Forms of Panel Data

- Consider a dataset in which each variable contains information on N panel units, each with T time-series observations.
- Mixed model vs. panel model (time, households, …)
- Micro panel (long panel) Macro panel (short panel), e.g. scanner data, world bank data, household survey…

$$
\begin{array}{cc}
 & \text{Cross section} \\
\begin{array}{c}\text{Time}\\ \text{series}\end{array} &
\begin{bmatrix}
y_{11} & y_{21} & \cdots & y_{i1} & \cdots & y_{N1} \\
y_{12} & y_{22} & \cdots & y_{i2} & \cdots & y_{N2} \\
\vdots & \vdots & \ddots & \vdots & & \vdots \\
y_{1t} & y_{2t} & \cdots & y_{it} & \cdots & y_{Nt} \\
\vdots & \vdots & & \vdots & \ddots & \vdots \\
y_{1T} & y_{2T} & \cdots & y_{iT} & \cdots & y_{NT}
\end{bmatrix}
\end{array}
$$

Panel Data Model: Type of Regressors

- Varying regressors x_{it}.

– Annual income for a person, temperature variation.

- Time-invariant regressors $x_{\mathrm{it}} = x_{\mathrm{i}}$ for all t. have zero within variation.

– Gender, race, education

- Individual-invariant regressors $x_{\mathrm{it}} = x_t$ for all i. have zero between variation.

– Time trend, inflation rate.

- Overall variation: variation over time and cross section
- Between variation: variation between cross section
- Within variation: variation over time

Overall, Between and Within Variations

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
ID	Time	Var	Individual mean	Overall mean	Overall deviation	Between deviation	Within deviation
i	t	x_{it}	$\bar{x}_i$	$\bar{x}$	$x_{it} - \bar{x}$	$\bar{x}_i - \bar{x}$	$x_{it} - \bar{x}_i$
1	2000	9	10	20	-11	-10	-1
1	2001	10	10	20	-10	-10	0
1	2002	11	10	20	-9	-10	1
2	2000	20	20	20	0	0	?
2	2001	20	20	20	0	0	?
2	2002	20	20	20	0	0	?
3	2000	25	30	20	5	10	-5
3	2001	30	30	20	10	10	0
3	2002	35	30	20	15	10	5

Data Summary

Measure	Definition	Mean	Min.	Max.	Std.D
$P_{i,t}^{w}$	**Real wholesale price**				
	Overall	723.126	347.342	1109.473	115.339
	Between		678.011	759.105	19.168
	Within		344.940	1105.530	113.789
$P_{i,t}^{f}$	**Real farm price**				
	Overall	539.747	303.512	987.800	105.430
	Between		518.237	564.270	9.594
	Within		296.547	963.277	105.007
$RM_{i,t}^{r-w}$	**Relative farm-wholesale margin**				
	Overall	.253	-.965	.458	.086
	Between		.199	.284	.019
	Within		-.972	.462	.084
$Temp_{i,t}$	**Temperature (degree Celsius)**				
	Overall	17.594	-6.900	39.600	9.820
	Between		10.536	27.462	4.211
	Within		-2.069	32.528	8.904

Overall variation is the movements over time and regions. Within variation denotes the movements over time. Between variation denotes the movements across regions.
Source: Own calculation based on data from State Livestock Affairs Logistics (2017) and Iran Meteorological Organization (2017).

Descriptive Analysis in Stata

- Import the data.
- Manage the panel settings of a dataset.
- Check data trend.
- What do you think about the variable trends, e.g. temperature?

An Overview of Stata

- To Start: Type clear in the command box to clear Stata’s memory.
- Do file: Instead of entering and running commands line-by-line in the command box, do-files allow the user to place commands in a text file and run them in batch.
- Data editor:
- Log-file: A log is a record of your Stata session.

Log using “file name.txt”, text append

- Help: ?? Or help …
- Install package: ssc install [name of package]

Panel Data Estimation Using Stata

- Setting panel data: xtset

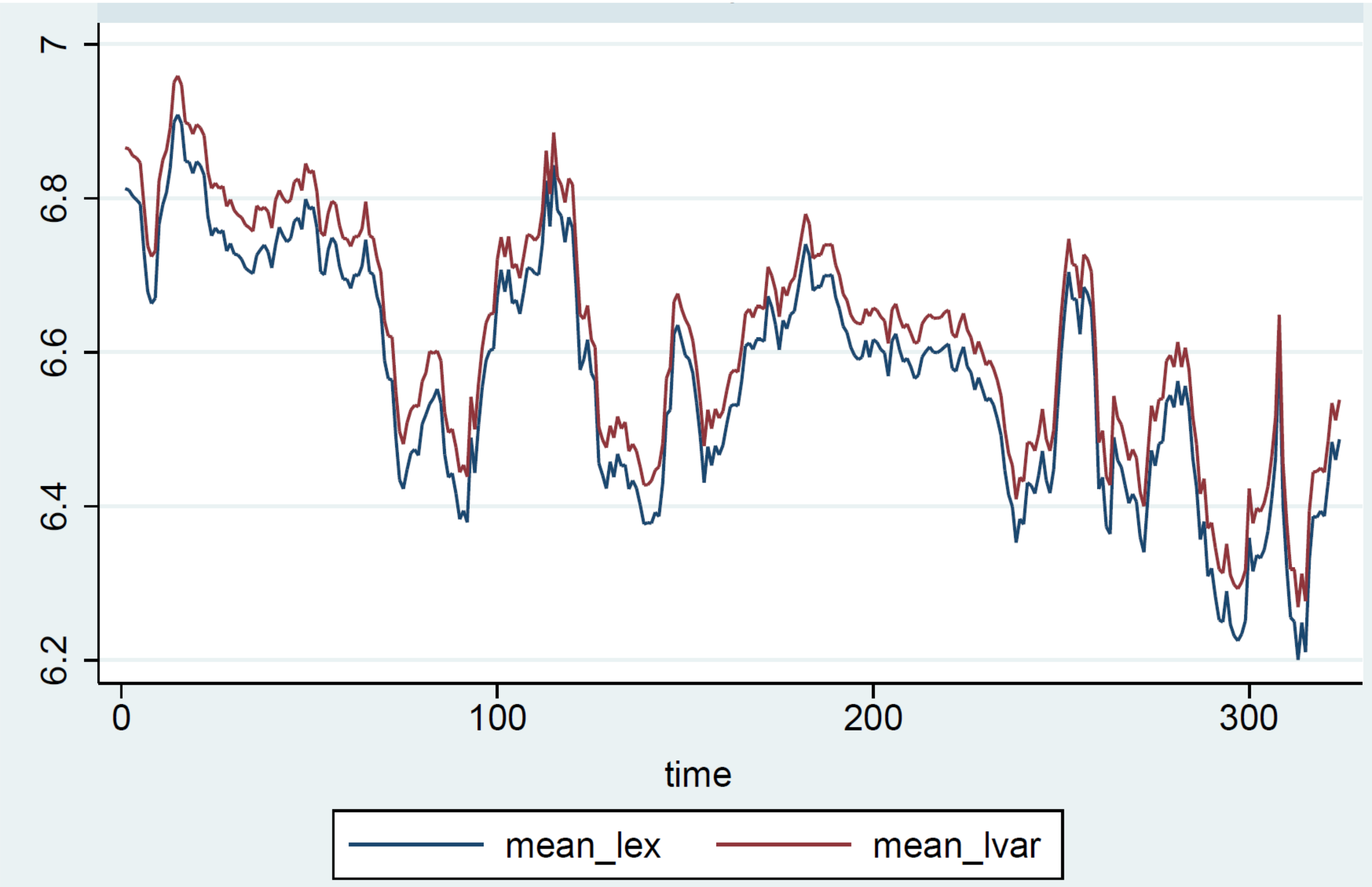
xtset id week

In this case “id” represents the entities or panels (i) and “week” represents the time variable (t).

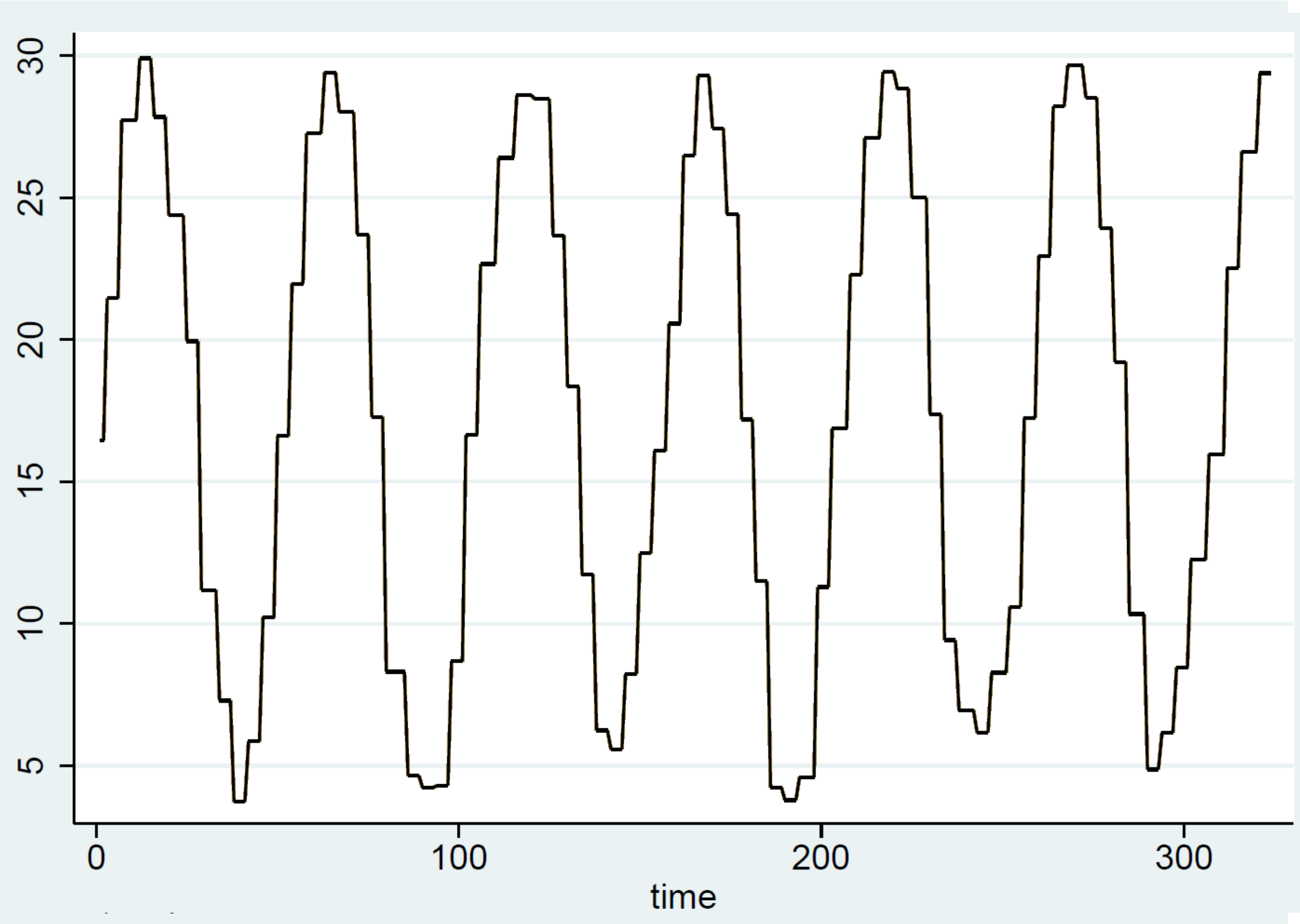
xtline var, overlay

xtline ex, overlay

Price Data



Temperature



Empirical Model

$$P^r - P^w = \mathrm{L}$$

Standard cost pass-through model:

$$\mathrm{P}_{\mathrm{it}}^{\mathrm{r}} = \alpha_{1\mathrm{i}} + \alpha_{2\mathrm{i}}\mathrm{P}_{\mathrm{it}}^{\mathrm{f}} + \mathrm{ECT}_{\mathrm{it}}$$

- $\mathbf{P}_{\mathbf{it}}^{\mathbf{r}}$: Retail prices of region i at time t.
- $\mathbf{P}_{\mathbf{it}}^{\mathbf{f}}$: Farm pices of region i at time t.
- $\mathbf{ECT}_{\mathbf{it}}$: Error correction term, idiosyncratic errors.
- $\boldsymbol{\alpha}_{\mathbf{1i}}$: Unobserved component (heterogeneity) Constant (size of transaction cost).
- $\boldsymbol{\alpha}_{\mathbf{2i}}$: Cost pass-through elasticty in long run.

