An Introduction to Error Correction Models (ECMs)

Outline

- What is an ECM
- Stationary and non-stationary data
- Estimating ECMs
- ECMs and price transmission analysis

Motivation

- The ECM is a type of time series model with which we can estimate the speed at which two or more time series return to their common equilibrium following a shock that disturbs this equilibrium
- The ECM can directly estimate both long-run and short-run effects of one time series on another
- Thus, ECMs introduce an element of theory to time series analysis, which is often criticized as being theory-free

The basic form of the ECM

- $\Delta Y_t = \delta + \gamma \Delta X_{t-1} + \alpha ECT_{t-1} + \varepsilon_t$

where ECT is the so-called „error correction term“ which measures deviations from the long-run equilibrium between Y and X

- ECMs can be estimated with OLS, although other, more sophisticated methods have been developed (e.g. Johansen‘s Maximum Likelihood approach)

Stationary and non-stationary data

- ECMs are most often estimated using non-stationary time series
- A stationary time series has a finite mean and variance that do not vary with time*
- $Y_t = \rho Y_{t-1} + \varepsilon_t$
  - If $0 < |\rho| < 1$, Y is stationary, otherwise not
  - If $|\rho| > 1$, then Y „explodes“ – not very plausible in economic settings
  - But what if $\rho = 1$?

* This is the definition of weak stationarity. Strong stationarity requires that higher order moments also be constant.

The case of $0 < |\rho| < 1$

$Y_t = \rho Y_{t-1} + \varepsilon_t$

$Y_1 = \rho Y_0 + \varepsilon_1$
$Y_2 = \rho Y_1 + \varepsilon_2$

Therefore
$Y_2 = \rho(\rho Y_0 + \varepsilon_1) + \varepsilon_2 = \rho^2 Y_0 + \rho\varepsilon_1 + \varepsilon_2$

$Y_3 = \rho Y_2 + \varepsilon_3$

Therefore
$Y_3 = \rho(\rho(\rho Y_0 + \varepsilon_1) + \varepsilon_2) + \varepsilon_3 = \rho^3 Y_0 + \rho^2\varepsilon_1 + \rho\varepsilon_2 + \varepsilon_3$

Ultimately
$Y_t = \rho^t Y_0 + \rho^{t-1}\varepsilon_1 + \rho^{t-2}\varepsilon_2 + \ldots + \rho^{t-t+1}\varepsilon_{t-1} + \rho^{t-t}\,\varepsilon_t$
$Y_t = \rho^t Y_0 + \sum_{i=1}^{t} \rho^{t-i}\varepsilon_i$

The case of $\rho = 1$

$Y_t = Y_{t-1} + \varepsilon_t$

$Y_1 = Y_0 + \varepsilon_1$
$Y_2 = Y_1 + \varepsilon_2$

Therefore
$Y_2 = (Y_0 + \varepsilon_1) + \varepsilon_2 = Y_0 + \varepsilon_1 + \varepsilon_2$

$Y_3 = Y_2 + \varepsilon_3$

Therefore
$Y_3 = (Y_0 + \varepsilon_1 + \varepsilon_2) + \varepsilon_3 = Y_0 + \varepsilon_1 + \varepsilon_2 + \varepsilon_3$

Ultimately
$Y_t = Y_0 + \varepsilon_1 + \varepsilon_2 + \ldots + \varepsilon_{t-1} + \varepsilon_t$
$Y_t = Y_0 + \sum_{i=1}^{t} \varepsilon_i$

Integrated time series

- When $\rho = 1$ we refer to the series as „integrated“
- Recall: $Y_t = Y_0 + \sum_{i=1}^{t} \varepsilon_i$
- As the time interval t becomes very short (i.e. we approach continuous time), this approaches an integral – therefore „integrated“
- An integrated time series is one type of non-stationary time series
- A time series is integrated of order *d* – referred to as I(*d*) – if it becomes stationary after being differenced *d* times
- If $Y_t = Y_0 + \sum_{i=1}^{t} \varepsilon_i$, then $\Delta Y_t = (Y_t - Y_{t-1}) = \varepsilon_t$, which is stationary
- So the time series $Y_t = \rho Y_{t-1} + \varepsilon_t$ with $\rho = 1$ is I(1)

