# The Error Correction Model

## 1  Setting up the EC model

We start from a simple, proportional, long-run equilibrium relationship between two variables:

$$Y_t = KX_t$$

We might think of $Y$ as inventory and $X$ as sales, or $Y$ as consumption and $X$ as income, or whatever. Of course a fully specified equilibrium model may well include more variables, and the equilibrium relationship need not be one of direct proportionality, but let's keep it simple.

The relationship above can be written in log form as

$$y_t = k + x_t \tag{1}$$

where we follow the convention of letting a lower-case letter designate the natural log of the variable represented by the corresponding upper case letter. (Taking logs reduces the multiplicative relationship to an additive one, which is a helpful mathematical simplification.)

Now let's write down a general dynamic relationship between $y$ and $x$:

$$y_t = \beta_0 + \beta_1 x_t + \beta_2 x_{t-1} + \alpha_1 y_{t-1} + u_t \tag{2}$$

By including lagged values of both $x$ and $y$ this specification allows for a wide variety of dynamic patterns in the data.

We now ask: Under what conditions is the generic dynamic equation (2) consistent with the long-run equilibrium relationship (1)? To assess this, we "zero out" the factors that could cause divergence from equilibrium, namely changes in $x_t$ and stochastic fluctuations, $u_t$. That is, we set $y_t = y^*$ and $x_t = x^*$ for all $t$, and set $u_t = 0$. Thus we get

$$
\begin{aligned}
y^* &= \beta_0 + \beta_1 x^* + \beta_2 x^* + \alpha_1 y^* \\
(1 - \alpha_1)y^* &= \beta_0 + (\beta_1 + \beta_2)x^* \\
y^* &= \frac{\beta_0}{1 - \alpha_1} + \frac{\beta_1 + \beta_2}{1 - \alpha_1}x^*
\end{aligned}
$$

If the above corresponds with equation (1) we have

$$
\begin{aligned}
\frac{\beta_0}{1 - \alpha_1} &= k \\
\frac{\beta_1 + \beta_2}{1 - \alpha_1} &= 1
\end{aligned}
$$

Suppose this is the case. The second relationship above means that $\beta_1 + \beta_2 = 1 - \alpha_1$. Let $\gamma$ denote the common value of these two terms. Then $\beta_2$ can be written as $\gamma - \beta_1$ and $\alpha_1$ can be written as $1 - \gamma$. Therefore equation (2) becomes

$$
\begin{aligned}
y_t &= \beta_0 + \beta_1 x_t + (\gamma - \beta_1)x_{t-1} + (1 - \gamma)y_{t-1} + u_t \\
y_t &= \beta_0 + \beta_1 x_t - \beta_1 x_{t-1} + \gamma x_{t-1} - \gamma y_{t-1} + y_{t-1} + u_t \\
y_t - y_{t-1} &= \beta_0 + \beta_1(x_t - x_{t-1}) + \gamma(x_{t-1} - y_{t-1}) + u_t
\end{aligned}
$$

So, finally,

$$\Delta y_t = \beta_0 + \beta_1 \Delta x_t + \gamma(x_{t-1} - y_{t-1}) + u_t \tag{3}$$

where $\Delta x_t \equiv x_t - x_{t-1}$. This is the characteristic "error correction" specification, where the change in one variable is related to the change in another variable, as well as the gap between the variables in the previous period.

## 2    Illustration: consumption function

To illustrate, let's take the data on U.S. per capita disposable income (Yt) and consumption expenditure (Ct), annual 1959–1994, from the Ramanathan data files (`data3-6`). We begin by generating the logs of Ct and Yt and the changes in the logs of these variables. The gretl "script" commands are as follows:

```
logs Ct Yt
ldiff Ct Yt
```

These commands generate the new variables l_Ct, l_Yt (logs) and ld_Ct, ld_Yt (the log-differences or $\Delta$s). We then create a variable representing the gap between log income and log consumption:

```
genr gap = l_Yt - l_Ct
genr gap_1 = gap(-1)
```

(The (-1) calls for the first lag to be used). We now specify a regression, to be estimated via OLS, corresponding to equation (3) above:

```
ols ld_Ct const ld_Yt gap_1
```

