpha \), where \( e_t = \alpha' x_t \) is \( I(0) \). There may be more than one such vector. The equilibrium is interpreted as \( \alpha' x_t = 0 \), thus \( e_t \) is interpreted as a measure of the error or deviation from equilibrium.

Engle and Granger (1987) stress the difference from earlier definitions. This definition of an ECM is explicitly multivariate. It does not distinguish between
endogenous and exogenous variables, though it may allow some inferences about
Granger causality, and current values do not appear on the right hand side.
Homogeneity or unit coefficients are not intrinsic to the definition and the vector
\( \alpha \) consists of unknown parameters to be estimated. However, these parameters
do not necessarily have a theoretical interpretation.

3. Theoretical considerations

At least two conclusions can be drawn from our examination of the evolution
of the concept of error correction models in the previous section.

The first is that different authors have different views as to what constitutes
an ECM. There seem to be at least three lines, one associated with Phillips,
the other associated with Sargan–Hendry, and the last associated with
Engle–Granger. Of the three approaches, Phillips, Sargan–Hendry seem to have
more common ground, in as much as they seem to be interested in dynamic
decision rules of economic agents. For them ECMs seem to be structural
representations of dynamic adjustment towards some equilibrium about which
economic theory can be informative. Engle, Granger and associates seem to view
ECMs not as structural decision rules, but as statistical representations only
(reduced forms). As a result, their definition of equilibrium is also statistical
(variables stay close together).

The second conclusion is that, notwithstanding the differences in emphasis
referred to above, all authors seem to envisage a rather limited role for economic
theory. In contrast to the Cowles Commission approach which has largely
dominated theoretical blueprints of how to specify econometric equations, the
ECM models are not directly derived from theory, and therefore the estimated
parameters only bear an indirect relation to theoretical parameters of interest.
This is clearly borne out by the fact that the leading practitioners are quite happy
with alternative interpretations. This is amply illustrated in both Sargan’s (1964)
paper and in Hendry’s (1980) contrast of the two methodologies. Practitioners
of this alternative approach seem to have to give \textit{ex post} interpretations to their
estimated parameters, rather than setting up from the start the relation between
their statistical and theoretical parameters of interest. One of the main problems
with this approach is that multiple interpretations are given much more emphasis
in this methodology, than in the Cowles Commission approach. This is clearly
illustrated in a well known problem in this literature, that of solving for long run
parameters.

3.1. ‘Interpreting’ long run parameters

The derivation of long-run parameters has prompted considerable controversy,
which seems to have been bred by the practice of using theory \textit{ex post} as a way
to interpret equilibrium relationships. However, the estimated ‘long-run
coefficients’ are likely to be a mixture of adjustment and expectational as well
as long run structural parameters.
To illustrate this suppose the 'long run' equilibrium in logs is:

\[ y^* = k + x \]  

Then assume an expectations augmented ECM short run relation given by:

\[ \Delta y_t = \beta (x_t^e - x_{t-1}) + \gamma (y_{t-1}^* - y_{t-1}) \]  

where \( x_t \) follows a stationary AR(1) process with mean \( \pi \), and superscript \( e \) denotes \( E(x_t \mid I_{t-1}) \). (12) is of exactly the same form as (8), only that \( x \) has been replaced by its expectation. The expectation of \( x \) is,

\[ x_t^e = \pi (1 - \rho) + \rho x_{t-1} \]  

Substituting (13) in (12), in terms of observables the relationship is:

\[ \Delta y_t = \beta \pi (1 - \rho) + \gamma k + (\beta \rho - \beta + \gamma) x_{t-1} - \gamma y_{t-1} \]  

with long run solution:

\[ y = \frac{\pi \beta (1 - \rho) + \gamma k}{\gamma} + \left[ 1 + \frac{\beta (\rho - 1)}{\gamma} \right] x \]  

The long run coefficient of \( x \) will reflect equilibrium, adjustment and expectations parameters. Thus, if one concentrates on (14) only, ignoring the further information given by (13), one would tend to get an estimate of the long run impact of \( x \) on \( y \) that is less than unity. 3 This bias will be avoided if one recognizes that since \( x \) follows an AR(1) process with mean \( \pi \), in the long run \( x \) ought to be replaced by \( \pi \). Then (15) reduces to,

\[ y = k + \pi = k + x \]  

In (16) the calculated long run elasticity will be the correct one.

The example above, based on expectations, is not the only way in which biases in the calculation of long run coefficients may arise. This problem can be illustrated by a second type of solution, that Davidson et al. (1978) use. This is based on steady state growth effects.

Abstract from expectations in equation (12), and consider the ECM as in equation (8). In a balanced growth path \( \Delta y = \Delta x = g \). Then the steady state solution of (8) depends on the growth rates:

\[ y = g(\beta - 1)/\gamma + \alpha/\gamma + x \]  

There are many observationally equivalent interpretations that can be given to the 'growth coefficient' \( (\beta - 1)/\gamma \). Davidson et al. supposed that the equilibrium depended on the growth rate:

\[ y^* = k + x + \psi g \]  

They then took \( (\beta - 1)/\gamma \) as an estimate of \( \psi \), and \( (\alpha/\gamma) \) as an estimate of \( k \). This was justified by reference to Life Cycle Hypothesis models where the savings ratio is a positive function of the growth rate, because of demographic effects. This interpretation was disputed and prompted a considerable controversy.
ALOGOukoFIS AND SMITH


A second interpretation is that the equilibrium is \( y = x + k \). Then \( k \) is estimated by \( \alpha/\gamma \), but the adjustment process:

\[ \Delta y_t = \beta \Delta x_t - \gamma(y_{t-1} - y_{t-1}^*) \]  \( (18b) \)

is not 'trend neutral' and leads to a steady state bias \( y - y^* = g(\beta - 1)/\gamma \) unless \( \beta = 1 \). The term involving the growth rate is then just a measure of the size of the steady state bias.

A third interpretation uses the same equilibrium condition as the second, but an adjustment process:

\[ \Delta y_t = g + \beta \Delta x_t - \gamma(y_{t-1} - y_{t-1}^*) \]  \( (18c) \)

where \( g \) is a constant, which is the decision-makers' correct estimate of the steady state growth rate. The estimated intercept is \( (g + \gamma k) \) and the extra term just offsets the steady state bias. Thus there is neither steady state bias nor a growth effect in the equilibrium, but \( \alpha/\gamma \) is a biased estimate of \( k \).

Again we have the position that interpretation of the estimated parameters is dependent on theoretical considerations about both the nature of the long run equilibrium, and the adjustment process itself. Parameterisations with quite different theoretical interpretations are observationally equivalent. It may be the case that other evidence can be used to decide between these interpretations (e.g. under the third interpretation we might expect the intercept to shift if the steady state growth shifted) but the estimates themselves cannot inform us about the appropriate interpretation.

What transpires from the above discussion is that there could be multiple interpretations of the estimated equations, even when one starts from the basic premise of an error correction statistical model with long run homogeneity between \( y \) and \( x \).

In view of the above, we shall proceed in the rest of this section to assess various ways in which prior theoretical considerations have been used to derive error correction models. We shall concentrate on the strand of the literature associated with the 'LSE approach', as this is the strand that gives relatively more attention to structural relations. Engle, Granger and associates are avowedly only interested in statistical representations (reduced forms).

3.2. Error correction models as optimal adjustment rules

Adjustment to equilibrium may come about through tatonnement-type processes as prices and quantities respond to imbalances between demand and supply or as the optimal response of individual agents in an intertemporal context. In the context of ECMs, Hendry and Anderson (1977), Hendry and von Ungern Sternberg (1981), Nickell (1985) and Pagan (1985) provide derivations of ECMs
starting from the last case. Common to all the derivations are quadratic loss functions that penalize both deviations from equilibrium and rapid adjustment.