Example with GRETL

- Generate a random normal series $RanY$
- Generate a random normal series $RanX$
- Look at graphs of $RanY$ and $RanX$, what do you see?
- Now generate the following cumulative sums using the „cum“ command (i.e. with $\rho = 1$)

$$Y_t = Y_{t-1} + RanY_t$$
$$X_t = X_{t-1} + RanX_t$$

- Look at graphs of Y and X, what do you see?
- Generate the first differences of Y and X, what is the result?

Why does this matter? (I)

- There are several important differences between I(0) (stationary) and I(1) (integrated) time series
- I(0) series „forget“ – they have finite memory, the effect of a shock is transitory
  - Recall: $Y_t = \rho^t Y_0 + \sum_{i=1}^{t} \rho^{t-i} \varepsilon_i$
  - Only the most recent shocks affect Y_t, earlier shocks fade with increasing exponents on ρ
- I(1) series never forget – they have infinite memory, the effect of a shock is permanent
  - Recall: $Y_t = Y_0 + \sum_{i=1}^{t} \varepsilon_i$
  - Y_t therefore equals its initial value, plus the sum of all shocks that have occurred since $t=0$.

Why does this matter? (II)

- Furthermore, many of the time series that we work with in empirical economic research (for example, many prices) behave as if they were I(1)
- Go back to the data you generated with GRETL and see for yourself

Why does this matter? (III)

- Most important, although the theory was not developed until the 1970s and 1980s, it has long been recognized that regression with non-stationary data can lead to „spurious regression“ – i.e. nonsense results
- Go back to GRETL:
  - Run the regression $RanY_t = \beta_0 + \beta_1 RanX_t$ – what do you expect, what do you get?
  - Run the regression $Y_t = \beta_0 + \beta_1 X_t$ – what do you expect, what do you get? Surprised?
  - With I(1) data, the likelihood of making a Type-II error (rejecting the true null hypothesis that there is no relationship between Y and X) increases dramatically

Integration and cointegration

- What we need is a means of distinguishing between spurious regression and a true relationship
- One formerly popular solution was to analyze I(1) time series in differenced form: $\Delta Y_t = \beta_0 + \beta_1 \Delta X_t$
- However, this assumes that X only has short-run effects on Y, and ignores the possibility that they share a long-run relationship
- If they do share a long-run relationship, estimating an equation in first differences is inefficient because it fails to account for information about this relationship
- When two integrated variables share a long-run equilibrium relationship, they are referred to as cointegrated

Integration, cointegration and random walks (I)

- Imagine a drunk staggering out of a pub. Each step is a random draw, and his position after t such steps is the sum of these random draws: $Y_t = \sum_{i=1}^{t} \varepsilon_i$
- This is sometimes referred to as a random walk – a random walk is an example of an I(1) or integrated, non-stationary time series
- Imagine a second drunk who leaves the pub at the same time: $X_t = \sum_{i=1}^{t} \varepsilon_i$
- If we regress Y on X (or X on Y) we run the danger of spurious regression. The paths taken by the two drunks are random and independent, but there is a high probability that we will reject the true null hypothesis that they are

Integration, cointegration and random walks (II)

- Now imagine that the first drunk is accompanied by his dog on a leash
- The leash establishes a „long-run equilibrium" between the drunk and his dog
- The dog's path (Z_t) is also a random walk, but it is driven by the first drunk's path: $Z_t = f(Y_t)$
- The drunk and his dog can become more or less separated temporarily
- But these deviations will be „corrected" by the leash
- When we regress the position of the dog on the position of the drunk, the result will not be spurious

Integration, cointegration and random walks (III)