The results are shown below:

### OLS estimates using the 35 observations 1960–1994
Dependent variable: ld_Ct

| Variable | Coefficient | Std. Error | $t$-statistic | p-value |
|---|---|---|---|---|
| const | −0.0172635 | 0.0129678 | −1.3313 | 0.1925 |
| ld_Yt | 0.819320 | 0.0934824 | 8.7644 | 0.0000 |
| gap_1 | 0.214349 | 0.124736 | 1.7184 | 0.0954 |

| | |
|---|---|
| Sum of squared residuals | 0.00328885 |
| Standard error of residuals ($\hat{\sigma}$) | 0.0101379 |
| Adjusted $\bar{R}^2$ | 0.698098 |
| $F(2, 32)$ | 40.3096 |
| Durbin–Watson statistic | 1.84671 |

Note that the model does not seem to suffer from autocorrelation (DW = 1.85). The coefficient on the lagged gap, which corresponds to $\gamma$ in equation (3), does not appear highly significant, but on a *one*-tailed test it is significant at the 5 percent level. And a one-tailed test is appropriate here: on theoretical grounds we expect a positive coefficient, and we can run $H_0$: $\gamma \leq 0$ versus $H_1$: $\gamma > 0$.

Why do we expect a positive value for $\gamma$, if the error-correction model is appropriate? Let's go back to equation (3). The idea is that, ceteris paribus, the dependent variable should converge towards its equilibrium level. Now "ceteris paribus" can be taken to mean: barring changes in $x$ and other disturbances ($u_t$). If we set $\Delta x_t$ and $u_t$ to zero, equation (3) then becomes

$$\Delta y_t = \beta_0 + \gamma(x_{t-1} - y_{t-1}) = \gamma(\frac{\beta_0}{\gamma} + x_{t-1} - y_{t-1}) \tag{4}$$

But notice that $\beta_0/\gamma$ is, by definition, $\beta_0/(1 - \alpha_1)$, which in turn corresponds to $k$ in equation (1)—see the figuring on page 1. So (4) is equivalent to

$$\Delta y_t = \gamma(k + x_{t-1} - y_{t-1})$$

and $k + x_{t-1}$ is nothing other than the equilibrium value of $y$ in period $t-1$, according to equation (1). Thus, suppose we had $k + x_{t-1} - y_{t-1} > 0$: this would mean that last period the *equilibrium* value of $y$, namely $k + x_{t-1}$, exceeded its *actual* value, or actual $y$ fell short of equilibrium. If "error correction" is going on, what should happen? Well, $y$ should increase (i.e. $\Delta y_t > 0$), heading towards equilibrium. In other words, the coefficient $\gamma$ should be positive.[1] It's like this:

change in $y$ = positive coeff. $\times$ (degree to which $y$ fell short of equilibrium last period)

---

[1] If we had specified the "gap" the other way round, as consumption minus income, we would then have expected a negative value for $\gamma$.

*Factor of proportionality*

Given the above regression results, how do we "retrieve" an estimate of the factor of proportionality, $K$, in the original equilibrium relationship? We've just noted that $k$ corresponds to $\beta_0/\gamma$. In terms of the consumption estimates reported above, this means the constant divided by the coefficient on the lagged `gap`, or $-0.017/0.214$. That will give the estimated value of $k$, but that's the natural log of what we want, so we have to exponentiate (take the anti-log). The best way to do the calculation is to let gretl take care of it, based on its internal representation of the coefficient values.

```
genr prop = exp($coeff(const)/$coeff(gap_1))
print prop
   prop = 0.922619
```

That is, according to these estimates, the implied long-run equilibrium has people spending 92% of their disposable income. And if consumer spending diverges from this equilibrium relationship with income, then, due to the positive error-correction coefficient, there will be a tendency for spending to adjust towards the target value.

*General comments*

The results, then, are "sensible," on the face of it, but the model is not without problems. Note that both the constant and the disequilibrium adjustment coefficient, $\gamma$, are not estimated with much precision: the standard errors are rather large in relation to the coefficient estimates. The point estimate for $\gamma$ seems low: it implies that only about 21 percent of last year's disequilibrium is made up in the current year, a sluggish adjustment. A 95 percent confidence interval for this parameter would be about $0.21 \pm 2(0.12)$ or $-0.03$ to $0.44$, which is uncomfortably wide. It may well be that the model is mis-specified, perhaps because it omits other relevant variables or because the relationship between income and consumption has not in fact remained constant over the 36-year span of the data.