Pagan begins with the problem of an agent who wishes to minimise a myopic quadratic loss function which penalizes both deviations of actual, $y_t$, from target values $y_t^*$ and the deviation of changes in the variable from some normal growth rate $\alpha_t$. We will treat $y_t$ as being the logarithm of some variable, so $\Delta y_t$ is a growth rate. In the usual derivation of the Partial Adjustment Model (PAM) $\alpha_t$ is set equal to zero so any growth is penalised. In this case $y_t$ is chosen to minimise

$$\Lambda = \frac{1}{2} (y_t - y_t^*)^2 + \frac{\theta}{2} (\Delta y_t - \alpha_t)^2$$

(19)

where $\theta$ is the ratio of the marginal cost of adjustment relative to the marginal cost of being away from equilibrium. The optimal $y_t$ that minimizes (19) is given by,

$$y_t = \phi^{-1} y_t^* + \phi^{-1} \theta y_{t-1} + \phi^{-1} \theta \alpha_t$$

(19a)

where $\phi = 1 + \theta$, so $1 - \phi^{-1} \theta = \phi^{-1} = \lambda$. The solution can be written:

$$\Delta y_t = \lambda (y_t^* - y_{t-1}) + (1 - \lambda) \alpha_t$$

(19b)

If $\alpha_t = 0$, we get the usual PAM expression,

$$\Delta y_t = \lambda (y_t^* - y_{t-1})$$

(20)

Note the similarity of (20) with Sargan’s version of the ECM in equation (6a). This version of Sargan’s model is none other than the partial adjustment model. (20) can also be written as,

$$\Delta y_t = \lambda (y_t^* - y_{t-1})$$

(20a)

The standard PAM has the property that the deviation from the long run equilibrium $(y - y^*)_t$, does not go to zero in steady state when the target grows at a constant rate. Its limit is $g(\lambda - 1)/\lambda$, where in steady state we have assumed that $\Delta y_t^* = g \neq 0$. Trend neutrality, zero error in steady state, requires that $\alpha_t = \Delta y_t^*$. It is not clear that trend neutrality is necessarily a desirable property. Nickell (1985) argues, using the forward looking cost of adjustment model discussed below, that with a discount factor less than unity, it will not be worth incurring the additional adjustment costs necessary to catch up completely with a growing target. However, this is in the context of a model where any growth is penalised, rather than only growth which differs from some normal level.

Pagan suggests three ways in which trend neutrality can be achieved: through an intercept, which picks up the constant $(1 - \lambda)g$, or by adding, as proxies for $\alpha_t$, either $\Delta y_{t-1}$ or $\Delta y_t^*$. Suppose we take $\Delta y_t^*$ as a proxy for $\alpha_t$. Then the model becomes,

$$\Delta y_t = (1 - \lambda) \Delta y_t^* + \lambda (y_t^* - y_{t-1}) = \Delta y_t^* - \lambda (y_{t-1} - y_t^*)$$

(21)

(21) could also be rewritten as,

$$y_t = y_t^* + (1 - \lambda) (y_{t-1} - y_t^*)$$

(21a)
which for $y_t^* = \beta' x_t$, is of exactly the same form as a static model with AR1 disturbances. Thus this model involves the common factor restrictions (COMFAC), emphasized by Sargan (1964) and Hendry and Mizon (1978). The autoregressive coefficient is equal to $(1 - \lambda)$. As argued in Hendry, Pagan and Sargan (1986), the usual PAM and the AR1 disturbance models can be seen as special cases of an ECM of the form of (18c).

The question that arises then is whether one could derive the general error correction form from a problem of optimization of quadratic loss function like (19). Hendry and von Ungern Sternberg (1981) justify this form of ECM by augmenting the loss function (19) by a term involving the product of the rate of change of the actual variable and its long run equilibrium value. Their argument is that if adjustment in the actual variable is in the same direction as the change in the equilibrium, then adjustment will be less costly.

$$\Lambda = 1/2(y_t - y_t^*)^2 + \theta_t/2(y_t - y_{t-1})^2 - \theta_2(y_t - y_{t-1})(y_t^* - y_{t-1}^*)$$

(22)

Minimizing (22) with respect to $y_t$, and re-arranging the first order condition yields the ECM, like in equation (18c), where,

$$\beta = (1 + \theta_2)/(1 + \theta_1)$$

$$\gamma = 1/(1 + \theta_1)$$

However, the additional term introduced by Hendry and von Ungern Sternberg (1981) is too ad hoc to be satisfactory. In addition, (22), like (19) is myopic, and therefore inconsistent with the more plausible theoretical assumption that agents are forward looking.

Nickell (1985), derives the ECM from a forward looking quadratic costs of adjustment model supplemented by particular forms of stochastic process driving the target. He also examines the Hendry-von Ungern Sternberg type loss function, but we will use the simpler partial adjustment form. Nickell assumes that agents have an infinite horizon, and minimize the present value of the one period losses given by (19).

$$L_t = \sum_{i=t}^{\infty} \delta^i \left[ \frac{1}{2} (y_i - y_i^*)^2 + \frac{\theta}{2} \Delta y_i^2 \right]$$

(23)

where $\delta$ is the discount factor. The Euler equation takes the form,

$$\delta y_{t+1} - (1 + \delta + \theta^{-1}) y_t + y_{t-1} = -y_t^*/\theta$$

(24a)

Solving the Euler equation requires finding the two roots, $\mu_1 < 1 < \mu_2$ which are the solutions of the characteristic equation,

$$\delta \mu^2 - (1 + \delta + \theta^{-1}) \mu + 1 = 0$$

Calling the stable root (i.e. the one that is less than unity) $\mu$, the optimal policy is then given by:

$$\Delta y_t = (1 - \mu)(\bar{y}_t - y_{t-1})$$

(24b)
the traditional partial adjustment form, but with a forward looking target,

\[ \dot{y}_t = (1 - \delta \mu)E \left[ \sum_{i=1}^{\infty} (\delta \mu)^i y_{t+i}^* \right] \]  \hspace{1cm} (24c)

where the expectation is taken at time \( t \).

To make this operational requires that it be augmented by a model for \( y_{t+i}^* \), that will be used to get the rational expectations solution. As Nickell (1985) shows, if \( y^* \) follows a random walk with drift \( g \), then (24b) becomes,

\[ \Delta y_t = (1 - \mu) \delta \mu g/(1 - \delta \mu) + (1 - \mu) \Delta y_t^* - (1 - \mu)(y_{t-1} - y_{t-1}^*) \]  \hspace{1cm} (25a)

which is just partial adjustment with an intercept to track the growing target, written in an error correction form.

Alternatively, assume that the rate of change of \( y_t^* \) follows a stationary AR(1) process, which is the same as \( y_t^* \) following a second order autoregression with one unit root. This can be written as,

\[ \Delta y_t^* = g(1 - \rho) + \rho \Delta y_{t-1}^* + \nu_t \]  \hspace{1cm} (25b)

where \( g \) is the steady state growth rate, and \( \rho \) is the persistence of the growth rate.

With this specification, (24b) becomes,

\[ \Delta y_t = \frac{\delta \mu (1 - \mu)(1 - \rho)}{(1 - \delta \mu)(1 - \rho \delta \mu)} \Delta y_t^* - (1 - \mu)(y_{t-1} - y_{t-1}^*) \]  \hspace{1cm} (25c)

which looks like a standard ECM. Note that if \( \rho \) is equal to one, the constant vanishes.