- Q: How do we distinguish between the two cases?
- A: Using the concept of cointegration.
- We say that two time series are cointegrated if:
  - they are both integrated of the same order, for example if both are I(1)
  - there is a linear combination of the two series that is I(0), i.e. stationary
- Cointegrated series are non-stationary, but they share a common non-stationarity because they are linked by a long-run relationship that does not permit them to drift too far apart
- The Granger Representation Theorem states that if two variables are cointegrated, then they must be linked by an ECM

Deriving the ECM (I)

Begin with a simple distributed lag model

$$Y_t = \phi_0 + \phi_1 X_t + \phi_2 X_{t-1} + \phi_3 Y_{t-1} + \varepsilon_t$$

Since, in long run equilibrium, $X_t = X_{t-1} = X^*$ and $Y_t = Y_{t-1} = Y^*$

$$Y^* = \phi_0 + \phi_1 X^* + \phi_2 X^* + \phi_3 Y^* + \varepsilon_t$$

$$Y^*(1 - \phi_3) = \phi_0 + X^*(\phi_1 + \phi_2) + \varepsilon_t$$

$$Y^* = \frac{\phi_0}{(1-\phi_3)} + \frac{(\phi_1+\phi_2)}{(1-\phi_3)} X^* + \frac{\varepsilon_t}{(1-\phi_3)}$$

Thus, the long-run equilibrium relationship between x and y is given by:

$$Y^* = \beta_0 + \beta_1 X^* + \frac{\varepsilon_t}{(1-\phi_3)},$$

where $\beta_0 = \frac{\phi_0}{(1-\phi_3)}$ and $\beta_1 \frac{(\phi_1+\phi_2)}{(1-\phi_3)}$.

Deriving the ECM (II)

Now start with the same distributed lag model

$$Y_t = \phi_0 + \phi_1 X_t + \phi_2 X_{t-1} + \phi_3 Y_{t-1} + \varepsilon_t$$

$$Y_t - Y_{t-1} = \phi_0 + \phi_1 X_t - \phi_1 X_{t-1} + \phi_1 X_{t-1} + \phi_2 X_{t-1} + \phi_3 Y_{t-1} - Y_{t-1} + \varepsilon_t$$

$$\Delta Y_t = \phi_0 + \phi_1 (X_t - X_{t-1}) + (\phi_1 + \phi_2) X_{t-1} + (\phi_3 - 1) Y_{t-1} + \varepsilon_t$$

$$\Delta Y_t = \phi_1 \Delta X_t + (\phi_3 - 1)\left[Y_{t-1} - \frac{(\phi_1 + \phi_2)}{(1 - \phi_3)} X_{t-1} - \frac{\phi_0}{(1 - \phi_3)}\right] + \varepsilon_t$$

This is the basic form of the error correction model. Recall (from the previous slide) that in the long run:

$Y^* = \beta_0 + \beta_1 X^*$, with $\beta_0 = \frac{\phi_0}{(1-\phi_3)}$ and $\beta_1 \frac{(\phi_1+\phi_2)}{(1-\phi_3)}$

Therefore, the term $\left[Y_{t-1} - \frac{(\phi_1+\phi_2)}{(1-\phi_3)} X_{t-1} - \frac{\phi_0}{(1-\phi_3)}\right]$ is equivalent to $Y_{t-1} - \beta_1 X_{t-1} - \beta_0$

Hence, the ECM relates changes in one variable to changes in the other, <u>and</u> to deviations from the long run equilibrium relationship between both variables.

Deriving the ECM (III)

$\Delta Y_t = \delta + \gamma \Delta X_{t-1} + \alpha ECT_{t-1} + \varepsilon_t$
where $ECT_{t-1} = Y_{t-1} - \beta_0 - \beta_1 X_{t-1}$, the deviation from the long-run equilibrium relationship in the previous period

- Note that the ECM is balanced in the sense that all variables in it are I(0)
- Therefore, no danger of spurious regression
- However, all variables in the ECM are I(0) only if Y and X are cointegrated, so that the linear combination $ECT = Y - \beta_0 - \beta_1 X$ is I(0)
- This is essentially the Granger Representation Theorem again: if Y and X are cointegrated, there must be a valid ECM that describes their behavior $\Leftrightarrow$ if there is a valid ECM for Y and X, they must be cointegrated