*Further interpretation*

Nonetheless, continuing to take the estimates at face value, consider what we can read from the coefficient on the change in the log of income, `ld_Yt`, which corresponds to the $\beta_1$ of equation (3). The point estimate is 0.819. What is this saying? (Notice that it is *not* our estimate of the equilibrium ratio of consumption to income.)

Let's go back to equation (3) once more, but this time run the thought experiment of setting the previous period's disequilibrium (and $u_t$) to zero. We're now asking, how would consumption behave if there were no adjustment to be made on account of a previous disequilibrium, and no random disturbance? "No previous disequilibrium" means that $y_{t-1} = k + x_{t-1}$. Then equation (3) becomes

$$
\begin{aligned}
\Delta y_t &= \beta_0 + \beta_1 \Delta x_t + \gamma(x_{t-1} - y_{t-1}) \\
\Delta y_t &= \beta_0 + \beta_1 \Delta x_t + \gamma[x_{t-1} - (k + x_{t-1})] \\
\Delta y_t &= \beta_0 + \beta_1 \Delta x_t - \gamma k \\
\Delta y_t &= \beta_0 + \beta_1 \Delta x_t - \gamma \frac{\beta_0}{\gamma} \\
\Delta y_t &= \beta_1 \Delta x_t
\end{aligned}
$$

Equation (1) implies that if equilibrium is to be maintained, $\Delta y_t$ should equal $\Delta x_t$ (i.e. the percentage change in consumption should equal the percentage change in income). But we are estimating that the proportional change in consumption is only 82 percent of the proportional change in income, absent any previous disequilibrium to correct. This is not actually a contradiction, but it's saying that when income rises, consumption tends to lag behind, creating a disequilibrium (which will in turn call for adjustment of consumption in subsequent periods). In other words, it's saying that changes in income are themselves a source of disequilibrium in this model (and not just the random disturbances represented by $u_t$).

*Error correction versus general dynamic model*

One further question. We said above (see page 1) that the error correction model represents a restriction on the general dynamic model given in equation (2). The restriction takes the form of the requirement that $\beta_1 + \beta_2 = 1 - \alpha_1$. This is a linear restriction, and it is testable, so let's test it. We have already estimated the restricted model (i.e. the error correction model) and its SSR is reported above. We need to estimate the unrestricted model—corresponding to equation (2)—and find its SSR, then form an *F*-statistic in the usual way:

$$F(1, \text{dfu}) = \frac{(\text{SSRr} - \text{SSRu})/1}{\text{SSRu}/\text{dfu}}$$

The numerator df = 1 because there is only one restriction. The denominator degrees of freedom (dfu) = 31 since there are 35 observations (after accounting for the lagged terms) and the unrestricted model involves estimating 4 parameters. It turns out that SSRu = 0.00319503 and $F(1, 31) = 0.910265$ with a p-value of 0.3474. We fail to reject $H_0 : \beta_1 + \beta_2 = 1 - \alpha_1$, and conclude that the error correction model is an acceptable restriction on the more general equation (2).

Here's the complete gretl program to generate the output discussed above.

```
open data3-6
logs Ct Yt
ldiff Ct Yt
genr gap = l_Yt - l_Ct
genr gap_1 = gap(-1)
# estimate error correction model
ols ld_Ct const ld_Yt gap_1
# retrieve factor of proportionality
genr prop = exp($coeff(const)/$coeff(gap_1))
print prop
genr SSRr = $ess
# now for the unrestricted model
lags l_Ct l_Yt
ols l_Ct const l_Yt l_Yt_1 l_Ct_1
genr SSRu = $ess
genr dfu = $df
genr Ftest = (SSRr-SSRu)*dfu/SSRu
pvalue F 1 dfu Ftest
```

## 3 Error correction in the stock market?

In view of goings-on in the stock market over recent years it might be interesting to inquire whether the behavior of the market (as represented, say, by the Dow–Jones Industrial Average) can be modeled as a case of "long-run equilibrium plus error correction".