Higher order autoregressive processes for \( \Delta y_t^* \) add further lags in that variable. Nickell comments that, 'Since it is almost a stylised fact that aggregate quantity variables in economics follows a second order autoregression with a root close to unity, we may expect to find the error correction mechanism appearing in many different contexts.' (p. 124). Nickell also shows that a random walk with a moving average error also gives rise to an error correction type equation. In each of the cases considered by Nickell the assumed processes for \( y_t^* \) can be used to disentangle the adjustment from the expectational parameters (see Alogoskoufis and Nissim 1981 for an early application to the consumption function, and Domowitz and Hakkio 1990 for an international study of money demand). Thus, shifts in the process generating \( y_t^* \) ought to be reflected in equations of the form of (25a) and (25b). This suggests tests in the spirit of the Lucas (1976) critique. Thus, one could in principle test whether error correction mechanisms are due to the kind of forward looking behaviour, by checking whether shifts in the process generating \( y_t^* \) are reflected in equations of the form of (25b), as is done in Alogoskoufis and Smith (1989) and Hendry (1988).

When Domowitz and Hakkio (1990) implemented this procedure, estimating a generalized version of (25c) that was derived from a cost function of the form of (22), they found that there is no empirical support for the Hendry–von
Ungern Sternberg term in the cost function. Without that extra term, the error correction model works well, and the cross equations restrictions cannot be rejected as long as the ECM parameters are allowed to shift with expectational parameters.

4. Specification and estimation of error correction models with an application to wage setting in the United Kingdom

In this section we proceed to discuss in more detail the specification of the alternative error correction models, and review alternative methods of estimation. We illustrate the different models and methods and the problems that arise with an example associated with the early evolution of the models, namely wage setting in the United Kingdom. The section is structured as follows: First, we discuss statistical representation of the ECM as a re-parameterization of the Dynamic Linear Regression (DLR) Model or of a Vector Autoregressive (VAR) model, involving cointegrating vectors. The approaches differ insofar as the DLR involves conditioning on current and lagged values, whereas the VAR involves conditioning on lagged values only. They both share a common weakness in that the economic interpretation of the estimated parameters is problematic, as neither approach provides a satisfactory solution to the identification problem. We next consider a structural approach, where economic theory considerations are brought in from the start, and they inform the process of estimation. In this approach there is a clear distinction between behavioural and expectational parameters. Three alternative structural models, corresponding to what we called the Philips, Sargan and Hendry interpretations in Section 2, are estimated. We discuss the relation between the treatment of expectations and the estimation method, and demonstrate the usefulness of non-linear estimation methods for structural error correction models.

4.1. Statistical representations of ECMs

In the earlier historical review we identified two more ‘statistical’, versions of the ECM. The Hendry interpretation of the ECM as a reparameterisation of a Dynamic Linear Regression (DLR) model, and the Granger interpretation of the ECM as a representation for a Vector Auto-Regression (VAR) involving cointegrating vectors. These approaches are statistical in that the initial objective is to estimate the parameters of a conditional distribution rather than an explicit theoretical model. The approaches differ in whether the distribution is conditional on current and lagged values (the DLR) or only on lagged values (the VAR).

The data used are five series for the UK 1855–1987. The variables are the logarithms of wages, \( w \), the Feinstein earnings measure; prices, \( p \), the consumers expenditure deflator; real GDP, \( y \); employment, \( l \), including armed forces, and labour force, \( n \). From these, we can also construct four derived variables: real wages, \( w - p \); productivity, \( y - l \); the unemployment rate, \( n - l \); and the share
of wages, \((w - p) - (y - I)\). There are well known problems of definition and interpretation with such long runs of data, but here we shall use them primarily to illustrate the specification and estimation of ECMs.

We shall describe the basic time series properties of the data by first estimating a univariate AR model for each variable, i.e. conditioning only on its own past values; then a VAR for all the variables, i.e. conditioning on the full set of past values; and finally the DLR for wages, which conditions on current and lagged variables. We shall also calculate estimates of the long-run or cointegrating vectors.

The univariate time series model has the form:

\[
\Delta x_t = \alpha_0 + \alpha_1 x_{t-1} + \alpha_2 \Delta x_{t-1} + \alpha_3 t
\]  

(26)

The estimates for the original variables are given in Table 1A, and for the derived variables in Table 1B. The F statistic testing against a further lag suggest that second order autoregressions seem to be adequate to describe these series. Table 1 also gives a test for serial correlation, the standard error of regression for the univariate AR model; the standard error of regression for the VAR discussed below, and the Augmented Dickey Fuller statistics (Dickey and Fuller 1981). These ADFs, calculated as the \(t\) statistics for \(\alpha_1\), indicate that the variables \(w, p, y, e\) and \(n\) are all \(I(1)\). This is in contrast to Hall (1989) who found, with post-war quarterly data, that \(w\) and \(p\) were \(I(2)\). The hypothesis that \(p\) and \(w\) were \(I(2)\) was tested and rejected. This indicates the sensitivity of the test to sample and the sampling frequency. In addition, the power of the test is low and it is not robust to parameter change, e.g. Perron (1990). The fragility of inferences about the degree of integration, or cointegration should be emphasised. The real wage, \(w - p\), and productivity, \(y - I\), also appear \(I(1)\).

However, \(n - I\) (the unemployment rate) and \((w - p) - (y - I)\) the share of wages are \(I(0)\). Again for post-war data Hall found \(u\), equivalent to \(n - I\) to be \(I(1)\). Given these tests, in what follows we shall regard all our series as \(I(1)\) except the share and the unemployment rate, which we shall treat as \(I(0)\). It is worth noting that since linear combinations of cointegrating vectors are also co-integrating vectors, other combinations such as \(w - p - y + n\) would also be cointegrating vectors.

Above we identified the cointegrating vectors from prior theoretical considerations, namely the definition of economically interesting ratios such as the unemployment rate and the labour share. Engle and Granger suggest a two step procedure for estimating the parameters of a cointegrating vector and the ECM. First run a static regression of \(y\) on \(x\), to obtain the elements of the cointegrating vector \(\alpha\). The residuals from this regression provide an estimate of the 'error' and can be tested for stationarity. If the errors are \(I(0)\), they can be used in a second stage in estimating the short run error correction mechanism. Table 5 gives the Engle–Granger estimate of the cointegrating regression, obtained by a regression of \(w\) on the levels of \(p y I\) and \(n\), intercept and trend. This had a DW of 0.4444 and an ADF of \(-4.89\), again confirming that there exists at least one cointegrating vector. In addition, since the share of wages and
unemployment rate seem to be $I(0)$, $w - p - y + l$ was run on $n - l$, intercept and trend. This is given as the restricted E–G estimates in Table 5.

A second order VAR was then estimated of the same form as (26), but making $x$ the 5 element vector, $w$, $p$, $y$, $l$, $n$. The derived variables are also regressed on lagged levels and changes of these 5 variables. Table 2 gives the residual correlation matrix for the VAR. It is clear that by far the highest

**Table IA.** Univariate 2nd order AR 1857–1987

\[ \Delta x_t = \alpha_0 + \alpha_1 x_{t-1} + \alpha_2 \Delta x_{t-1} + \alpha_3 l \]

<table>
<thead>
<tr>
<th>$x$</th>
<th>$w$</th>
<th>$p$</th>
<th>$y$</th>
<th>$l$</th>
<th>$n$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha_0$</td>
<td>$-0.02$</td>
<td>$-0.05$</td>
<td>$0.56^*$</td>
<td>$0.68^*$</td>
<td>$0.28$</td>
</tr>
<tr>
<td></td>
<td>($0.02$)</td>
<td>($0.04$)</td>
<td>($0.25$)</td>
<td>($0.26$)</td>
<td>($0.15$)</td>
</tr>
<tr>
<td>$\alpha_1 \times 10^2$</td>
<td>$-0.08$</td>
<td>$-0.66$</td>
<td>$-5.99^*$</td>
<td>$-7.32^*$</td>
<td>$-3.01$</td>
</tr>
<tr>
<td></td>
<td>($0.71$)</td>
<td>($0.70$)</td>
<td>($2.78$)</td>
<td>($2.88$)</td>
<td>($1.65$)</td>
</tr>
<tr>
<td>$\alpha_2$</td>
<td>$0.60^*$</td>
<td>$0.71^*$</td>
<td>$0.26^*$</td>
<td>$0.36^*$</td>
<td>$0.41^*$</td>
</tr>
<tr>
<td></td>
<td>($0.07$)</td>
<td>($0.07$)</td>
<td>($0.09$)</td>
<td>($0.08$)</td>
<td>$0.08$</td>
</tr>
<tr>
<td>$\alpha_3 \times 10^3$</td>
<td>$0.36$</td>
<td>$0.35$</td>
<td>$1.06^*$</td>
<td>$0.39^*$</td>
<td>$0.15$</td>
</tr>
<tr>
<td></td>
<td>($0.27$)</td>
<td>($0.18$)</td>
<td>($0.48$)</td>
<td>($0.18$)</td>
<td>($0.10$)</td>
</tr>
</tbody>
</table>

**Notes:** Asymptotic Standard Errors are below estimated coefficients. $AUT_1$ is the F version of the LM test for first order residual autocorrelation. $F$ is a test of the exclusion of $\Delta x_{t-2}$. $ADF$ is the Augmented Dickey Fuller $\tau$ statistic on $x_{t-1}$.