The vector error correction model (VECM)

- What if the drunk's dog is big and also determines their joint path?
- In other words, why just one equation for Y as a function of X?
- Solution: the VECM

$$\begin{bmatrix} \Delta Y_t \\ \Delta X_t \end{bmatrix} = \begin{bmatrix} \varphi_1 \\ \varphi_2 \end{bmatrix} + \begin{bmatrix} \alpha_1 \\ \alpha_2 \end{bmatrix} [Y_{t-1} - \beta_0 - \beta_1 X_{t-1}] + \sum_{i=1}^{k} \begin{bmatrix} \delta_{1j} & \rho_{1j} \\ \delta_{2j} & \rho_{2j} \end{bmatrix} \begin{bmatrix} \Delta Y_{t-i} \\ \Delta X_{t-i} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{bmatrix}$$

- Or, in price transmission analysis

$$\begin{bmatrix} \Delta p_t^A \\ \Delta p_t^B \end{bmatrix} = \begin{bmatrix} \varphi_1 \\ \varphi_2 \end{bmatrix} + \begin{bmatrix} \alpha_1 \\ \alpha_2 \end{bmatrix} [p_{t-1}^A - \beta_0 - \beta_1 p_{t-1}^B] + \sum_{i=1}^{k} \begin{bmatrix} \delta_{1j} & \rho_{1j} \\ \delta_{2j} & \rho_{2j} \end{bmatrix} \begin{bmatrix} \Delta p_{t-i}^A \\ \Delta p_{t-i}^B \end{bmatrix} + \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{bmatrix}$$