Empirical Model

$$P_{it}^{r} = \alpha_{1i} + \alpha_{2i} P_{it}^{f} + \alpha_{3i} Temp_{it} + \alpha_{4i} \left(P_{it}^{f} \cdot Temp_{it} \right) + ECT_{it}$$

- Constant (size of transaction cost) = $\alpha_{1i} + \alpha_{3i} \overline{Temp}_{it}$
- Cost pass-through rate in long-run= $\alpha_{2i} + \alpha_{4i} \overline{Temp}_{it}$
- If $\alpha_{4i} > 0$ →?
- If $\alpha_{3i} > 0$ →?

Empirical Model

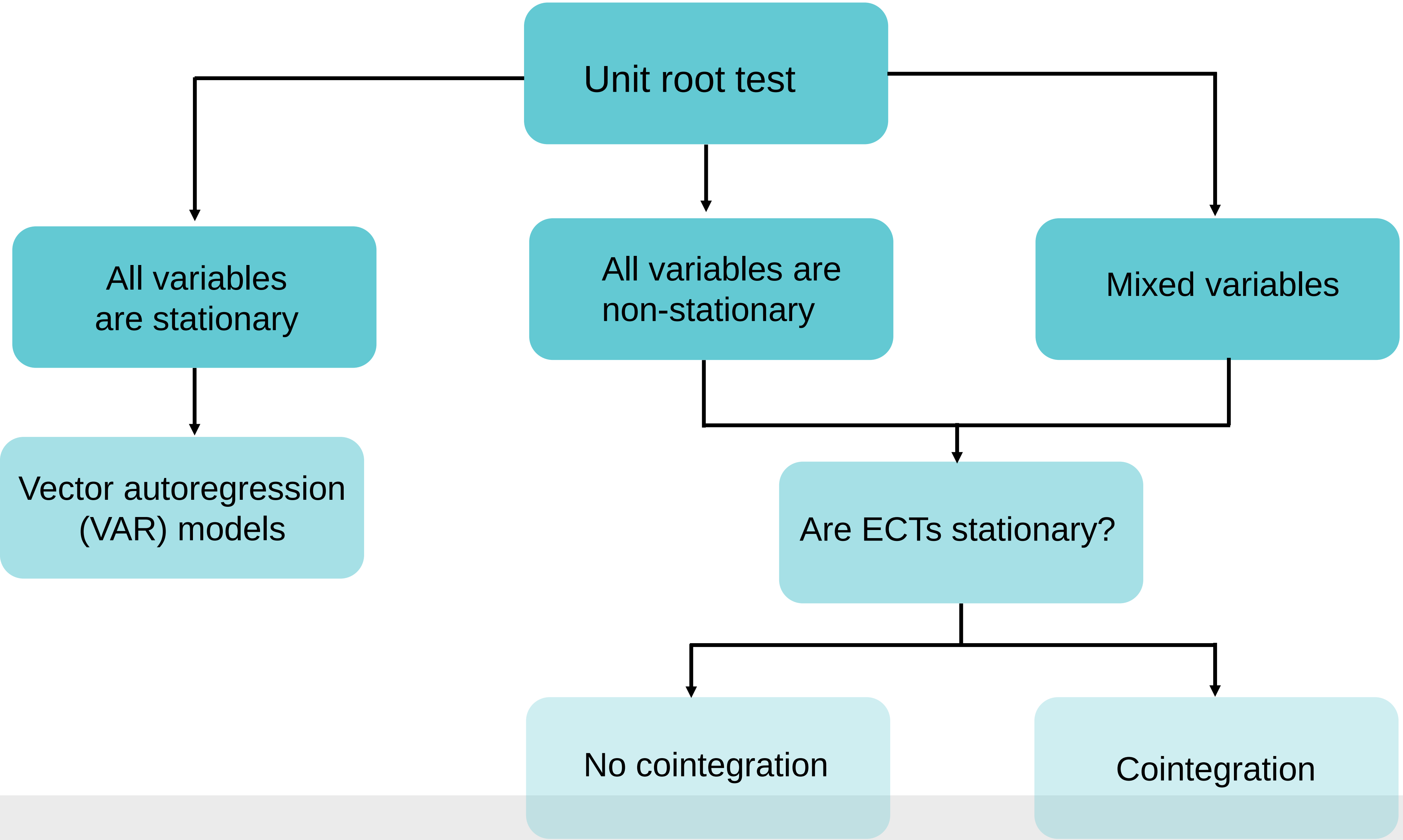
Price data are normally non-stationary→ **spurious regressio**

$$P_{it}^{r} = \alpha_{1i} + \alpha_{2i} P_{it}^{f} + ECT_{it}$$

Engle and Granger (1987): if ECT_{it} is stationary, then we have long-run relationship→ **co-integration**

1) long-run estimation → 2) check the error term→ 3)short-run estimation

Cointegration Methods



Empirical Strategy

- The two-step approach proposed by Engle and Granger (1987).

- First step: we first estimate the long-run relationship between wholesale and farmgate prices, considering temperature (i.e. $\mathrm{Temp}_{i,t}$),

$$\mathrm{P}_{\mathrm{it}}^{\mathrm{r}} = \alpha_{1\mathrm{i}} + \alpha_{2\mathrm{i}}\mathrm{P}_{\mathrm{it}}^{\mathrm{f}} + \alpha_{3\mathrm{i}}Temp_{\mathrm{it}} + \alpha_{4\mathrm{i}}\left(\mathrm{P}_{\mathrm{it}}^{\mathrm{f}} \cdot Temp_{\mathrm{it}}\right) + \mathrm{ECT}_{\mathrm{it}}$$

Then, we check if $\mathrm{ECT}_{\mathrm{it}}$ is stationary!?

- Second step: We estimate short-run relationship using an Error Correction Model!?

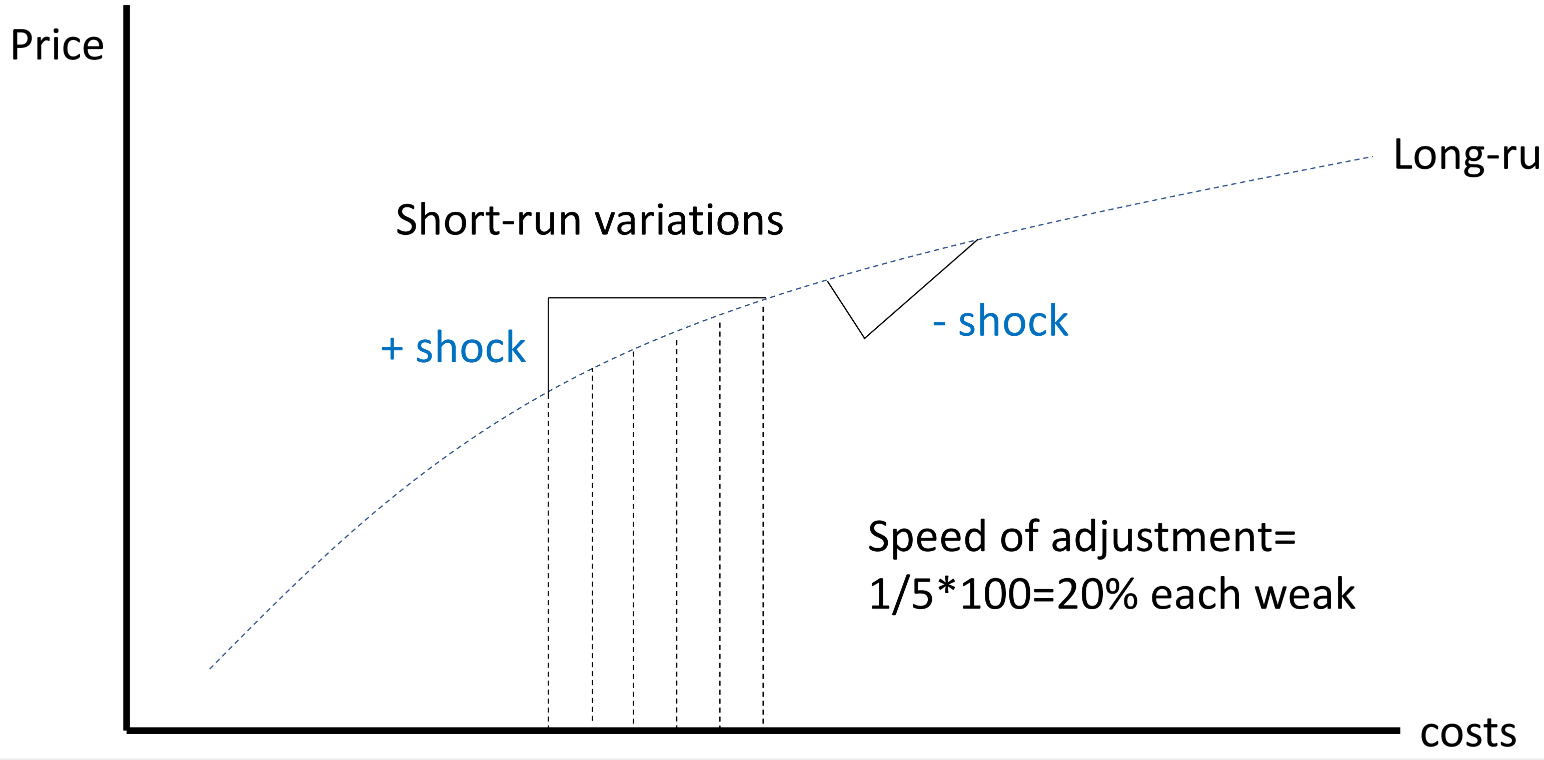
Symmetric Short-run Price Adjustment

$$\Delta p_{i,t}^{r} = \rho_{\mathrm{i}} + \gamma_0 \mathrm{ECT}_{\mathrm{i,t-1}} + \gamma_1 \mathrm{ECT}_{\mathrm{i,t-1}} Temp_{\mathrm{it}} + \sum_{\mathrm{r=1}}^{\mathrm{p}} \lambda_r^{j} \Delta \mathrm{P}_{\mathrm{i,t-r}}^{\mathrm{r}} + \sum_{\mathrm{r=0}}^{\mathrm{q}} \sigma_{\mathrm{r}}^{\mathrm{j}} \Delta \mathrm{P}_{\mathrm{i,t-r}}^{f} + \nu_{\mathrm{i,t}}^{\mathrm{j}}$$

- $\mathbf{ECT_{i,t-1}}$**:** Lagged error correction term of the long-run equilibrium.
- $\boldsymbol{\gamma_0}$: Speed of adjustment toward long-run equilibrium. $P^r - \alpha_1 + \alpha_2 P^w = \mathrm{ECT_t}$.
- $\boldsymbol{\gamma_1}$: The impact of Temperature on the speed of adjustment.
- Speed of adjustment**:** $\frac{\partial \Delta p_{i,t}^{r}}{\mathrm{ECT_{i,t-1}}} = \boldsymbol{\gamma_0} + \boldsymbol{\gamma_1} \overline{Temp}_{\mathrm{it}}$

Speed of Adjustment

- If the speed of adjustment is 0.25: the target variable adjusts at a speed of 25% per week to equilibrium.



Model Estimation

- **Time series models**

Non-linear ARDL, VECM, Threshold VECM, Switching Models.

- **Panel models**
- Dynamic panel models, Fixed-effects, random-effects models, panel VECM, Mean Group Models, Fully modified least squares (FM-OLS), Instrumental variable estimators, Threshold Panel Models.
- **We look at**

How to detect a suitable way to estimate the empirical model.

Panel Data Model

$$\boldsymbol{y}_{\mathbf{it}} = \boldsymbol{\alpha} + \sum_{j=2}^{k} \boldsymbol{\beta}_{j} \boldsymbol{X}_{jit} + \mu_{\mathbf{i}} + \boldsymbol{\varepsilon}_{\mathbf{it}}$$

Note: If the $\boldsymbol{X}_{j}$ are so comprehensive that they capture all relevant characteristics of individual i, $\mu_{\mathbf{i}}$ can be dropped and, then, pooled OLS may be used. <u>But, this is situation is very unlikely.</u>

In general, dropping $\mu_{\mathbf{i}}$ leads to missing variables problem: bias!