**Table IB.** Univariate 2nd order AR 1857–1987

\[ \Delta x_t = \alpha_0 + \alpha_1 x_{t-1} + \alpha_2 \Delta x_{t-1} + \alpha_3 l \]

<table>
<thead>
<tr>
<th>$x$</th>
<th>$n - l$</th>
<th>$w - p$</th>
<th>$y - l$</th>
<th>$(w - p) - (y - l)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha_0$</td>
<td>$0.01$</td>
<td>$-0.03$</td>
<td>$0.002$</td>
<td>$0.85^*$</td>
</tr>
<tr>
<td></td>
<td>($0.01$)</td>
<td>($0.09$)</td>
<td>($0.008$)</td>
<td>($0.24$)</td>
</tr>
<tr>
<td>$\alpha_1$</td>
<td>$-0.16^*$</td>
<td>$0.004$</td>
<td>$-0.007$</td>
<td>$-0.13^*$</td>
</tr>
<tr>
<td></td>
<td>($0.05$)</td>
<td>($0.01$)</td>
<td>($0.02$)</td>
<td>($0.04$)</td>
</tr>
<tr>
<td>$\alpha_2$</td>
<td>$0.22^*$</td>
<td>$0.07$</td>
<td>$0.02$</td>
<td>$0.28^*$</td>
</tr>
<tr>
<td></td>
<td>($0.09$)</td>
<td>($0.09$)</td>
<td>($0.09$)</td>
<td>($0.08$)</td>
</tr>
<tr>
<td>$\alpha_3 \times 10^3$</td>
<td>$0.02$</td>
<td>$0.11$</td>
<td>$-0.17$</td>
<td>$0.20^*$</td>
</tr>
<tr>
<td></td>
<td>($0.04$)</td>
<td>($0.18$)</td>
<td>($0.04$)</td>
<td>($0.08$)</td>
</tr>
</tbody>
</table>

**Notes:** Asymptotic Standard Errors are below estimated coefficients. $AUT_1$ is the F version of the LM test for first order residual autocorrelation. $F$ is a test of the exclusion of $\Delta x_{t-2}$. $ADF$ is the Augmented Dickey Fuller $\tau$ statistic on $x_{t-1}$.
contemporaneous correlations are between wages and prices as one might expect. This implies that including current prices in a wage equation or vice versa will produce large, though potentially spurious, improvements in fit.

The existence of cointegrating vectors implies Granger-causality between the variables. To pursue this causality tests were conducted by deleting variables in turn from equations of the VAR. The results, also shown in Table 2, which focus on the relationship between wages and the other variables, are clear cut. Whereas \( p, y, l, \) and \( n \) are all Granger causal with respect to \( w \), \( w \) is not Granger causal with respect to any of these variables. Such Granger non-causality of nominal wages with respect to the other variables, is a condition required for the strong exogeneity (see Engle et al. 1983) of these other variables with respect to wages. However, as Osborn (1984) points out (p. 94), Granger Causality relates to the final equations of an econometric system, and this information is different in nature from the economic causation used in building a structural model.

Table 3 provides estimates for an unrestricted second order DLR model explaining wages in terms of other four variables. It also gives the long run solution estimated directly using the ‘Bewley’ (1979) transform, discussed by Wickens and Breusch (1988). None of the diagnostic statistics for serial-correlation, functional form, normality or heteroskedasticity suggest any problem, but the model is clearly over-parameterised, with only half the 16 estimated parameters being ‘significant’. While it would certainly be possible to conduct an *ad hoc* specification search which restricted and reparameterised the DLR to produce a more parsimonious model, we will not follow that route. It is noticeable that, as the residual correlation matrix of the VAR suggested, conditioning on current variables substantially improves the fit, reducing the SER from 3.81 to 1.93. In addition, the long-run coefficients of prices, output and employment are not significantly different from unity.

The Granger–Engle procedure cannot deal with cases, such as this, where
Table 3. Dynamic Linear Regression

Dependent Variable $w_t$

<table>
<thead>
<tr>
<th></th>
<th>$x_t$</th>
<th>$x_{t-1}$</th>
<th>$x_{t-2}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$p$</td>
<td>0.9638*</td>
<td>-0.9778*</td>
<td>0.1601</td>
</tr>
<tr>
<td></td>
<td>(0.0603)</td>
<td>(0.1510)</td>
<td>(0.01135)</td>
</tr>
<tr>
<td>$y$</td>
<td>0.2743*</td>
<td>-0.5124*</td>
<td>0.3701*</td>
</tr>
<tr>
<td></td>
<td>(0.0754)</td>
<td>(0.1113)</td>
<td>(0.0900)</td>
</tr>
<tr>
<td>$l$</td>
<td>-0.1294</td>
<td>0.7455*</td>
<td>-0.5301*</td>
</tr>
<tr>
<td></td>
<td>(0.1234)</td>
<td>(0.1656)</td>
<td>(0.1150)</td>
</tr>
<tr>
<td>$n$</td>
<td>-0.3501</td>
<td>-0.7115</td>
<td>-0.0698</td>
</tr>
<tr>
<td></td>
<td>(0.2511)</td>
<td>(0.3744)</td>
<td>(0.2307)</td>
</tr>
<tr>
<td>$w$</td>
<td>1.0048*</td>
<td>-0.1354</td>
<td>0.4558</td>
</tr>
<tr>
<td></td>
<td>(0.0846)</td>
<td>(0.0840)</td>
<td>(0.5980)</td>
</tr>
<tr>
<td>intercept</td>
<td>1.5902*</td>
<td>trend $\times 10^3$</td>
<td>0.4558</td>
</tr>
<tr>
<td></td>
<td>(0.7024)</td>
<td>(0.5980)</td>
<td></td>
</tr>
</tbody>
</table>

$= R^{-2} = 0.9998$

FUN (1.114) 1.32

NOR (2) 0.6432

HET (1,129) 0.1973

*absolute t statistic greater than 2.

Long Run Solution:

$w = 12.17 + 1.1185p + 1.0105y + 0.6586(t - 2.2322n + 0.0035t)$

(4.74) (0.677) (0.5491) (0.7488) (0.0047)

Notes. AUT, FUN and HET are F-type diagnostics testing respectively against first order residual autocorrelation, non-linearity and heteroskedasticity. NOR is the skewness–kurtosis diagnostic of Bera–Jarque, testing against non-normality.

there is more than one cointegrating vector. A maximum likelihood procedure for estimation and inference about all the cointegrating vectors is described in Johansen and Juselius (1990). Suppose $x_t$ is a $1 \times p$ vector. The VAR can be written:

$$\Delta x_t = \sum \delta_t \Delta x_{t-1} + \pi x_{t-1} + u_t$$  \hfill (27)

where $\pi$ is a $p \times p$ matrix. If $\pi$ is of full rank $p$, this implies that the vector process $x_t$ is stationary. If the rank of $\pi$ is zero, i.e. $\pi$ is the null matrix, then the level terms have no effect and a model in first differences is appropriate. If the rank of $\pi$ is $r$ between zero and $p$, then there are $p \times r$ matrices $\gamma$ and $\alpha$ such that

$$\pi = \gamma \alpha'$$  \hfill (28a)

i.e. there are $r$ cointegrating vectors

$$e_t = \alpha' x_t$$  \hfill (28b)

and (27) can be written in terms of the errors:

$$\Delta x_t = \sum \delta_t \Delta x_{t-1} + \gamma e_{t-1} + u_t$$  \hfill (28c)

with feedback coefficients or loadings $\gamma$. 