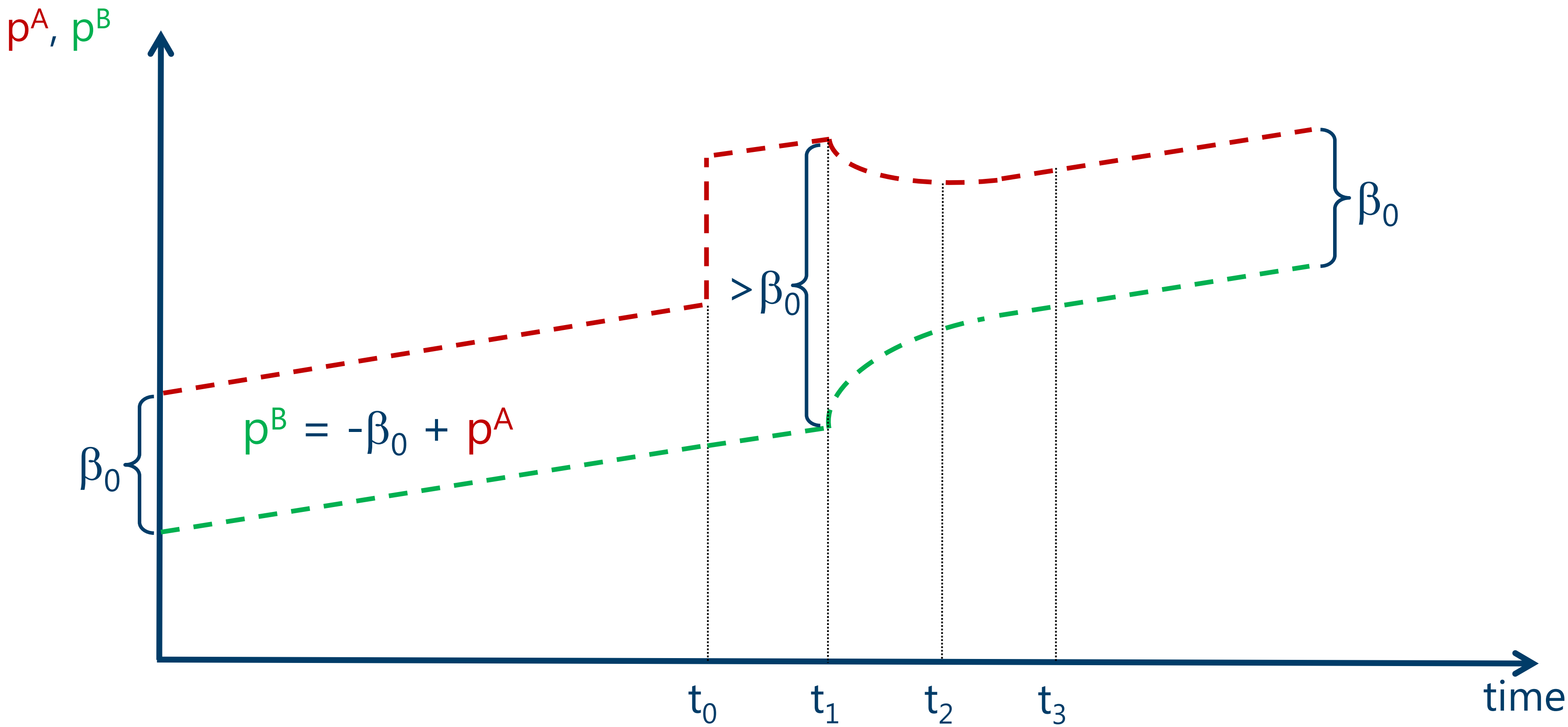
Adjustment parameters

Coefficient of long-run price transmission

Visualising cointegration in a VECM (I)

$$\begin{bmatrix} \Delta p_t^A \\ \Delta p_t^B \end{bmatrix} = \begin{bmatrix} \varphi_1 \\ \varphi_2 \end{bmatrix} + \begin{bmatrix} \alpha_1 \\ \alpha_2 \end{bmatrix} [p_{t-1}^A - \beta_0 - \beta_1 p_{t-1}^B] + \sum_{i=1}^{k} \begin{bmatrix} \delta_{1j} & \rho_{1j} \\ \delta_{2j} & \rho_{2j} \end{bmatrix} \begin{bmatrix} \Delta p_{t-i}^A \\ \Delta p_{t-i}^B \end{bmatrix} + \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{bmatrix}$$