- We usually think of $\mu_{\mathbf{i}}$ as contemporaneously exogenous to the conditional error. That is, $\mathrm{E}[\boldsymbol{\varepsilon}_{\mathbf{it}}|\mu_{\mathbf{i}}] = \mathbf{0}$, t=1,…,T.
- A stronger assumption: Strict exogeneity can also be imposed. Then

$\mathrm{E}[\boldsymbol{\varepsilon}_{\mathbf{it}}|x_{\mathbf{i1}}, \dots, x_{\mathbf{iT}}, \mu_{\mathbf{i}}] = \mathbf{0}$, t=1,…,T.

Panel Data Model

$$y_{it} = \alpha_1 z_i + \alpha_2 x_{it} + \varepsilon_{it}$$

x_{it}: a vector of observed characteristics

z_i: a vector of unobserved characteristics

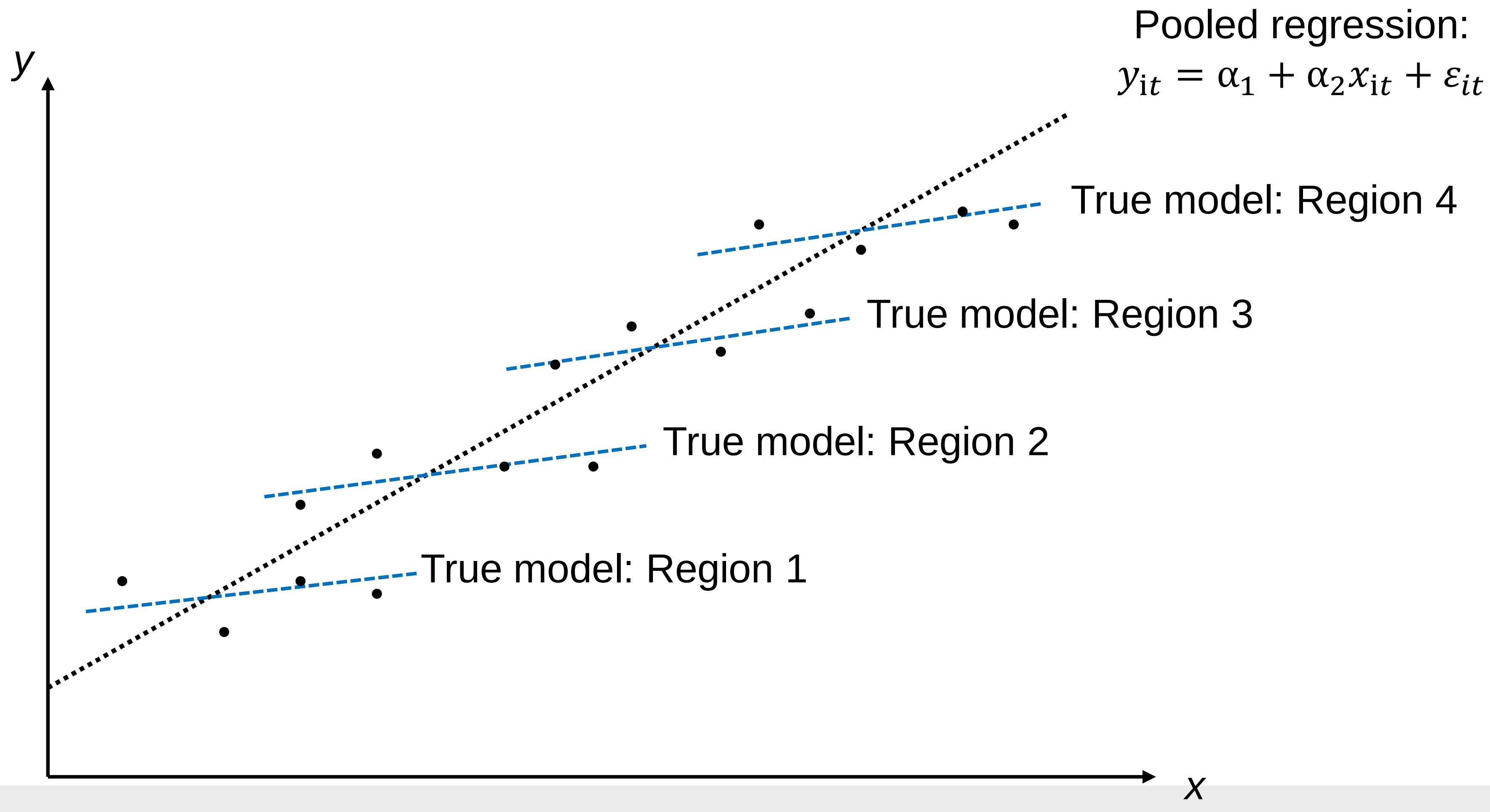
Depending on how we model the heterogeneity in the panel, we have different models.

z_i is constant and uncorrelated with x_{it}! Dependence on the y_{it} may enter through the variance. That is, repeated observations on individual i are linearly independent. In this case,

$$y_{it} = \alpha_1 + \alpha_2 x_{it} + \varepsilon_{it}$$

Panel Model Type: Pooled Model

Pooled regression may result in heterogeneity bias:



Panel Model Type: Fixed Effects Model (One-way)

$$y_{\mathrm{it}} = \alpha_1 z_{\mathrm{i}} + \alpha_2 x_{\mathrm{it}} + \varepsilon_{it}$$

- The unobservable effects (z_{i}) are correlated with included variable; pooled OLS will be inconsistent. What “inconsistency” means!?

The distribution of $\widehat{\alpha}_n$ doesn’t collapse on α as n becomes large→ $plim\widehat{\alpha}_n \neq \alpha$

- Assume $\mathrm{E}[\boldsymbol{z}_{\mathbf{i}}|x_{\mathbf{it}}] = \alpha_{\boldsymbol{i}}$ → it does not vary with t. Then,

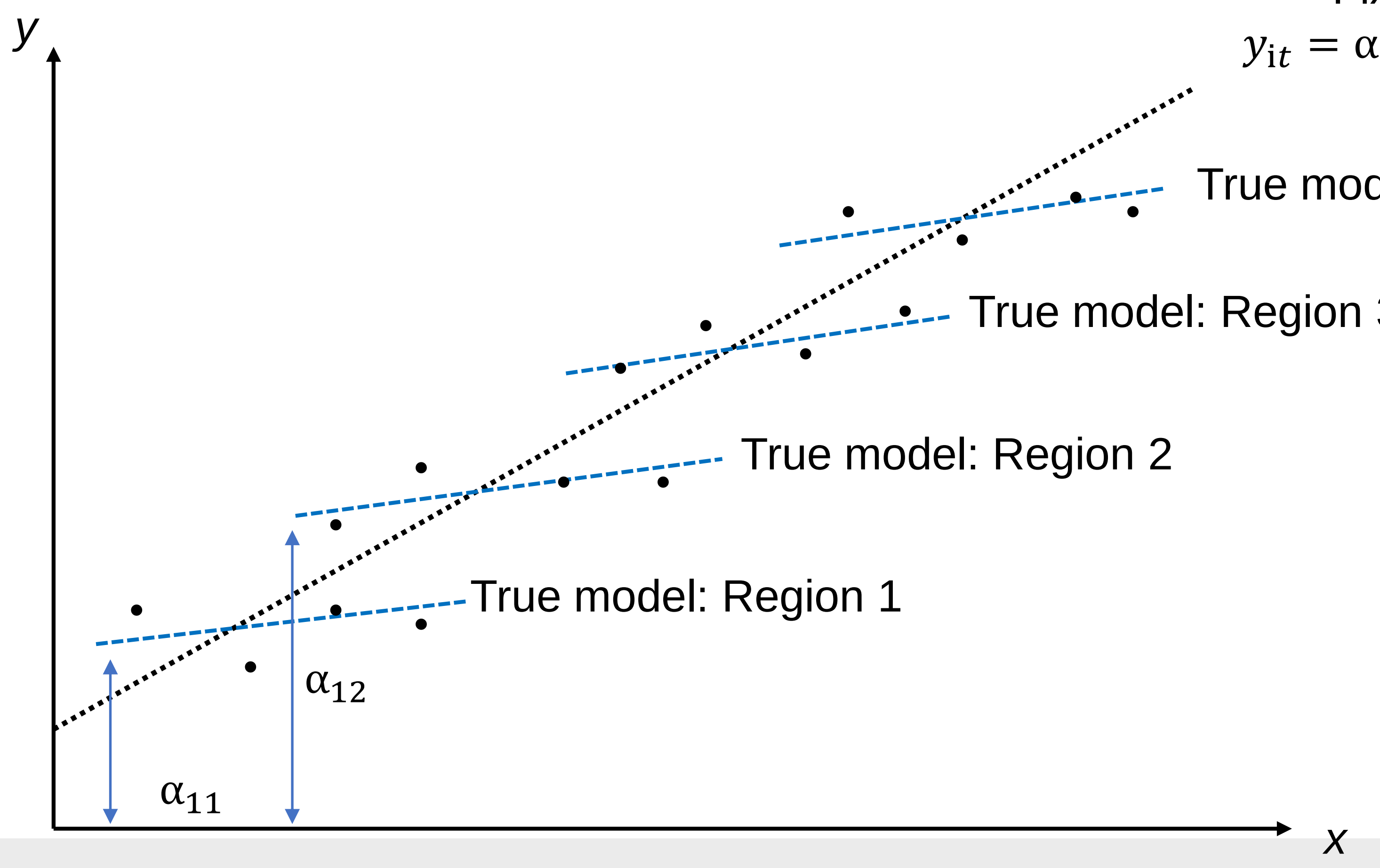
$$y_{\mathrm{it}} = \alpha_{1i} + \alpha_2 x_{\mathrm{it}} + \varepsilon_{it}$$

- The regression line is raised/lowered by a fixed amount for each individual i. In econometrics terms, this is the source of the fixed-effects.
- OLS can be used to estimates α_{1i} and α_2 consistently

Panel Model Type: Fixed Effects Model (One-way)

Fixed effect:

$$y_{\mathrm{it}} = \alpha_{1i} + \alpha_2 x_{\mathrm{it}} + \varepsilon_{it}$$



Fixed-effect Model

$$y_{it} = \alpha_{1i} + \alpha_2 x_{it} + \varepsilon_{it} \qquad \mathrm{i} = 1, \ldots, \mathrm{N}; \mathrm{t} = 1, \ldots, \mathrm{T}$$

With i denoting housholds, regions, firms, etc. and t denoting time.

- **One-way error component model:**

$$\mathrm{u_{it}} = \mu_\mathrm{i} + \upsilon_\mathrm{it}$$

μ_i is the unobservable individual-specific effect and υ_it is the reminder disturbance.

- **Two-way error component model:**

$$\mathrm{u_{it}} = \mu_\mathrm{i} + \lambda_\mathrm{t} + \upsilon_\mathrm{it}$$

λ_t is the unobservable individual-invariant effect

Estimation Within–Group Variation

- How do we estimate it?

We have to remove fixed effect unobserved factor! How!?

- The simplest way to allow each region to have its own intercept is to create a set of dummy (binary) variables, one for each region i, and include them as regressors.
- Alternatively, first difference (FD) models.
- They involve averaging the original equation over time (LSDV) or first differencing (FD) to wipe out the unobserved

On-Way Fixed-effect Model

One can obtain the LSDV (Least Square Dummy Variables) estimator, by premultiplying the model by Q and performing OLS on the resulting transformed model:

$$Qy = QX\beta + Q\upsilon$$

The Q matrix wipes out the individual effects.

$$Q\mu_i = 0$$

$$y_{it} = X_{it}\beta + \alpha_i + \varepsilon_{it}$$

Demeaning,

$$y_{it} - \bar{y}_i = (X_{it} - \bar{X}_i)\beta + \alpha_i - \alpha_i + (\varepsilon_{it} - \bar{\varepsilon}_i)$$

$$\bar{y}_i = \frac{1}{T}\sum_{t=1}^{T} y_{it}, \qquad \bar{X}_i = \frac{1}{T}\sum_{t=1}^{T} X_{it}, \qquad \bar{\varepsilon}_i = \frac{1}{T}\sum_{t=1}^{T} \varepsilon_{it}, \qquad \alpha_i = \bar{\alpha}_i$$

Least-squares Dummy Variables

- In dynamic panel data models, dummy variables may be introduced to the least squares to explain the effect of each individual unit of a cross section which is unobserved but correctly specifies the model of relation.

- In the LSDV estimation the individual effect is assumed to be fixed over time in each individual.

- in small sample case i.e. short time period, the LSDV estimator is inconsistent owing to the incidental parameters problem.