The estimation method first concentrates the likelihood to remove the effects of the right hand side differences by regressing $\Delta x_t$ and $x_{t-1}$ on $\Delta x_{t-1}$. These regressions give $p \times 1$ residual vectors $R_{0t}$, from the $\Delta x_t$ regression, and $R_{1t}$, from the $x_{t-1}$ regression; with Covariance Matrices:

$$S_{ij} = T^{-1} \sum_{t=1}^{T} R_{it}R_{jt}'; \quad i, j = 0, 1; \quad t = 1 \ldots T. \quad (29a)$$

Then the solution to the equation:

$$| \lambda S_{11} - S_{10}S_{00}^{-1}S_{01} | = 0,$$

(29b)
gives ranked eigenvalues $\lambda_1, \ldots, \lambda_p$; and a matrix of eigenvectors $V = (v_1, \ldots, v_p)$. The estimates of $\alpha$ that maximise the likelihood function under the hypothesis $x = \gamma \alpha'$ is then the eigen matrix $(v_1, v_r)$. These have to be normalised to be interpreted. Johansen also provides a test based on $\lambda_i$ for determining $r$, the number of cointegrating vectors. These tests use $-T \ln(1 - \lambda_i)$ and $-T \Sigma \ln(1 - \lambda_i)$, and suggest that there are at least two cointegrating vectors in our data, given by the first two columns of the eigen-matrix.

Table 4 gives the results from the Johansen procedure. These were obtained from the PCFIML module of PC-GIVE, which gives you the Johansen estimates automatically when you estimate a VAR (provided you told it that you were an expert). The first two Johansen cointegrating vectors are given in Table 5

| $\lambda_i$ | 0.213 | 0.187 | 0.119 | 0.046 | 0.002 |
| $-T \ln(1 - \lambda_i)$ | 31.38 | 27.14 | 16.60 | 6.18 | 0.22 |
| $-T \Sigma \ln(1 - \lambda_i)$ | 81.52 | 50.14 | 23.00 | 6.40 | 0.22 |

| $p$ | +12.48 | -15.00 | +14.70 | +10.89 | +0.39 |
| $y$ | 11.01 | -22.57 | +12.14 | -0.79 | +7.15 |
| $n$ | -37.22 | +28.89 | +16.01 | +5.96 | +21.22 |
| $l$ | +25.68 | +16.64 | -1.15 | -7.05 | -13.26 |
| $w$ | -10.25 | +16.79 | -11.91 | -8.03 | +1.10 |

Table 5. Alternative Long Run Estimates, for wage equation

| $p$ | $y$ | $l$ | $n$ | $\eta$ |
| 1.08 | 0.97 | -0.26 | -0.95 | -0.19 |
| 1.07 | 0.97 | -0.26 | -0.95 | -0.19 |
| 0.89 | 1.34 | -0.99 | -1.72 | -1.71 |
| 1.12 | 1.01 | +0.66 | -2.23 | -0.57 |
renormalised to make the coefficient of wages unity. With post-war data, using the Johansen procedure on \( w - p, y, l, u \) and \( h \) (the logarithm of weekly hours worked), Hall (1989) identified the cointegrating vector:

\[
(w - p) = 1.099 (y - l) - 0.562u - 0.49h
\]

where \( u \equiv n - l \). This is similar to our first vector, except for the effect of hours.

The various procedures give us a number of different estimates of the long run relationship between \( w \) and \( p, y, l \) and \( n \). Were target real wages determined by productivity with a unit elasticity (e.g. as the marginal product equation of a Cobb–Douglas), unemployment, and trend we would expect a long run relationship of the form:

\[
(w - p)^* = (y - l) - \psi u + \delta_0 + \delta_1 t;
\]  

(20a)

thus in the long run solutions:

\[
w = \beta_0 p + \beta_1 y + \beta_2 l + \beta_3 n + \beta_4 t
\]  

(30b)

the restrictions \( \beta_0 = 1, \beta_1 = 1; \) and \( 1 + \beta_2 + \beta_3 = 0 \); should be satisfied. Table 5 summarises the possible long-run estimates. These are obtained from: the Engle–Granger static levels regression; the regression of the share on the unemployment rate and a trend, which satisfies the restrictions above; the first two Johansen cointegrating vectors normalised on wages, and finally the long run solution from the dynamic Linear regression model. It is clear that, except for the second Johansen vector, \( \beta_0 \) and \( \beta_1 \) are not far from unity, and, except for the second Johansen and the DLR, \( \eta = 1 + \beta_2 + \beta_3 \) is not too far from zero. However, although economically plausible, these restrictions on the long-run solution would be rejected both for the VAR, \( F(3,119) = 3.45 \), and for the DLR, \( F(3,115) = 5.14 \). The reason for this seems to be that this measure of the share of wages shifted up to a new level after the second world war. If a dummy variable which takes the value unity after 1945 is introduced, the restrictions cannot be rejected at the 5% level, for the VAR \( F(3,119) = 1.92 \), for the DLR \( F(3,114) = 2.51 \).

Both the DLR and the VAR typically characterise the time series properties of the data by a highly parameterised model. In contrast, more theoretical versions of the ECM tend to imply strong restrictions which leave relatively few parameters to be estimated.

In conclusion, as our example above has shown, the range of statistical techniques utilized have not provided us with anything more than we would have got by taking the economically meaningful linear combinations of the variables and looking at their graphs. In addition, the earlier discussion emphasised that it may be misleading to impute structural interpretations to summaries of the statistical characteristics of the data, such as those provided by the long run solution of the DLR, the pattern of Granger Causality and the cointegrating vectors. The estimated coefficients are likely to be complicated mixtures of equilibrium, adjustment and expectations parameters and disentangling them.
poses substantial identification problems. This suggests that there are advantages in starting from more structural models.

4.2. Specification of structural error correction models

We next turn to such structural error correction models. The Phillips ECM for wage adjustment is given by equation (4). It postulates that wage inflation is equal to 'equilibrium' or 'target' wage inflation \( w^* \), minus three terms related to the lagged unemployment rate, its change, and the acceleration of current unemployment. The issue that needs to be resolved before one proceeds to discuss estimation of this version is what determines 'equilibrium' wage inflation.

Different authors take different views about the process of wage determination, and therefore about the underlying model of equilibrium wages. Phillips (1958) seems to think of equilibrium wages as being a function of prices. He seems to view the effect of unemployment as a manifestation of excess supply of labour that will only affect the adjustment process of wages, but not the equilibrium level of real wages. We can therefore assume for the Phillips model that equilibrium real wages are independent of the unemployment rate. Then, equilibrium wage inflation is determined by expected price inflation, and the rate of change of equilibrium real wages. The following simple model captures this.

\[
\Delta w_t^* = \omega_t + \rho_t^e
\]  

where \( \omega \) is the exogenous equilibrium real wage, and superscript \( e \) denotes the expectation of the relevant variable. We have deliberately used the expectation of the price level, in order to highlight the role of the modelling of expectations. Clearly, if workers had full information about current variables at the time when wages are set, then one could substitute the actual price level for the expectation. If they did not, then some process of expectations formation has to be assumed.