What might be a plausible model for long-run equilibrium in this case? Well, corporate stocks ultimately derive their value from the fact that they are claims on the profits made by firms. Thus it seems reasonable to suppose that stock prices should reflect the present discounted value of the future (expected) stream of corporate profit. A simple (perhaps too simple) proxy for this would be the *current* level of (after-tax) corporate profits, divided by a long-term interest rate such as the rate on 10-year Treasury bonds. This would be the "correct" figure if people expected the current level of corporate profits to persist into the indefinite future: the present value of an asset which promises to pay a given sum $F$ each year, forever, is $F/r$ where $r$ denotes the appropriate discount rate.

My hypothesis, then, is that in the long run the Dow (as an aggregate measure of the value of the stock market) should be proportional to after-tax corporate profits divided by the long-term interest rate. We want to allow the possibility that the Dow is not equal to this long-run equilibrium value at

4

all times, but that if it diverges from this value the "error" will tend to be corrected over time. The simplest model that captures this is equation (3), reproduced below.

$$\Delta y_t = \beta_0 + \beta_1 \Delta x_t + \gamma(x_{t-1} - y_{t-1}) + u_t$$

Here $y_t$ represents the log of the Dow, and $x_t$ represents the log of the present value of corporate profits as described above.

The analysis was conducted on a quarterly basis, with data obtained from the Federal Reserve Bank of St Louis (corporate profits), the Federal Reserve Board of Governors (10-year Treasury bond rate) and `economagic.com` (the closing value of the Dow). The results are shown below (where `ld_dj` denotes the log-difference of the Dow–Jones average, `ld_cprof` denotes the log-difference of discounted corporate profits, and `djgap_1` denotes the lagged value of the difference between the log of discounted profit and the log of the Dow).

**OLS estimates using the 220 observations 1953:3–2008:2**
Dependent variable: ld_dj

|  | Coefficient | Std. Error | $t$-ratio | p-value |
|---|---|---|---|---|
| const | −0.00549884 | 0.00689003 | −0.7981 | 0.4257 |
| ld_cprof | 0.190529 | 0.0482375 | 3.9498 | 0.0001 |
| djgap_1 | 0.0538393 | 0.0152960 | 3.5198 | 0.0005 |

| | |
|---|---|
| Sum of squared residuals | 0.746783 |
| Standard error of the regression ($\hat{\sigma}$) | 0.0586634 |
| Unadjusted $R^2$ | 0.111245 |
| Adjusted $\bar{R}^2$ | 0.103054 |
| $F(2, 217)$ | 13.5809 |
| Durbin–Watson statistic | 1.68157 |

LM test for autocorrelation up to order 4 –
 Null hypothesis: no autocorrelation
 Test statistic: LMF = 1.43485
 with p-value = $P(F(4, 213) > 1.43485) = 0.223534$

The error-correction parameter (i.e. the coefficient on `djgap_1`) has the expected positive sign. The DW statistic does not suggest that autocorrelation is a problem; neither does the LMF test statistic for autocorrelation up to the fourth order.

As in the discussion of the consumption function above, we can retrieve an estimate of the factor of proportionality in the hypothesized long-run equilibrium relationship by dividing the constant by the `djgap_1` coefficient and exponentiating. This gives a value of 0.9029. Thus we estimate that the equilibrium value of the Dow index is about 90 percent of the present value of the corporate profit stream.

The residuals from the regression above are shown in Figure 1. The stock-market crash of 1987 stands out as a large negative residual: we infer that, so far as our model is concerned, this crash was not a case of "error correction" (otherwise the model would have predicted it) but rather an "error".