Estimation of Two Way Fixed Effects

- A further extension is to allow the intercept to vary across the different time periods:

$$y_{\mathrm{it}} = \sum_{i=1}^{N} \alpha_{1\mathrm{i}} D_{\mathrm{it}} + \sum_{i=1}^{T} \alpha_{2\mathrm{i}} T_{\mathrm{it}} + X_{\mathrm{it}}\beta + \varepsilon_{\mathrm{it}}$$

- The time dummy coefficients can allow the regression function to shift over time to capture changes in technology, government regulation, tax policy, external influences (wars…) etc.

Unobserved Factors

- “The key insight is that if the unobserved variable does not change over time, then any changes in the dependent variable must be due to influences other than these fixed characteristics.” (Stock and Watson, 2003, p.289-290).

- “In the case of time-series cross-sectional data the interpretation of the beta coefficients would be “…for a given country, as X varies across time by one unit, Y increases or decreases by β units”.

Random-effect Model

- The fixed effects model assumes that each group (region) has a non-stochastic group-specific component to y.

- But these unobservable effects may be stochastic (i.e. random). The Random Effects Model attempts to deal with this:

$$y_{it} = \alpha + X'_{it}\beta + \mu_i + \varepsilon_{it}$$

- μ_i, is treated as a component of the random error term. μ_i is the element of the error which varies between groups but not within groups. ε_{it} is the element of the error which varies over group and time.

Random-effect Model

- Estimation of the random effects model cannot be performed by OLS, generalised least squares (GLS) must be used.
- We could also introduce an error component which varies across time periods but not across groups – two way random effects.

– $\mathrm{E}(\mu_i) = \mathrm{E}(\varepsilon_{it}) = 0$

– Both components homoscedastic; $\mathrm{E}(\mu_i^2) = \sigma_v^2 \qquad \mathrm{E}(\varepsilon_{it}^2) = \sigma_\varepsilon^2$

– Both independent of regressors; $\mathrm{E}(\mu_i x_{it}) = \mathrm{E}(\varepsilon_{it} x_{it}) = 0$

– No cross group correlation; $\mathrm{E}(\mu_i \mu_j) = 0 \text{ if } i \neq j$

– No autocorrelation; $\mathrm{E}(\varepsilon_{it} \varepsilon_{js}) = 0 \text{ if } t \neq s \text{ or } i \neq j$

Estimate the Long-Run Equilibrium in Stata

- OLS estimate, command:

Fixed effect: xtreg varlist, fe robust

Random effect: xtreg varlist, re robust

OLS regression→ xi: regress lex lfr i.id

- Rho is the interclass correlation of the error . Rho is the fraction of the variance in the error due to the individual-specific effects. It approaches 1 if the individual effects dominate the idiosyncratic error $\rho_{\varepsilon} = \mathrm{cor}\left(\varepsilon_{\mathrm{it}} \varepsilon_{\mathrm{js}}\right)$

the errors are correlated with the regressors in the fixed effect.

t-values test the hypothesis that each coefficient is different from 0. To reject this, the t-value has to be higher than 1.96 (for a 95% confidence).

Number of region

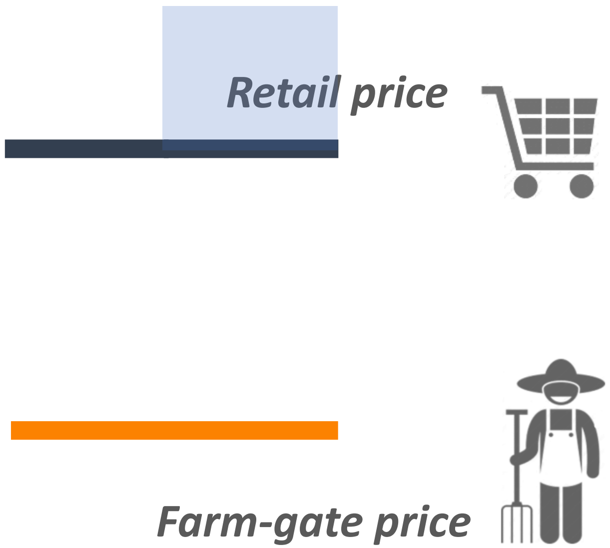
```
. xtreg lex lfr temp lfrtemp , fe  robust

Fixed-effects (within) regression               Number of obs      =      9,720
Group variable: id                              Number of groups   =         30

R-sq:                                           Obs per group:
     within  = 0.6903                                         min =        324
     between = 0.2572                                         avg =      324.0
     overall = 0.6782                                         max =        324

                                                F(3,29)            =    1928.43
corr(u_i, Xb)  = 0.0058                         Prob > F           =     0.0000

                                 (Std. Err. adjusted for 30 clusters in id)
------------------------------------------------------------------------------
             |               Robust
         lex |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         lfr |   .8143476    .016714    48.72   0.000     .7801636    .8485316
        temp |   .0412066   .0042842     9.62   0.000     .0324444    .0499688
     lfrtemp |  -.0065287   .0006864    -9.51   0.000    -.0079326   -.0051248
       _cons |   1.460863   .1040476    14.04   0.000     1.248062    1.673664
-------------+----------------------------------------------------------------
     sigma_u |  .02359275
     sigma_e |   .0886436
         rho |  .06615139   (fraction of variance due to u_i)
------------------------------------------------------------------------------
```

Coefficients of the regressors.

It is a test (F) to see whether all the coefficients in the model

29.7% of the variance is due to differences across panels.

Results



Independent variables	$\ln P^{w}_{i,t}$ Fixed Effects	$\ln P^{w}_{i,t}$ Random Effects	$\ln P^{w}_{i,t}$ Mean Group
$\ln P^{f}_{i,t}$ Log farm price	.814*** (.016)	…	
$Temp_{i,t} \cdot \ln P^{f}_{i,t}$ Temperature· log farm price	**-.006** (.102)**		
$Temp_{i,t}$ Temperature	**.041*** (.004)**		
Constant	1.460*** (.086)		
Wald tests	1928*** [F test]		…

Panel Data Estimation: Tests for Poolability

One can test for joint significance of the dummy variables:

Panel model: $y_{\mathrm{it}} = \alpha_{1i} + \alpha_2 x_{\mathrm{it}} + \varepsilon_{it}$

Pool model: $y_{\mathrm{it}} = \alpha_1 + \alpha_2 x_{\mathrm{it}} + \varepsilon_{it}$

Null hypothesis: $H_0\colon \alpha_{1\mathrm{i}} = 0, i = 1, \ldots, \mathrm{N}$

$$\mathrm{F}_0 = \frac{(\mathrm{RRSS} - \mathrm{URSS})/(\mathrm{N} - 1)}{\mathrm{URSS}/(\mathrm{NT} - \mathrm{N} - \mathrm{K})} \sim \mathrm{F}_{\mathrm{N}-1,\mathrm{N}(\mathrm{T}-1)-\mathrm{K}}$$

Choosing Fixed or Random Effect Models?

α_{1i} should be treated as a random effect or a fixed effect?

In modern econometric, “random effect” is synonymous with zero correlation between the observed explanatory variables and the unobserved effect: $\mathrm{Cov}(\alpha_{1i}, x_{it}) = 0$

Fixed-effects will not work well with data for which within-cluster variation is minimal or for slow changing variables over time

Choosing Fixed or Random Effect Models?

- With large *T* and small *N* there is likely to be little difference, so FE is preferable as it is easier to compute
- With large *N* and small *T*, estimates can differ significantly. If the cross-sectional groups are a random sample of the population RE is preferable. If not the FE is preferable.
- If the error component, μ_i, is correlated with *x* then RE is biased, but FE is not.
- For large *N* and small *T* and if the assumptions behind RE hold then RE is more efficient than FE.

Hausman‘s test

RE holds under the null

```
xtreg  y x1, fe
estimates store fixed
xtreg  y x1, re
estimates store random
hausman fixed random
```

```
. hausman fixed random

                 ---- Coefficients ----
             |      (b)          (B)            (b-B)     sqrt(diag(V_b-V_B))
             |     fixed        random       Difference          S.E.
-------------+----------------------------------------------------------------
          x1 |    2.48e+09     1.25e+09        1.23e+09        6.41e+08
------------------------------------------------------------------------------
                           b = consistent under Ho and Ha; obtained from xtreg
            B = inconsistent under Ha, efficient under Ho; obtained from xtreg

    Test:  Ho:  difference in coefficients not systematic

                  chi2(1) = (b-B)'[(V_b-V_B)^(-1)](b-B)
                          =        3.67
                Prob>chi2 =      0.0553
```

If this is < 0.05 (i.e. significant) use fixed effects.

Cross-sectional Dependence Test

Assume,

$$y_{it} = \alpha_i + \beta x_{it} + \mu_{it} \ , i = 1, . . , N \ t = 1, \dots . , T$$

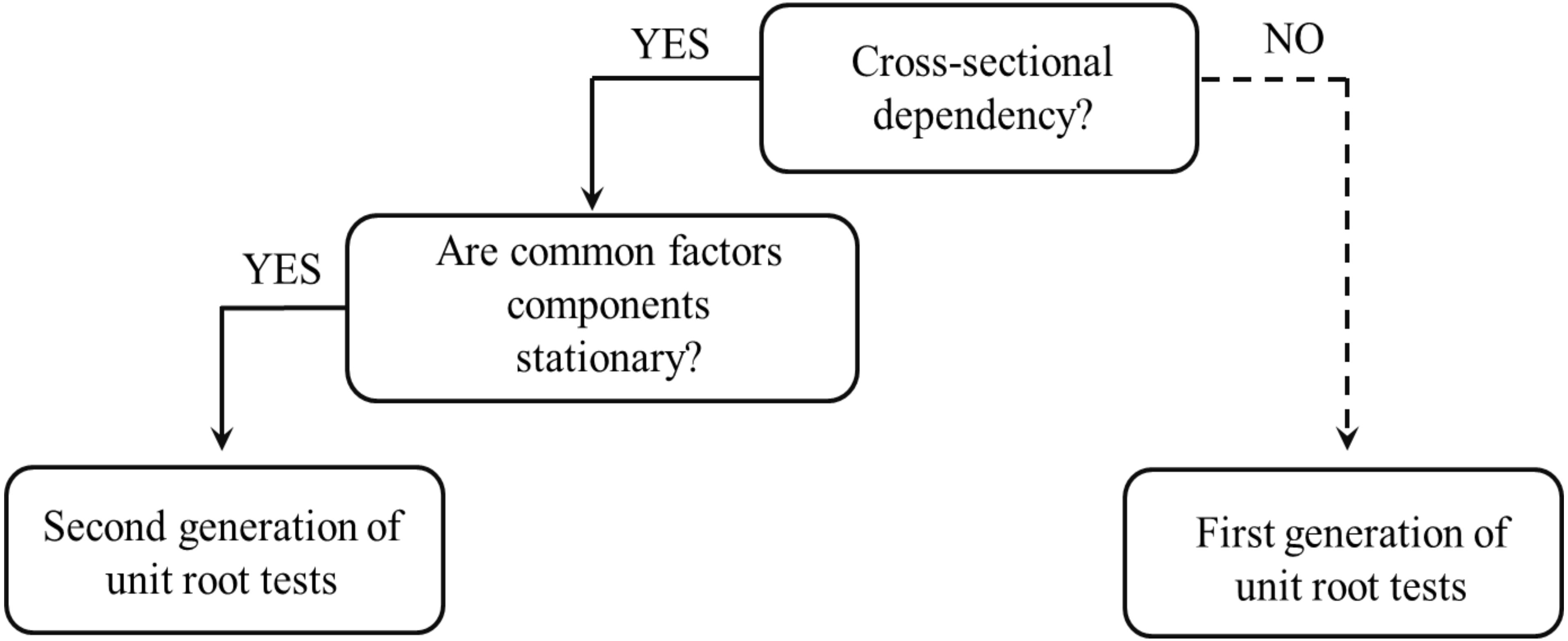
$$H_0: \ \rho_{ij} = \mathrm{cor}(u_{it}, u_{jt}) = 0 \text{ for } i \neq j$$
$$H_1: \ \rho_{ij} = \mathrm{cor}(u_{it}, u_{jt}) \neq 0 \text{ for } i \neq j$$

ρ_{ij}: the product-moment correlation coefficients of the residuals

$$\hat{\rho}_{ij} = \hat{\rho}_{ji} = \frac{\sum_t^T \hat{u}_{it}\hat{u}_{jt}}{(\sum_t^T \hat{u}_{it}^2)^{0.5} \ (\sum_t^T \hat{u}_{jt}^2)^{0.5}}$$

Possible tests: Pesaran's CD test, Friedman's test, Frees' test,

Panel Unit Root Test



Consider cross sectional dependencies, Hadri (2002)

Ignore cross sectional dependencies, …

Cross-sectional Dependence Test in Stata

xtcd varlist

Unit Root Test

The situation is different, when cross-sectional dependence is caused by common dynamic factors.