Applying the first difference operator to (31), and assuming that equilibrium real wages grow at the rate \( g \), we get,

\[
\Delta w_t^* = g + \Delta p_t^e
\]  

Substituting (32) in (4), we get the following expectations augmented version of the Phillips error correction model:

\[
\Delta w_t = g + \Delta p_t^e + k_d \Delta^2 u_t + k_p \Delta u_{t-1} + k_i u_{t-1}
\]  

We next move on to the Sargan specification. We shall utilize the version referred to in equation (6b). Sargan has a union wage setting model in mind, and for him equilibrium real wages for the union depend on the unemployment rate. Substituting for the 'equilibrium' real wage from the second equation in (6b) into the first, we get,

\[
\Delta w_t = \Delta p_t^e - \gamma u_{t-1} - \gamma (w - p - \omega)_{t-1}
\]  

Finally, the Hendry specification will be based on equation (8), substituting \( w_t \)
for $y_t$, and $w_t^*$ for $x_t$. This gives us,

$$\Delta w_t = \alpha + \beta \Delta w_t^* - \gamma (w - w^*)_{t-1} \quad (35')$$

Substituting the Sargan version of the equilibrium wage $w^*$ in (35'), we end up with,

$$\Delta w_t = (\alpha + \beta g) + \beta \Delta p_t^* - \beta \eta \Delta u_t^* - \gamma (w - p - \omega + \eta u)_t, \quad (35)$$

(35) will form the basis for our models of the ‘Hendry’ version of ECMs.

4.3. The treatment of expectations

Before one proceeds to estimate the three version of the ECM model (33), (34) and (35), a number of issues have to be resolved.

First and foremost is the treatment of expectations. The approach in the ECM literature has been to largely ignore problems having to do with expectations. In the context of our example, this amounts to replacing actual for expected price inflation in each of the three equations, and estimate the equations by a single equations method. With this treatment of expectations, the alternative ECM equations to be estimated take the form:

$$\Delta w_t = g + \Delta p_t + k_d \Delta^2 u_t + k_p \Delta u_{t-1} + k_i u_{t-1} \quad (33a)$$

$$\Delta w_t = \Delta p_t - \gamma \Delta u_t - \gamma (w - p - \omega)_{t-1} \quad (34a)$$

$$\Delta w_t = (\alpha + \beta g) + \beta \Delta p_t - \beta \eta \Delta u_t - \gamma (w - p - \omega + \eta u)_{t-1} \quad (35a)$$

This practice is widespread in the ECM literature, mainly because of its simplicity. However, it should not be adopted uncritically, especially before one thinks about the identification of the resulting equations.

As an illustration of this point consider a frequent complement to wage equations, which is none other than a price equation. Assume that prices are a markup on unit labour costs, and that productivity follows a random walk. Then, the price equation becomes,

$$\Delta p_t = \Delta w_t + v_t \quad (36)$$

where $v_t$ is the innovation to productivity. Any of the three ECM equations (33a), (34a) or (35a) is underidentified in the context of a system with (36). If the wage price system is viewed as a complete two equation model the order condition is not satisfied, while if they are viewed as part of a larger macromodel, the order condition may be satisfied, but the rank condition will not. Such considerations very seldom seem to find their way into applied research based on error correction models, because, despite the focus on weak exogeneity, the choice of theoretical parameters of interest is barely discussed. Yet, poor identification may seriously prejudice ones inferences.

To highlight one simple alternative to this practice of replacing expectations by current variables, consider an alternative assumption. This alternative is based on the hypothesis that wages are set for one year in advance. With this
assumption, expectations about current price inflation or the change in the unemployment rate have to be based on information available up to the end of the previous period \((t-1)\). This hypothesis is a version of the Gray (1976) and Fischer (1977) model of wage adjustment. Assume for simplicity that expectations are based on the univariate time series properties of the series in question. Since the price level has been shown to have been AR(2) with a unit root, and the unemployment rate has been shown to have been AR(2) but stationary (see Tables 1A and 1B), we shall make the following assumption about the process of expectations formation:

\[
\Delta p_t^e = E(\Delta p_t | I_{t-1}) = \phi_0 + \phi_1 \Delta p_{t-1} \tag{37}
\]

\[
\Delta u_t^e = E(\Delta u_t | I_{t-1}) = \psi_0 + \psi_1 \Delta u_{t-1} + \psi_2 u_{t-1} \tag{38}
\]

Our focus on univariate forecasting equations for these variables should not be much of a problem. Comparison of standard errors of the AR and VAR models in Tables 1A and 1B shows that the SER of the VAR is only marginally smaller than the AR for \(p\) and \(u\).

Substituting (37) and, where applicable, (38) in the three alternative error correction models (33), (34) and (35), we end up with the following models of wage inflation:

\[
\Delta w_t = (g + \phi_0) + \phi_1 \Delta p_{t-1} + k_d \Delta u_t + k_p \Delta u_{t-1} + k_\eta u_{t-1} \tag{33b}
\]

\[
\Delta w_t = \phi_0 + \phi_1 \Delta p_{t-1} - \gamma \eta u_{t-1} - \gamma (w - p - \omega)_{t-1} \tag{34b}
\]

\[
\Delta w_t = \xi + \beta \phi_1 \Delta p_{t-1} - \beta \eta \psi_1 \Delta u_{t-1} + (\beta \psi_2 - \gamma) \eta u_{t-1} - \gamma (w - p - \omega)_{t-1} \tag{35b}
\]

where \(\xi = \alpha + \beta g + \phi_0 - \beta \eta \psi_0\).

These equations are identified even in models with price equations of the form of (36). The timing assumption that wages are set before prices suffices in this case to ensure identification, as current prices are excluded from the wage equations.

The point we want to stress, and which is starkly illustrated by our example, is that the treatment of expectations has important implications for any econometric model, and that ECMs by themselves do not offer an easy escape from the problems that arise. Structural ECMs are at best models of short run adjustment, and as any other such model, there is a need for auxiliary assumptions about expectations formation. The nature of these assumptions may affect critically the empirical success or failure of the final econometric model.

4.4. Restrictions

The second problem before (33), (34) and (35) are estimated relates to the type of restrictions that they imply for unrestricted dynamic models of the same order. It is often argued that ECMs result in fairly unrestricted and flexible econometric models that are easy to estimate. We shall argue that if the formal
derivations are taken seriously, this is not necessarily the case. We shall briefly
deal with the Phillips model and then concentrate on the Sargan and Hendry
versions of the ECM.

The Phillips version clearly implies the restriction that both the dependent
variable and the variables that affect its equilibrium enter in differenced form.
What enters in a fairly unrestricted fashion is the variable that measures the
error, or disequilibrium, which in our example is unemployment.

Let us then concentrate on the Sargan and Hendry versions of ECMs. To see
the restrictions that they imply, consider first the general form of Hendry’s
model.

\[ \Delta y_t = \beta \Delta y_t^* - \gamma (y - y^*)_{t-1} + \epsilon_t \]  \( (39) \)

where \( \epsilon_t \) is a white noise residual, and \( y^* \) denotes the equilibrium value of the
dependent variable \( y \). Sargan’s model is a restricted version of (39), for which
\( \beta = 0 \).

The equilibrium value will in general be specified as,

\[ y^* = \alpha_0 + \alpha' x \]  \( (40) \)

where \( \alpha \) and \( x \) are \( k \times 1 \) vectors, the \( x_i \)s being the variables that determine the
equilibrium value of \( y \).

Substituting (35) in (34), we get,

\[ \Delta y_t = \gamma \alpha_0 + \beta \alpha' \Delta x_t - \gamma y_{t-1} + \gamma \alpha' x_{t-1} + \epsilon_t \]  \( (41) \)

(41) is clearly a restricted version of the first order dynamic linear model,

\[ \Delta y_t = a_0 + b_0' \Delta x_t + a_1 y_{t-1} + b_1 x_{t-1} + \nu_t \]  \( (41a) \)

where \( b_0, b_1 \) are \( k \times 1 \) vectors and \( \nu_t \) is white noise. The restrictions are given
by,

\[ \frac{b_{0i}}{b_{1i}} = \frac{b_{0j}}{b_{1j}} = \frac{\beta}{\gamma}, \quad \forall i, j = 1, 2, \ldots, k \]  \( (41b) \)

where \( b_{0i} \) and \( b_{1i} \) is the \( i \)th element of the \( b_0 \) and \( b_1 \) vector respectively.