Figure 2 shows the actual value of the Dow alongside the equilibrium value implied by the model above (that is, 0.9029 times the present value of the corporate profit stream). Interestingly, the run up in the Dow over the 1990s appears not as a movement above equilibrium, but rather as a "catching up" with an equilibrium value that exceeded the actual value for much of the period. (I'm not sure I'd want to place too much stress on this result without further testing of the adequacy of the model.)
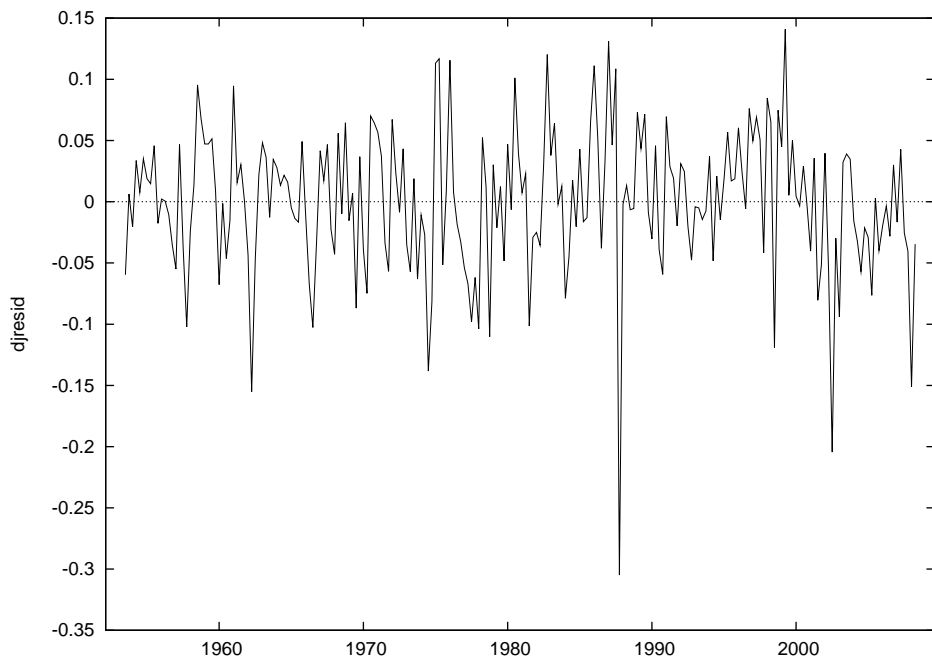
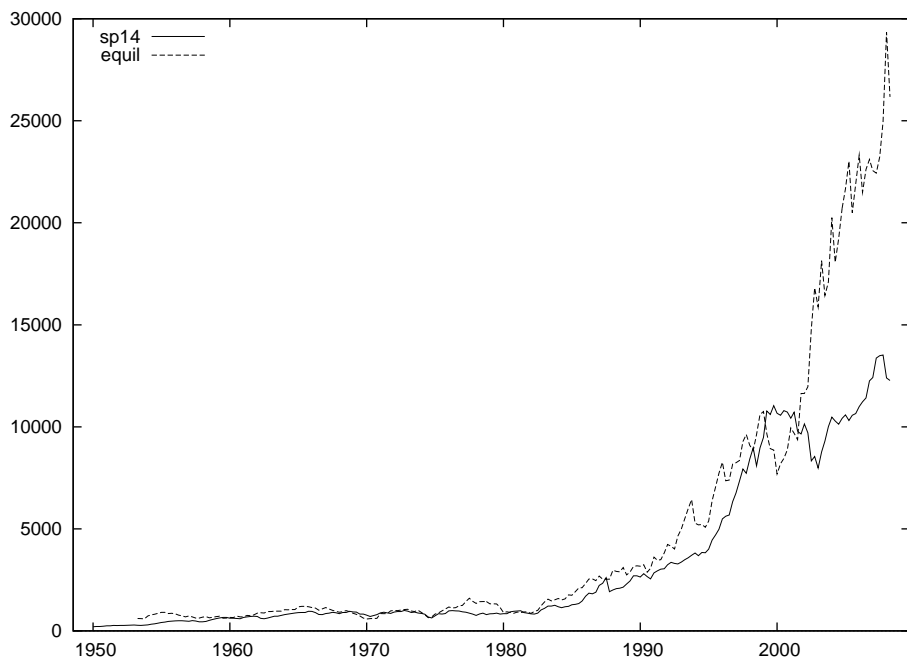Figure 1: Residuals from Dow error-correction regression



Figure 2: Actual and estimated equilibrium values of the Dow