Breitung and Das (2008) consider the following three cases:

(1) both, the common and idiosyncratic components are nonstationary;

(2) common factors are I(1) and the idiosyncratic components are I(0); and

(3) the common factors are I(0) and the idiosyncratic components are I(1)

<u>A variable is nonstationary if the common factors and/or the idiosyncratic components are nonstationary </u>(Breitung & Das,2008).

Unit Root Test in Stata

$$\ln \overline{P}_t^r = \frac{1}{n} \sum_i \ln P_{i,t}^r$$

Xtunitroot hadri var, robust

Xtunitroot hadri d.var, robust

Xtunitroot fisher var, dfuller lags(3)

What we should to do next?

Cointegration Test

The residual-based cointegration tests majorly suffer from a common-factor restriction which can cause a significant loss of power (Banerjee et al., 1998).

In order to solve this problem, Westerlund (2007) proposes a general test that is based on structural dynamics and, thus, does not assume any common-factor restriction.

$$\Delta \ln P_{it}^{r} = \rho_i d_t + \gamma_{i0}(ECT_{it-1}) + \sum_{j=1}^{p} \sigma_{ij}^{r} \Delta \ln P_{it-j}^{r} + \sum_{j=0}^{q} \sigma_{ij}^{w} \Delta \ln P_{it-j}^{w} + \nu_{it}$$

Where ECT_{it-1} is error terms defined as $\ln P_{it-1}^{r} - \beta_i \ln P_{it-1}^{w}$, and d_t denotes the deterministic component

H_0: no cointegration

Cointegration test in Stata

Westerlund (2008):

Xtwest varlist, lags()

$$\Delta \ln P_{it}^{r} = \rho_i d_t + \gamma_{i0}(ECT_{it-1}) + \sum_{j=1}^{p} \sigma_{ij}^{r} \Delta \ln P_{it-j}^{r} + \sum_{j=0}^{q} \sigma_{ij}^{w} \Delta \ln P_{it-j}^{w} + \nu_{it}$$

First, if we assume heterogeneity between cross sections in terms of error correction coefficients (mean group test), then the alternative is defined as $H_1: \gamma_i < 0$ for at least one i. The second option is when we assume the error correction coefficients do not vary over cross sections (panel tests).

Measures of Goodness of Fit

- There is no generally accepted standard for measures of goodness of fit.
- R2 within describes the goodness of fit for the observations that have been adjusted for their individual means (maximized by LDSV). How much of the variation in the dependent variable **within** regions is captured by your model.
- The R2 between describes the goodness of fit for the N different individual means. How much of the variation in the dependent variable **between** regions is captured by your model
- Finally, the R2 overall corresponds to the usual R2 of OLS regression. weighted average of within and between R2.

Random Coefficients Model

We can introduce heterogeneity through coefficients:

$$y_{\mathrm{it}} = X_{\mathrm{it}}\beta_{\mathrm{i}} + \alpha_{\mathrm{i}} + \varepsilon_{\mathrm{it}}$$
$$\beta_{\mathrm{i}} = (\beta + b_{\mathrm{i}})$$

b_{i} is a random vector that induces parameter variation.

The coefficients are different for each individual, like Mean Group models (discussed in the following slides)

It is possible to complicate the model by making them different through time:

$$\beta_{\mathrm{i}t} = (\beta + b_{\mathrm{i}}) + \varphi_t$$

Mean Group Model

Fixed effect: $\ln \mathrm{DP}_{\mathrm{it}}^{\mathrm{r}} = \alpha_{1\mathrm{i}} + \boldsymbol{\alpha_2} \ln \mathrm{DP}_{\mathrm{it}}^{\mathrm{f}} + \mu_{\mathrm{it}}$

The assumption of homogeneity of slope parameters is often inappropriate.

Mean group: $\ln \mathrm{DP}_{\mathrm{it}}^{\mathrm{r}} = \alpha_{1\mathrm{i}} + \boldsymbol{\alpha_{2i}} \ln \mathrm{DP}_{\mathrm{it}}^{\mathrm{f}} + \mu_{\mathrm{it}}$

The MG estimator (see Pesaran and Smith 1995) relies on estimating N time-series regressions and averaging the coefficients, whereas the PMG estimator (see Pesaran, Shin, and Smith 1997, 1999) relies on a combination of pooling and averaging of coefficients.

Mean Group Model

$$lnDP_{it}^{r} = \alpha_{1i} + \alpha_{2i} lnDP_{it}^{f} + \mu_{it}$$

$$\widehat{\alpha}_1 = \frac{1}{N}\sum_{i} \widehat{\alpha}_{1i} \quad \text{(MG estimate of constant)}$$

$$\widehat{\alpha}_2 = \frac{1}{N}\sum_{i} \widehat{\alpha}_{2i} \quad \text{(MG estimate of price coefficient)}$$

$\widehat{\alpha}_{1i}$ and $\widehat{\alpha}_{2i}$ are the OLS estimator of α_{1i} and α_{2i}

The mean group estimator provides consistent estimates of the parameters' averages for a large sample size ($\frac{\sqrt{N}}{T} \rightarrow 0$) (Pesaran and Smith, 1995).

Pooled Mean Group Model

- The PMG estimator is an "intermediate estimator:

It allows the intercepts, short-run coefficients and error variances to differ freely across cross-section while the long-run coefficients are constrained to be the same.

1. Estimating a fixed effect model for the long-run equilibrium
2. Store the error term
3. Estimating an ARDL model for each cross-section

Pooled Mean Group Model in Stata

Estimating long-run equilibrium

- Pooled Mean Group

xtpmg *depvar* [*indepvars*] [*if*] [*in*] [**,** *options*]

- Mean Group

xtmg *varlist* [*if*] [*in*] [**, trend robust cce aug imp full level(***num***) res(***string***)**]

Short-run Price Adjustment

So far, we estimated long-run Equilibrium: $\mathrm{P}_{\mathrm{it}}^{\mathrm{r}} = \alpha_{1\mathrm{i}} + \alpha_{2\mathrm{i}} \mathrm{P}_{\mathrm{it}}^{\mathrm{f}} + \mathrm{ECT}_{\mathrm{it}}$

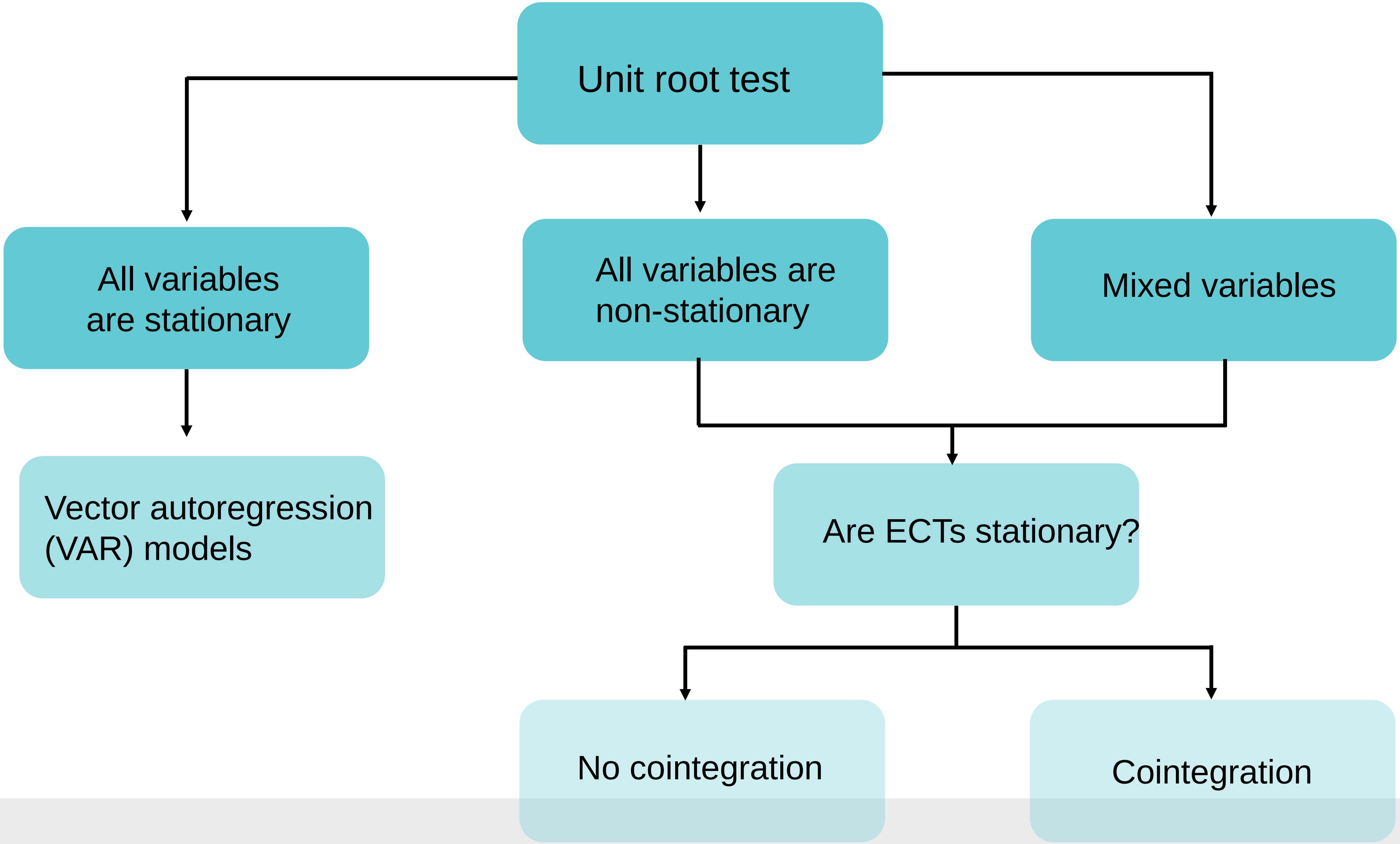
Variables at level

Variables are nonstationary

Engle and Granger (1987): if $\mathrm{ECT}_{\mathrm{it}}$ is stationary, then we have long-run relationship→ **co-integration**

1) long-run estimation → 2) check the error term→ 3)short-run estimation using Error Correction Model (ECM)