Clearly there are \( k - 1 \) restrictions. The reason that such non-linear restrictions
do not appear in the Sargan model of error correction is that this model implies
the \( k \) restrictions \( b_{0i} = 0, \forall i = 1, 2, \ldots, k \). What the above suggests is that the
Hendry ECM implies one restriction less than the Sargan model, independently
of the number of \( x_i \)s. However, as opposed to the linear restrictions of the
Sargan model, these restrictions are non-linear. This is something that is very
seldom recognized, as most applications of ECMs seem to deal with the case of
long run linear homogeneity (e.g. consumption and income, wages and prices)
in which these restrictions become linear, as \( b_i = b_j \) for all \( i \) and \( j \).

In the context of the wage inflation model we have been considering, the
equivalent of the unrestricted first order dynamic linear regression model is given
by,

\[ \Delta w_t = b_0 + b_1 \Delta p_t^* + b_2 \Delta u_t^* + b_3 (w - \omega)_{t-1} + b_4 p_{t-1} + b_5 u_{t-1} \]  \( (42) \)
The theory assumes long-run homogeneity between wages and prices. Thus, it implies the linear restriction $b_3 = -b_4$. This particular linear restriction is not a restriction of the ECM, but a restriction emanating from the theory. It is worth emphasizing this, as in many places (e.g. Hendry, Pagan and Sargan 1984) it appears as if long run homogeneity restrictions is something special to the error correction model.

The Sargan ECM implies the additional linear restrictions that $b_1 = 1$ and $b_2 = 0$. These are indeed ECM restrictions. The first stems from the assumption that the adjustment takes place in expected real wages, and the second from the assumed nature of the adjustment process. If the assumption was that the adjustment was in nominal wages only, then $b_1$ would also have been zero. In any case, the Sargan model implies two linear restrictions for the general dynamic model of order one. A third restriction stems from the assumed long run homogeneity between wages and prices.

The Hendry ECM implies that $b_1/b_2 = b_4/b_5$. This is a non-linear restriction. Again, the additional linear restriction that $b_4 = 1$ stems from the assumed linear homogeneity between nominal variables.

Having cleared up these issues, we can now move on to estimation.

4.5. Estimation

Before we present estimates of the alternative models, there is a remaining loose end, namely the measurement of the equilibrium real wage $\omega$. To keep matters simple, and given the statistical investigation in section 5.1, we have assumed that this is proportional to average labour productivity. We have also allowed for a shift in the factor of proportionality after World War II, by including in all equations a dummy variable that takes the value 0 before 1946, and 1 after. This is intended to capture the rise in the share of labour in the post-war period suggested before.

Ordinary Least Squares estimates of equations (33a), (34a) and (35a) which are based on the assumption of contemporaneous information about prices and unemployment are in Table 6, in columns Ia, IIa and IIIa. Ordinary Least Squares estimates of the model according to which wages are set in advance, and are therefore based on lagged information are in columns Ib, IIb and IIIb.

As one would expect, the current information assumption results in a much better fit for all equations. This is hardly surprising, as these estimates are based on a richer conditioning set which includes innovations in prices and unemployment. However, the diagnostics of these equations are worse than for the models that condition on lagged variables only, and the responsiveness of wages to unemployment is very poorly estimated in all cases. This latter problem may be a reflection of under-identification, as in the example we offered above. The restriction of a unitary short run elasticity between wages and prices cannot be rejected by the relevant $t$-tests in any of the equations. If one were to rely on these estimates, one would have to reject the underlying theory, and conclude
Table 6. OLS Estimates of Alternative ECMs Wage Inflation in the UK: 1857–1987

Dependent Variable $\Delta w_t$

<table>
<thead>
<tr>
<th></th>
<th>(Ia)</th>
<th>(Ib)</th>
<th>(Ila)</th>
<th>(Ilb)</th>
<th>(IIa)</th>
<th>(IIIb)</th>
</tr>
</thead>
<tbody>
<tr>
<td>constant</td>
<td>0.012</td>
<td>0.023</td>
<td>0.520</td>
<td>1.735</td>
<td>0.518</td>
<td>1.466</td>
</tr>
<tr>
<td></td>
<td>(0.004)</td>
<td>(0.007)</td>
<td>(0.262)</td>
<td>(0.474)</td>
<td>(0.266)</td>
<td>(0.466)</td>
</tr>
<tr>
<td>$\Delta p_t$</td>
<td>1.069</td>
<td>1.016</td>
<td>1.016</td>
<td>0.040</td>
<td>0.041</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.038)</td>
<td></td>
<td></td>
<td>(0.038)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\Delta p_{t-1}$</td>
<td>0.882</td>
<td>0.751</td>
<td>0.730</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.067)</td>
<td>(0.074)</td>
<td>(0.072)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\Delta^2 u_t$</td>
<td>0.072</td>
<td>-0.914</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.109)</td>
<td>(0.190)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\Delta u_{t-1}$</td>
<td>-0.263</td>
<td>-1.293</td>
<td></td>
<td></td>
<td>-0.609</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.317)</td>
<td>(0.237)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$u_{t-1}$</td>
<td>0.012</td>
<td>-0.099</td>
<td>-0.050</td>
<td>-0.098</td>
<td>-0.096</td>
<td>-0.051</td>
</tr>
<tr>
<td></td>
<td>(0.061)</td>
<td>(0.106)</td>
<td>(0.105)</td>
<td>(0.106)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\Delta u_t$</td>
<td></td>
<td>-0.004</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.108)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$(w - P - \omega)_{t-1}$</td>
<td>-0.078</td>
<td>-0.263</td>
<td>-0.078</td>
<td>-0.222</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.040)</td>
<td>(0.072)</td>
<td>(0.041)</td>
<td>(0.071)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

R²         | 0.883 | 0.638 | 0.886 | 0.617 | 0.887 | 0.646  |
SER        | 0.023 | 0.040 | 0.022 | 0.041 | 0.022 | 0.040  |
DW         | 1.796 | 1.896 | 1.872 | 1.815 | 1.874 | 2.045  |

LM-tests

<table>
<thead>
<tr>
<th></th>
<th>AUT (1)</th>
<th>LIN (1)</th>
<th>HET (1)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1.295</td>
<td>6.706</td>
<td>8.385</td>
</tr>
<tr>
<td></td>
<td>(0.351)</td>
<td>(0.329)</td>
<td>(4.736)</td>
</tr>
<tr>
<td></td>
<td>0.565</td>
<td>6.182</td>
<td>8.749</td>
</tr>
<tr>
<td></td>
<td>(1.687)</td>
<td>(0.603)</td>
<td>(0.646)</td>
</tr>
<tr>
<td></td>
<td>0.551</td>
<td>6.596</td>
<td>8.763</td>
</tr>
<tr>
<td></td>
<td>(0.147)</td>
<td>(0.473)</td>
<td>(2.280)</td>
</tr>
</tbody>
</table>

Note: AUT, LIN and HET are LM-tests for first order autocorrelation, functional form and heteroskedasticity. They are asymptotically distributed as $\chi^2(1)$. $\omega$ has been approximated by the log of average labour productivity and a shift dummy after 1946.

that the response of wage inflation to unemployment is zero both in the short and the long run.