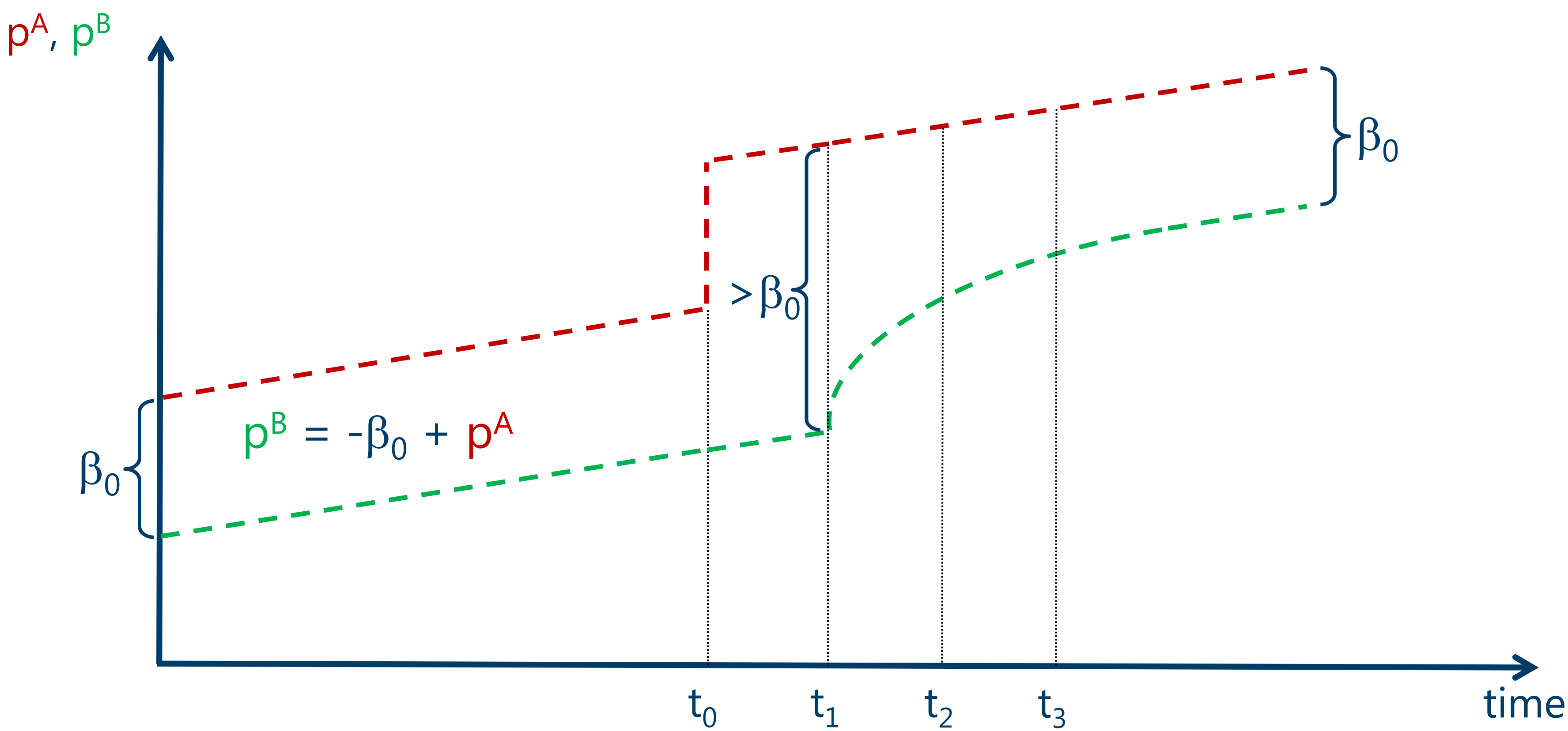
Example, p^A is the world market price, p^B is the domestic price in a large exporting country



Visualising cointegration in a VECM (I)

$$\begin{bmatrix} \Delta p_t^A \\ \Delta p_t^B \end{bmatrix} = \begin{bmatrix} \varphi_1 \\ \varphi_2 \end{bmatrix} + \begin{bmatrix} \alpha_1 \\ \alpha_2 \end{bmatrix} [p_{t-1}^A - \beta_0 - \beta_1 p_{t-1}^B] + \sum_{i=1}^{k} \begin{bmatrix} \delta_{1j} & \rho_{1j} \\ \delta_{2j} & \rho_{2j} \end{bmatrix} \begin{bmatrix} \Delta p_{t-i}^A \\ \Delta p_{t-i}^B \end{bmatrix} + \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{bmatrix}$$

Example, p^A is the world market price, p^B is the domestic price in a small exporting country



Interpretation

$$\begin{bmatrix} \Delta p_t^A \\ \Delta p_t^B \end{bmatrix} = \begin{bmatrix} \varphi_1 \\ \varphi_2 \end{bmatrix} + \begin{bmatrix} \alpha_1 \\ \alpha_2 \end{bmatrix} [p_{t-1}^A - \beta_0 - \beta_1 p_{t-1}^B] + \sum_{i=1}^{k} \begin{bmatrix} \delta_{1j} & \rho_{1j} \\ \delta_{2j} & \rho_{2j} \end{bmatrix} \begin{bmatrix} \Delta p_{t-i}^A \\ \Delta p_{t-i}^B \end{bmatrix} + \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{bmatrix}$$

- In most cases we expect α_1 to be negative and/or α_2 to be positive to ensure <u>error correction</u> rather than <u>error amplification</u>
- Together, $|\alpha_1| + \alpha_2$ measure the speed of error correction (always relative to the frequency of the data)
- Generally we expect α_1, α_2, and $|\alpha_1| + \alpha_2$ to be greater than zero and less than one (in price transmission applications)