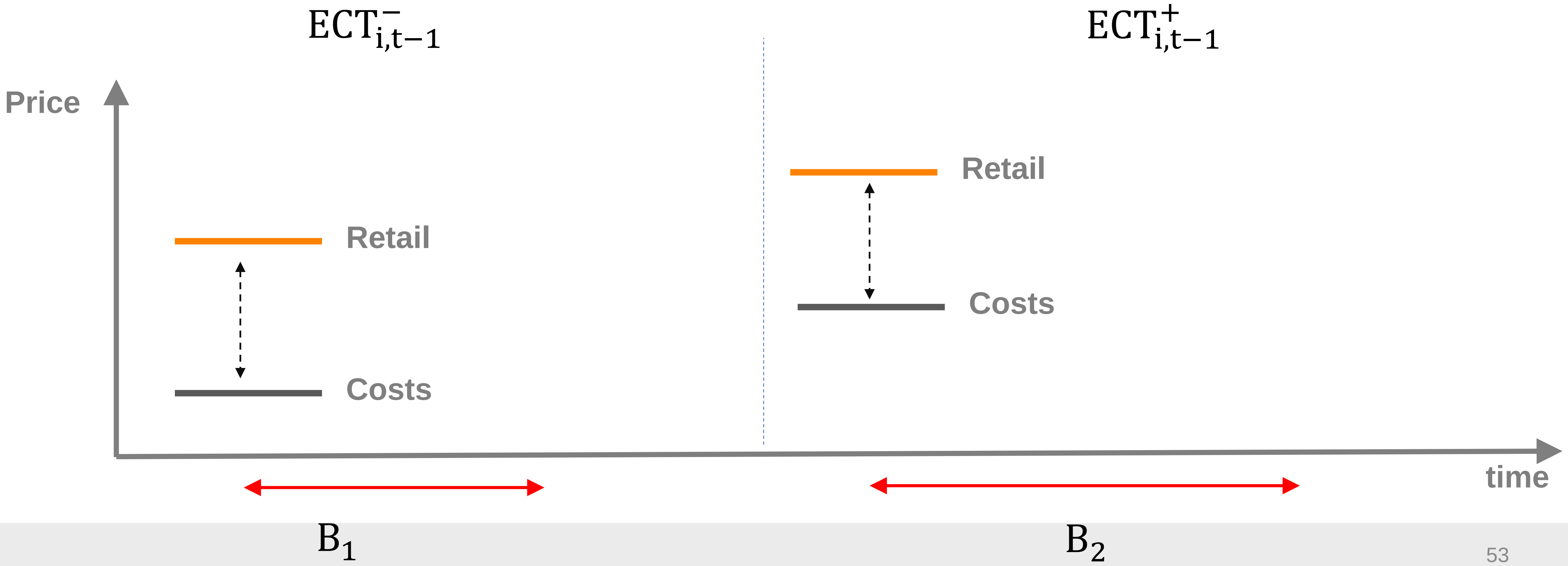
Cointegration Methods



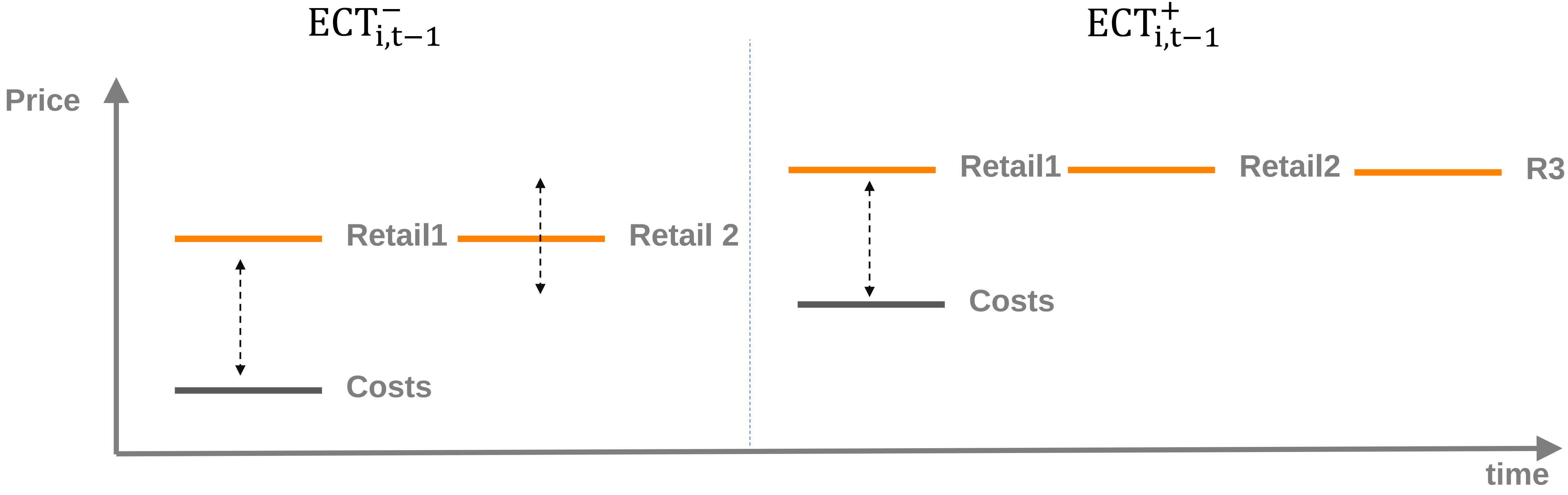
Short Run vs. Long Run

$$P^r = \alpha_1 + \alpha_2 P^w + \mathrm{ECT_t}$$

How fasts retail price reacts!?



Short-run Price Adjustment



Short Run vs. Long Run

- What is the implication of short-run or long-run effects of extreme weather effect!?

- **Long-run effect:** it might be permanent, the variable is I(1)

- **Short-run effects:** it is not permanent, the variable is I(0), or seasonality (we will discuss about seasonality in the next slides).

Symmetric Short-run Price Adjustment

We estimated $\mathrm{ECT}_{\mathrm{i,t-1}}$ from the fist step.

Second step: Error Correction Model (ECM)

$$\Delta P^{r}_{i,t} = \rho_i + \gamma_{0,i} ECT_{i,t-1} + \sum_{j=1}^{p} \sigma^{r}_{i,j}\, \Delta P^{r}_{i,t-j} + \sum_{j=0}^{q} \sigma^{w}_{i,j}\, \Delta P^{w}_{i,t-j} + \upsilon_{i,t}$$

$$\Delta P^{r}_{i,t} = \rho_i + \gamma_{0,i} ECT_{i,t-1} + \sigma^{r}_{i,1} \Delta P^{r}_{i,t-1} + \sigma^{w}_{i,0} \Delta P^{w}_{i,t} + \sigma^{w}_{i,1} \Delta P^{w}_{i,t-1} + \upsilon_{i,t}$$

Positive deviations from the long-term equilibrium if $\mathrm{ECT}_{\mathrm{i,t-1}} \geq 0$

Negative deviations from the long-term equilibrium if $\mathrm{ECT}_{\mathrm{i,t-1}} < 0$

Adding Interaction Terms

$$\Delta P_{i,t}^{r} = \rho_i + \gamma_0 ECT_{i,t-1} + \gamma_1 ECT_{i,t-1} \times Temp_{i,t} + \sum_{j=1}^{p} \sigma_j^r \, \Delta P_{i,t-j}^{r} + \sum_{j=1}^{p} \sigma_j^r \, \Delta P_{i,t-j}^{r} \times Temp_{i,t}$$

$$+ \sum_{j=0}^{q} \sigma_j^w \, \Delta P_{i,t-j}^{w} + \sum_{j=0}^{q} \sigma_{,j}^w \, \Delta P_{i,t-j}^{w} \times Temp_{i,t} + \upsilon_{i,t}$$

Effects of temperature on speed of adjustment: $\frac{\partial \, \Delta P_{it}^{r}}{\partial ECT_{i,t-1}} = \gamma_0 + \gamma_1 \times Temp_{i,t}$

If $\gamma_1 < 0$→ higher speed of adjustment

If $\gamma_1 > 0$→ lower speed of adjustment

Short-run estimate in Stata

- Similar to the long-run equation, but with lagged variables.

$\Delta \rightarrow$ in Stata d.

$\sum_{j=1}^{3} \sigma_j^r \, \Delta P_{i,t-j}^r \rightarrow$ in Stata d.l(1/3). $P_{i,t}^r$

```
xtreg d.var d.reg , fe robust
xtreg d.var d.reg , re robust
```

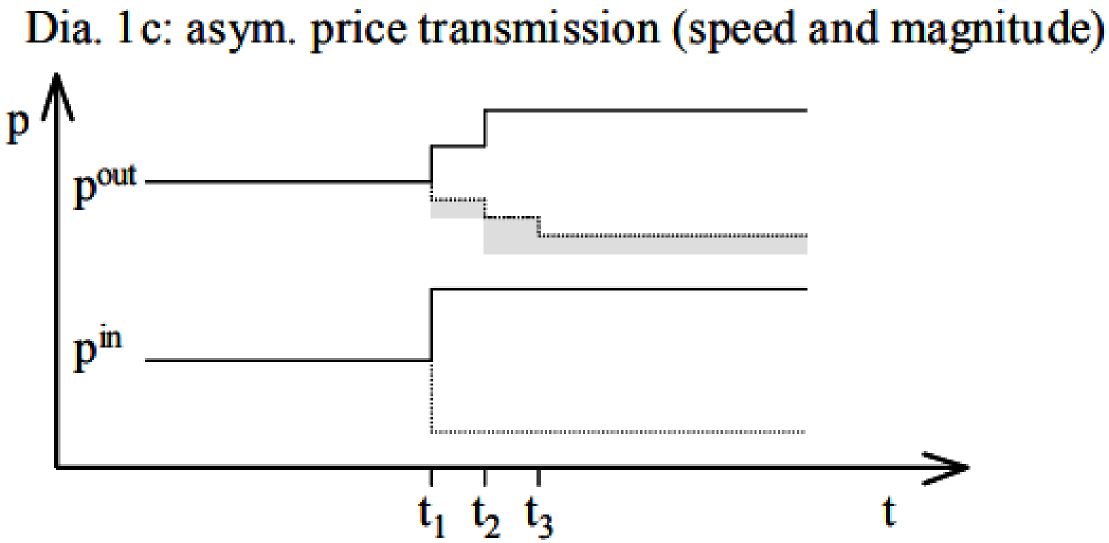
Rockets and Feathers Effects

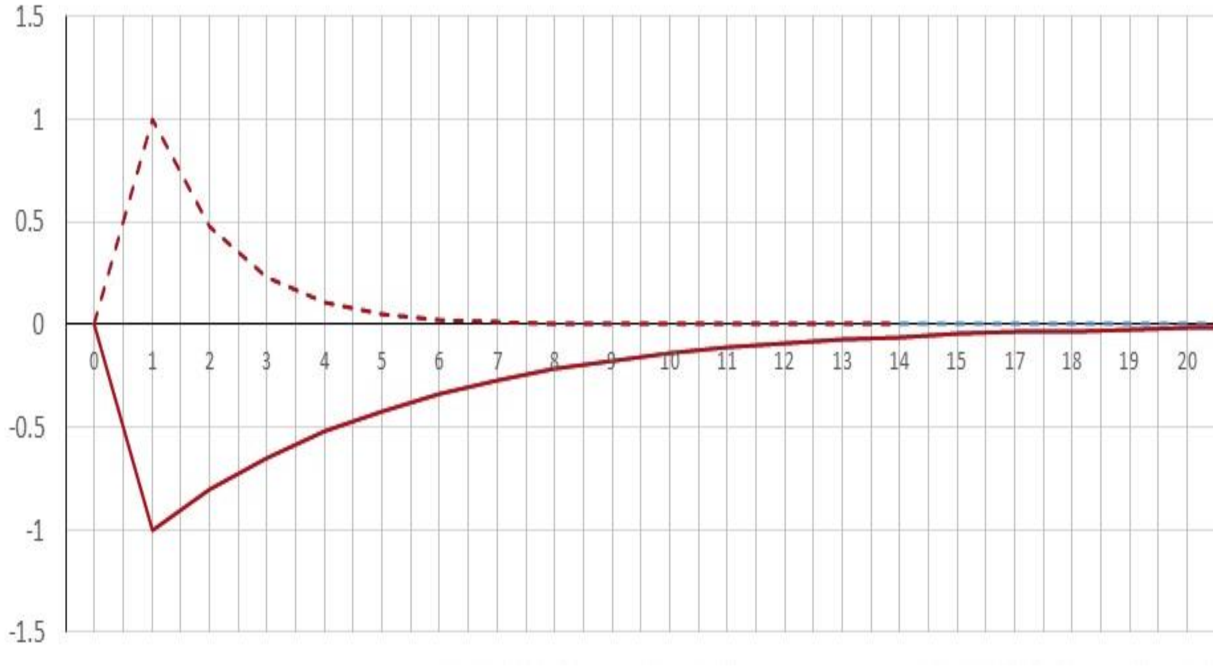
Prices rise like rockets but fall like feathers.

This stylized fact of many markets is confirmed by many empirical studies.

According to Peltzman's comprehensive study of 165 producer goods and 77 consumer goods, "In two out of three markets, output prices rise faster than they fall"

Asymmetric Short-run Price Adjustment





Asymmetric Short-run Price Adjustment

We estimated $\mathrm{ECT}_{i,t-1}$ from the previous model.