Estimation under the assumption that wage setters base their expectation on lagged inflation and unemployment (columns Ib, IIb and IIIb) result in parameters that are more in accord with the underlying theory. The fit of all the equations is worse, but this is a direct consequence of the fact that the conditioning set is smaller, since we are not conditioning on current innovations. The diagnostics show fewer signs of mis-specification, while the point estimates suggest negative short run elasticities of wage inflation to changes in unemployment. The effect of the level of unemployment on wage inflation is not statistically significant for any of the specifications, whereas the short run price elasticity is not significantly different from unity for the Mark I specification only. For the Mark II and III specifications it is significantly different from one at conventional significance levels.

To conclude from Table 6, the assumption that wage setters only use lagged
information appears to result in fewer mis-specification problems, and the estimates are consistent with prior theoretical expectations in all versions of ECM. Thus, these results provide an illustration of the importance of decisions about the treatment of expectations, as well as the possible dangers of not tackling the identification problem.

We next concentrate on the more theoretically satisfactory estimates, and address the question of whether the additional (overidentifying) restrictions implied by the Mark II (Sargan) and III (Hendry) version of ECMs are satisfied for wage inflation in the United Kingdom. These additional ECM restrictions are binding in this case because of the rational expectations hypothesis. We shall illustrate these restrictions for the Mark III ECM model, which is the more general one.

The Mark III model to be estimated and tested, is the following.

\[
\Delta w_t = \xi + \beta \phi_1 \Delta p_{t-1} - \beta \psi_1 \Delta u_{t-1} + (\beta \psi_2 - \gamma) \eta u_{t-1} - \gamma (w - p - \omega)_{t-1} + v_{1t} \\
\Delta p_t = \phi_0 + \phi_1 \Delta p_{t-1} + v_{2t} \\
\Delta u_t = \psi_0 + \psi_1 \Delta u_{t-1} + \psi_2 u_{t-1} + v_{3t}
\]

where the \( v \)'s are (cross-correlated) disturbances.

(37') and (38') are maintained hypotheses, as we concentrate on the structure of the wage equation. The three equation model above is a restricted version of a more general three equation model consisting of (37'), (38'), and an unrestricted wage equation of the form,

\[
\Delta w_t = a_0 + a_1 \Delta p_{t-1} + a_2 \Delta u_{t-1} + a_3 (w - \omega)_{t-1} + a_4 p_{t-1} + a_5 u_{t-1} + \epsilon_t
\]

The restrictions implied by the system of (35b'), (37') and (38') for this unrestricted system are,

\[
a_3 = -a_4 \\
a_5 = \frac{\psi_2}{\psi_1} a_3 \frac{\phi_1}{\phi_1}
\]

(43a) is the usual long-run linear homogeneity restriction (in this case between wages and prices), while (43b) are the non-linear ECM restrictions highlighted above, in conjunction with the cross equation rational expectations restrictions. Note that without these additional restrictions the ECM restrictions would not be binding in this case. Apart from the constant, (35b') has six parameters to be estimated, whereas (35'') has five parameters. Thus, without the cross equation (rational expectations) restrictions (35b') is underidentified. It is the combination of the ECM restriction with the rational expectations hypothesis that produces an identified model. In fact the full model is overidentified, as it implies the two restrictions (43a) and (43b).

Table 7 presents the structural estimates of the model. Estimation of the three equation system is by Iterative Seemingly Unrelated Regressions (Iterative
Table 7. Estimates of Structural Parameters
Sample: 1857–1987

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate 1</th>
<th>Estimate 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta$</td>
<td>1.029</td>
<td>1.000</td>
</tr>
<tr>
<td>$\gamma$</td>
<td>-0.098</td>
<td>-0.099</td>
</tr>
<tr>
<td>$\eta$</td>
<td>-2.306</td>
<td>-2.489</td>
</tr>
<tr>
<td>$\phi_1$</td>
<td>0.813</td>
<td>0.816</td>
</tr>
<tr>
<td>$\psi_1$</td>
<td>0.158</td>
<td>0.155</td>
</tr>
<tr>
<td>$\psi_2$</td>
<td>-0.112</td>
<td>-0.111</td>
</tr>
</tbody>
</table>

Likelihood Ratio Test 2.821 3.243

Note: Estimation is by Iterative Seemingly Unrelated Regressions. The likelihood ratio test in column (I) tests the restrictions (43a), (43b), while the likelihood ratio test in column (II) tests those two, plus the additional restriction that $\beta = 1$.

SURE). The estimates are in accordance with the theoretical priors, and the two restrictions implied by the model cannot be rejected by a likelihood ratio test. The estimate of $\beta$ is not statistically different from one, (see column II where this restriction is imposed) and the long-run responsiveness of real wages to unemployment is correctly signed and statistically significant at 10%. This is to be contrasted with the single equations estimates in Table 6 (IIIb), where one could not reject the hypothesis of a zero long run effect of the level of unemployment at this significance level. Furthermore, the non-linear overidentifying restrictions of this version of the model are not rejected at conventional significance levels.

It is worth noting that $\eta$, the structural estimate of the long run responsiveness of real wages to unemployment turns out to be of the same order of magnitude as in the Johansen–Juselius cointegrating vectors, and the long run solution of the unrestricted dynamic linear regression model in Table 5, but of a different order of magnitude than the Engle–Granger cointegrating vectors. It is also worth noting that since the estimate of $\beta$, the short run responsiveness of nominal wages to equilibrium nominal wages, is not statistically different from 1, the estimated ECM is observationally equivalent to an expectational model incorporating the Common Factor Restrictions.

To conclude, these results highlight the significance of imposing the non-linear overidentifying restrictions of Mark III of ECM, as well as the importance of assumptions about the information set on which the expectations of economic agents are assumed to be based. Inferences about the theoretical parameters of
interest may be affected in important ways by decisions concerning these aspects of error correction models.

5. Conclusions

As a particular parameterisation of the dynamic linear regression model (DLR) or vector autoregressions (VARs), error correction models (ECMs) are an effective way of characterising the dynamic multivariate interactions characteristic of economic data. In this use, they are atheoretical models in the sense of Cooley and Leroy (1985), and there is no particular problem in substituting one observationally equivalent form of the model for another. One implication of the irrelevance of parameterisation is that the only questions that can be asked of the model are those which have the same answer for all observationally equivalent versions of the model. For the purposes of data description, and often forecasting, this is perfectly adequate. For other purposes, this is not true.

In particular, theoretical considerations about forms of adjustment, expectations and equilibrium will determine which is the interesting economic parameterisation. In relatively few cases will this correspond to some unrestricted conditional distribution. The choice of parameters of interest is a product of prior theoretical specification. It will reflect considerations such as a belief that they are autonomous or deep, and thus likely to be stable; that they can be interpreted in terms of economic theory, and thus can be compared with prior expectations; and that they are useful, and thus can be used for policy analysis or other purposes.

To quote Frisch’s editorial in the first issue of Econometrica, ‘If we are not to get lost in the overwhelming, bewildering mass of statistical data that are now becoming available, we need the guidance and help of a powerful theoretical framework. Without this no significant interpretation and coordination of our observations will be possible.’ We would interpret ECMs as contributing to the powerful theoretical framework, rather than to the bewildering mass of statistical data. As an economic hypothesis about the nature of adjustment, they imply strong and testable restrictions on dynamic models. Although these restrictions have been widely ignored, they can make a significant contribution to the interpretation of the data as our examples illustrate.

Acknowledgements

We would like to acknowledge the comments of two anonymous referees and the editors of this Journal.

Notes

1. This list is not exhaustive. Yet another interpretation is provided by Winder and Palm (1989).
2. The former terminology is used in Hendry, Pagan and Sargan (1984), while the latter is used by Spanos (1986). In what follows we shall adopt the Spanos terminology.

3. It will only be unbiased if $\rho = 1$. The term bias is used rather loosely here, since the standard estimator of the long run coefficient does not possess any finite sample moments.

References

Perron, P. (1990), The great crash, the oil price shock and the unit root hypothesis, Econometrica, 57, 1361–1401.