Positive deviations from the long-term equilibrium: $\mathrm{ECT}^{+}_{i,t-1}$, if $\mathrm{ECT}_{i,t-1} \geq 0$

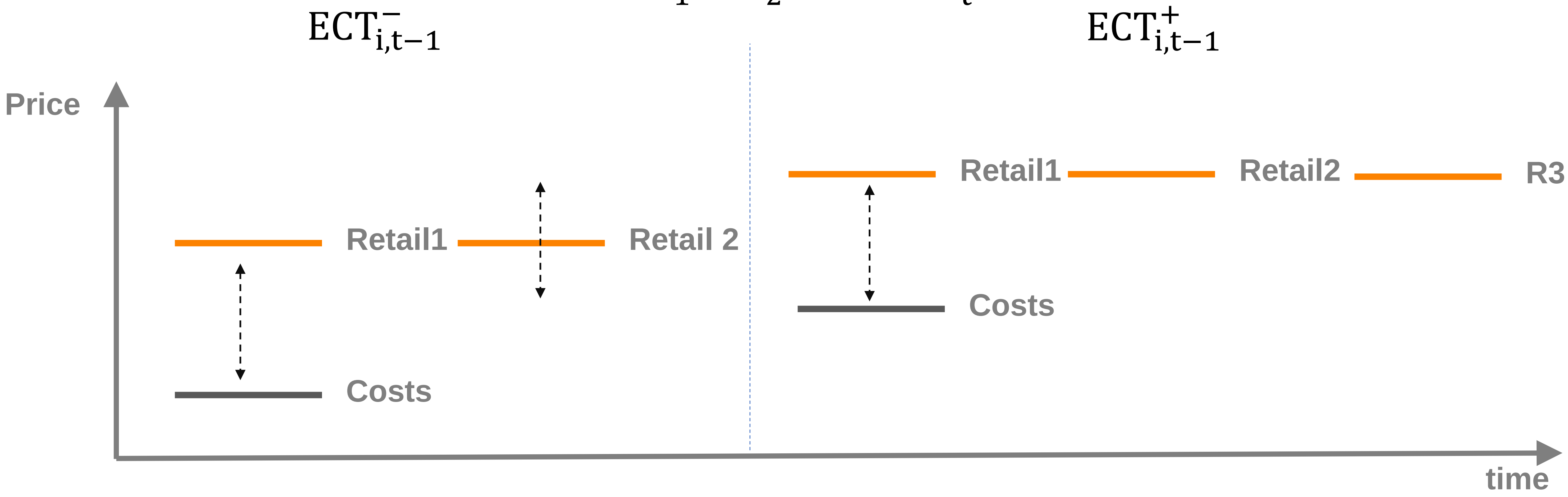
Negative deviations from the long-term equilibrium: $\mathrm{ECT}^{-}_{i,t-1}$, if $\mathrm{ECT}_{i,t-1} < 0$

$$\Delta P_{i,t} = \gamma_0^{+}\mathrm{ECT}^{+}_{i,t-1} + \gamma_1^{-}\mathrm{ECT}^{j,-}_{i,t-1} + \cdots + \rho_i \quad + \nu_{it}$$

$$\Delta P_{i,t} = \gamma_0^{+}\mathrm{ECT}^{+}_{i,t-1} + \gamma_1^{-}\mathrm{ECT}^{j,-}_{i,t-1} + \gamma_2^{+}\mathrm{Temp}_{i,t-1}\mathrm{ECT}^{+}_{i,t-1} + \gamma_3^{-}\mathrm{Temp}_{i,t-1}\mathrm{ECT}^{-}_{i,t-1} + \cdots + \rho_i \quad + \nu_{it}$$

Rockets and Feathers Effects

$$P^r - \alpha_1 + \alpha_2 P^w = \mathrm{ECT_t}$$



Parameter Interpretation

$$\Delta P_{i,t} = \gamma_0^+ ECT_{i,t-1}^+ + \gamma_1^- ECT_{i,t-1}^- + \boldsymbol{\gamma_2^+} Temp_{i,t-1} ECT_{i,t-1}^+ + \boldsymbol{\gamma_3^-} Temp_{i,t-1} ECT_{i,t-1}^- + \cdots + \rho_i + \nu_{it}$$

More asymmetric pattern

$$\frac{\partial\ \Delta lnP_{it}^r}{\partial ECT_{i,t-1}^+} = \gamma_0^+ + \boldsymbol{\gamma_2^+}\mathrm{Temp}_{t-1} \rightarrow \text{if } \gamma_2^+ > 0$$

$$\frac{\partial\ \Delta lnP_{it}^r}{\partial ECT_{i,t-1}^-} = \gamma_1^- + \boldsymbol{\gamma_3^-}\mathrm{Temp}_{t-1} \rightarrow \text{if } \gamma_3^- < 0$$

More symmetric pattern

$$\gamma_2^+ < 0$$

$$\gamma_3^- > 0$$

- Rockets and feathers effect: $|\gamma_0^+| < |\gamma_0^-|$

Asymmetric Price Adjustment in Stata

- Similar to the long-run equation, but with lagged variables.

$\Delta \rightarrow$ in Stata d.

$\sum_{j=1}^{3} \sigma_j^r \, \Delta P_{i,t-j}^r \rightarrow$ in Stata d(1/3). $P_{i,t}^r$

- Xtreg

Results

The higher speed of adjustment $\boldsymbol{\gamma_1} < 0$

Independent variables	$\Delta\ln DP^{r}_{i,t}$ Fixed Effects	$\Delta\ln DP^{r}_{i,t}$ Random Effects	$\Delta\ln DP^{r}_{i,t}$ Mean Group
$ECT_{i,t-1}$ Error correction term	-.041*** (.004)		
$ECT_{i,t-1}.Temp_{i,t}$ Error correction term · search costs	**-.349*** (.043)**		
$\Delta\ln P^{f}_{i,t}$ First difference log farm price	.662*** (.022)		
:	:		
Constant	.000*** (.000)		
Wald tests	1658.96*** [F test]	[chi2]	[chi2]

Asymptotic Bias of the Panel OLS (Fixed Effect)

- We now consider a dynamic panel data model, in the sense that it contains (at least) one lagged dependent variables, e.g. short-run equation.

- If lagged dependent variables appear as explanatory variables, strict exogeneity of the regressors no longer holds. The LSDV is no longer consistent when n tends to infinity (Nickell 1981).

- Only for the special case in which the regressors are strictly exogenous and the dynamics are homogeneous across members.

Source: Hurlin, 2018

Dynamic Panel Bias

What are the solutions:

Consistent estimator of parameters can be obtained by dynamic models:

- The IV (Instrumental Variable) approach,
  - Anderson and Hsiao, (1982),
- The GMM (Generalized Method of Moment) approach
  - Arenallo and Bond, (1985).

The weak instrument variable problem!?

Dynamic Panel Data

Arellano and Bond (AB) Approach

- An estimator designed for situations with:
  - ‘small T, large N’ panels: few time periods and many individual units, Macro panels.

  - Right-hand variables that are not strictly exogenous: correlated with past and possibly current realizations of the error fixed individual effects, implying unobserved heterogeneity heteroskedasticity and autocorrelation within individual units’ errors, but not across them.

- The instrumental variables should satisfy some properties:

Relevance (high correlation), exogeneity.

Nickell Bias

- The bias of the LSDV estimator in a dynamic model is generally known as dynamic panel bias or Nickellís bias (1981).
- A challenge arises with the one-way fixed effects model in the “small T, large N" context.
- As Nickell (Econometrica, 1981) shows, this arises because the demeaning process which subtracts the individual’s mean value of y and each X from the respective variable creates a correlation between regressor and error.

Nickell bias rules of thumb: T/N →0

Fully Modified OLS (FMOLS)

- Though ordinary least square (OLS) estimates are super-consistent with cointegrated variables, their finite-T bias can be large in the presence of endogenous feedback (Phillips and Hansen, 1990; Pedroni, 2000).
- In presence of endogeneity problem, the error terms can be serially correlated which would result in the dependence of OLS estimators on nuisance parameters.

$$y_{it} = \alpha_i + \beta x_{it} + \mu_{it}$$
$$x_{it} = x_{it-1} + \varepsilon_{it}$$

Vector error process: $\xi_{it} = (\mu_{it}, \varepsilon_{it})'$, covariance matrix Ω_i

Fully Modified OLS (FMOLS)

- FMOLS estimator is derived by making endogeneity correction (by modifying regressor x_{it}) and serial correlation correction (by modifying long run covariance matrix, Ω_{i}). The resulting final estimator is expressed as follows:

- $\hat{\beta}_{\mathrm{it}} - \beta = \left(\sum_{i=1}^{N} \hat{L}_{22i}^{-2} \sum_{i=1}^{T} (x_{\mathrm{it}} - \bar{x}_{\mathrm{i}})^2\right)^{-1} \sum_{i=1}^{N} \hat{L}_{11i}^{-1} \hat{L}_{22i}^{-1} \left(\sum_{i=1}^{T} (x_{\mathrm{it}} - \bar{x}_{\mathrm{i}}) \mu_{\mathrm{it}} - T\, \hat{\gamma}_{\mathrm{i}}\right)$

The constant term is removed!

Dynamic Ordinary Least Squares (DOLS)

- Panel DOLS involves augmenting the panel cointegrating regression equation with cross-section specific lags and leads of to eliminate the asymptotic endogenity and serial correlation.

- Then, the panel DOLS model is applied for a robustness check. The panel DOLS has some better sample properties than the panel FMOLS, such as a less bias estimator in small sample size (Rahman, 2017; (Kasman and Duman, 2015).

FMOLS and DOLS in Stata

“xtcointreg” command in Stata helps to estimate the long-run relationship between panel variables, using two options of the panel FMOLS and DOLS model.

Xtcointreg

xtcointtest

xtpmg

Short-run vs. Long-run Effects

Figure 2. Increase in Average Global Temperature and Contributions of Key Factors
(Deviations from 1880–1910 average, degrees Celsius)

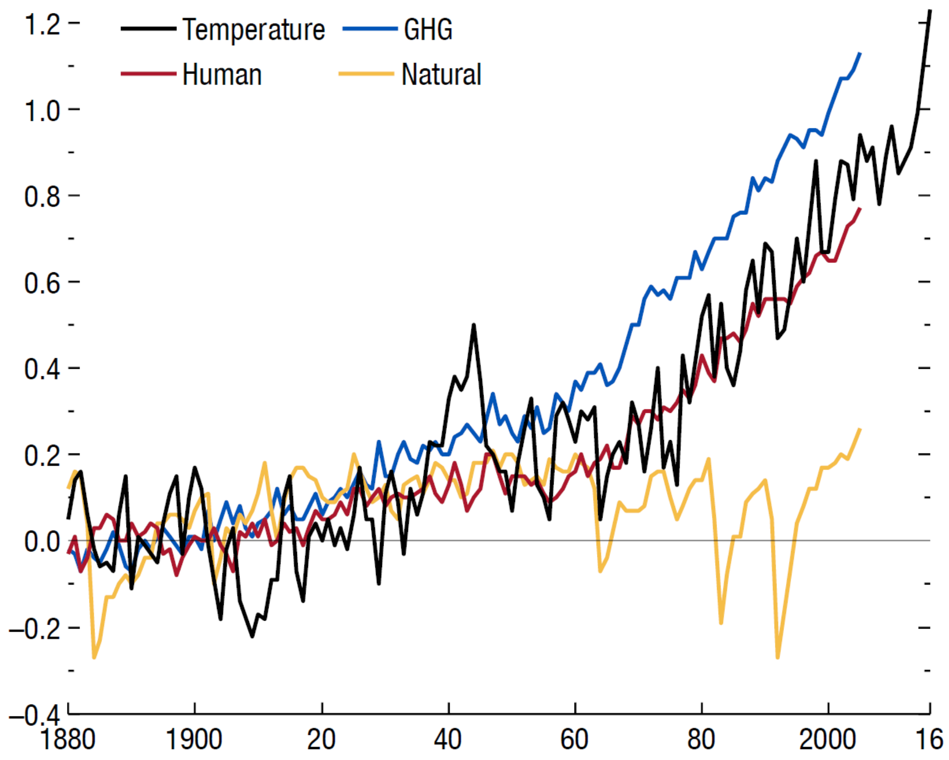
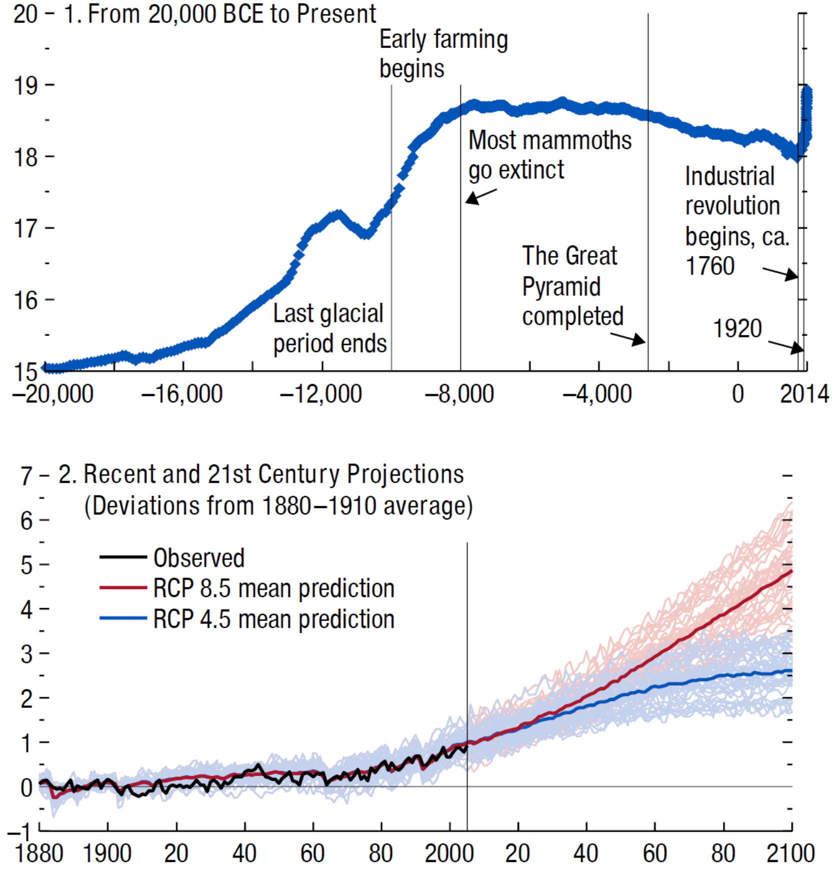


Figure 1. Average Global Temperature
(Degrees Celsius)

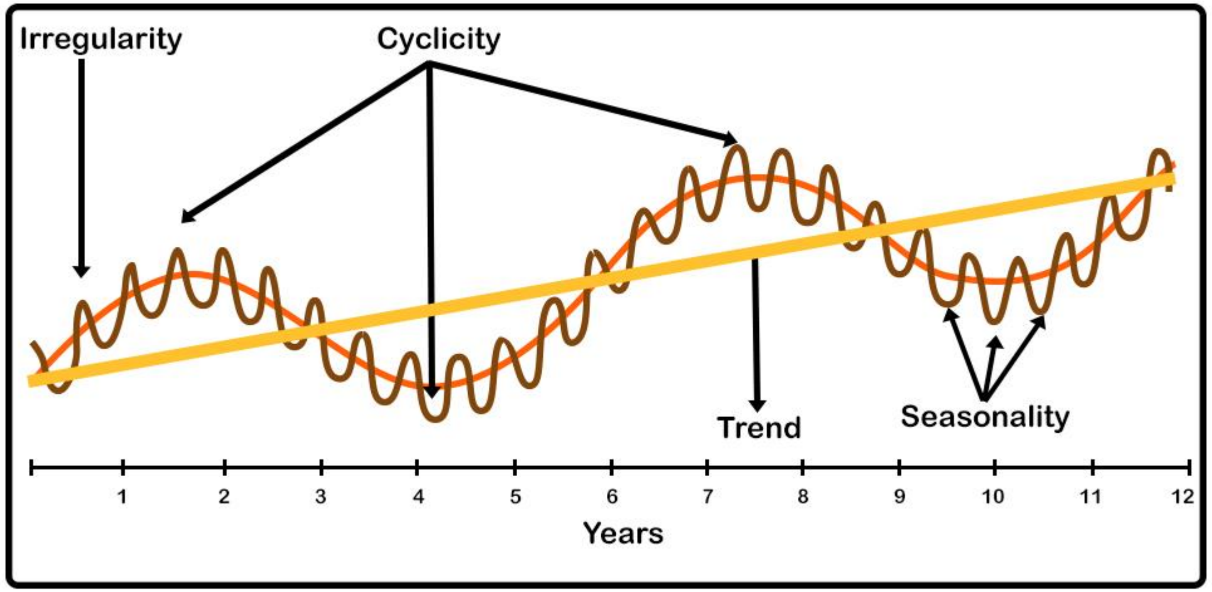


Source: IMF, 2018

Seasonality in Economics

- *"... systematic, although not necessarily regular, intra-year movement caused by changes of the weather, the calendar, and timing of decisions..."* (Hans Franses).

$Y_t = S_t + C_t + T_t + \varepsilon_t$



S_t: Seasonal component

C_t: Cyclical component

T_t: Trend component

ε_t: Stochastic/unexplained component

Seasonality in Economics

- Three major time series models to investigate seasonal effects:

1. deterministic seasonality (dummy variable),

2. stationary stochastic seasonality,

– Generated by a seasonal ARMA: $\varphi(L^s)y_t = \theta(L^s)\varepsilon_t \rightarrow \mathrm{L}y_t = y_{t-1}, L^s y_t = y_{t-s}$

3. seasonal unit roots (Ghysels et al., 2001).

If a series has a seasonal unit root, the seasonal component is changing across time in ways that are hard to predict.

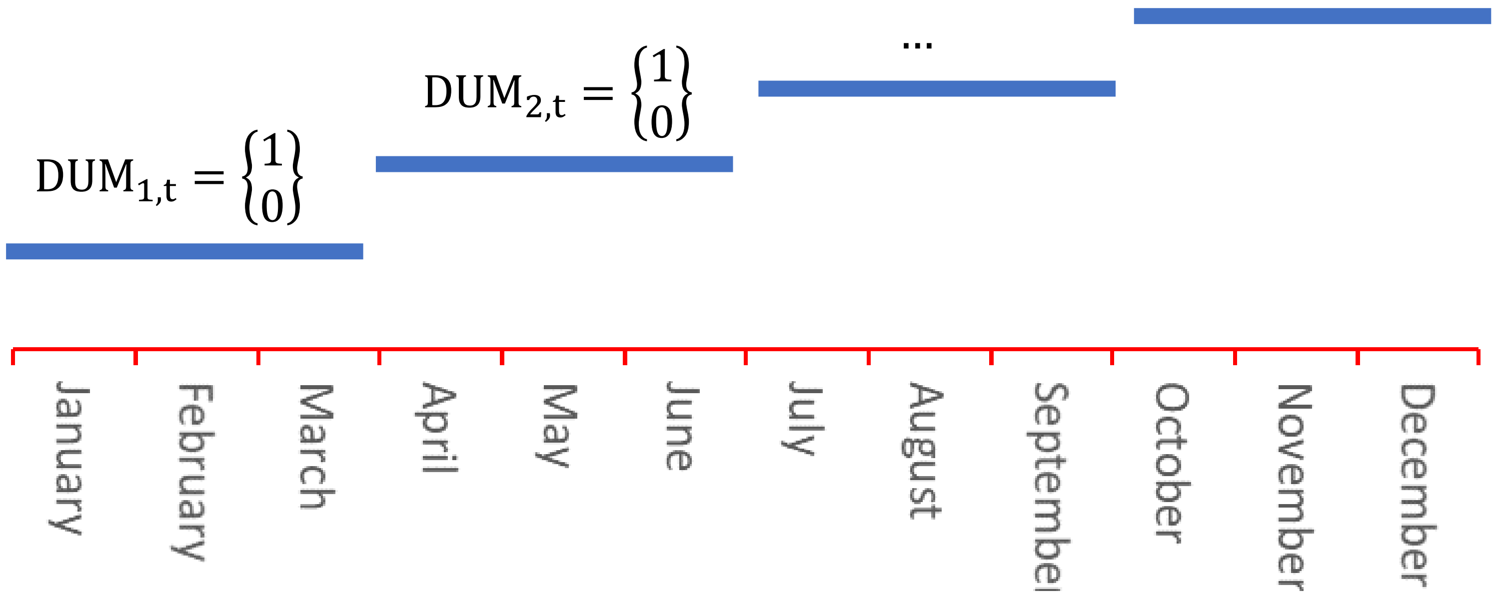
one example of a seasonal unit root is a process where the seasonal peak is changing through time.

Seasonality in Economics

- Seasonality can be modelled in two ways: deterministically or stochastically.
  – Calendar effects or climatic phenomena and can be removed from data by the seasonal adjustment procedures

Deterministic seasonality

- By deterministic fixed dummies defining seasons (e.g. Mehta and Chavas, 2008).
- The familiar seasonal dummy variable regression: $y_{\mathrm{t}} = \alpha + \sum_{i=1}^{s-1} \gamma_{\mathrm{i}}\, \mathrm{D}_{i\mathrm{t}} + \mathrm{u}_{\mathrm{t}}$

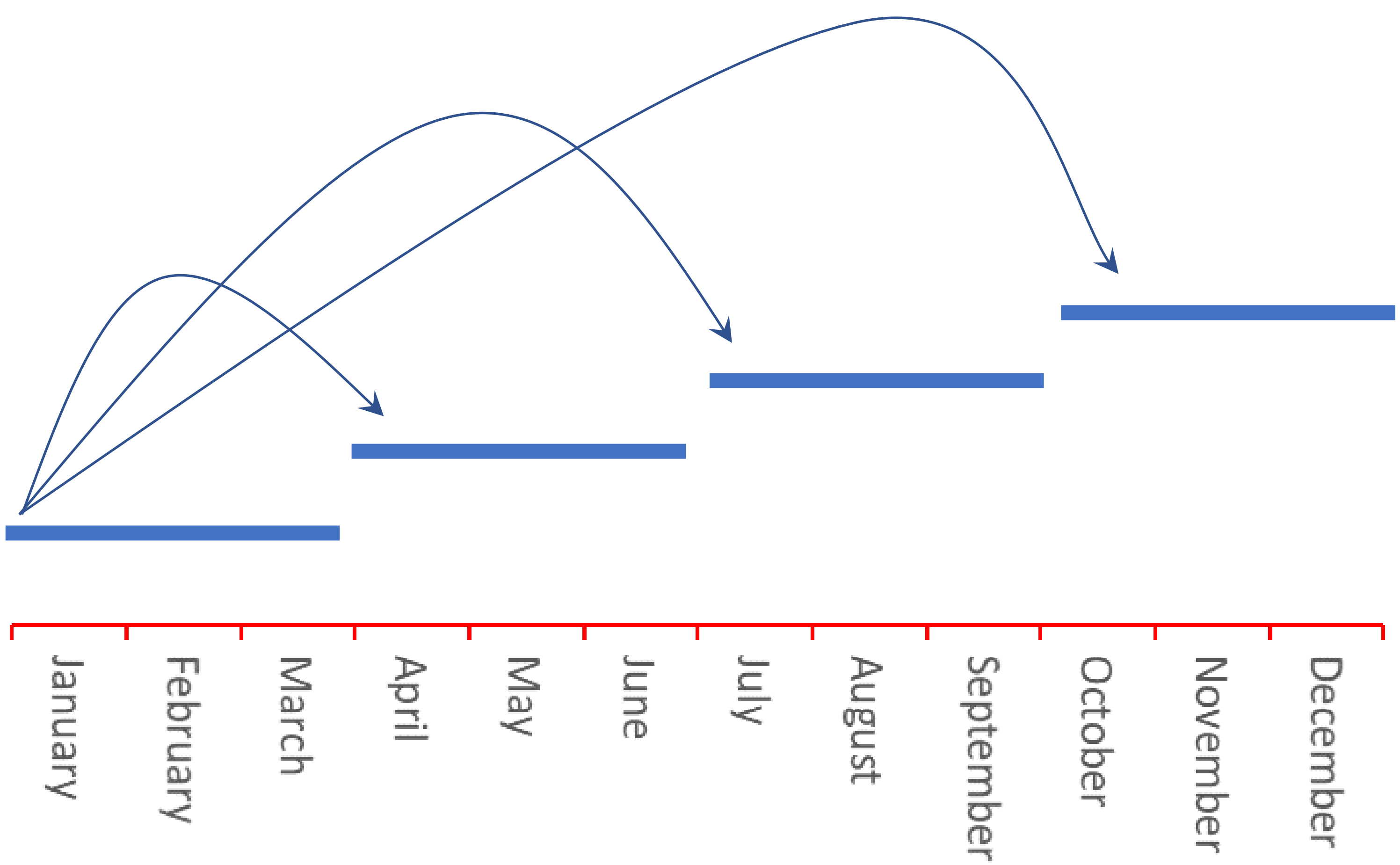


Deterministic seasonality in Stata

- Defining Dum1 for spring, Dum2 for Summer, Dum3 for fall and Dum4 for Winter.

- Estimate the long-run equilibrium using fixed effect, random effect models and mean group.

Deterministic vs. Stochastic Seasonality



Seasonal Unit Roots

- Using seasonal statistical tests such as HEGY-type tests inspired by Hylleberg et al. (1990) and Canova–Hansen (Canova and Hansen, 1995).
- Consider a basic autoregressive polynomial,

$\varphi(B) = 1 - B^s$,
$y = y_t - y_{t-1} = y_t - By_t = (1 - B)y_t$

Where B denotes a lag operator, and “s” is the number of time periods in a seasonal pattern which repeats regularly. S=4 for seasons, S=24 for ??

Seasonality in Economics

Challenges:

- This approach appears to be inaccurate if seasons cannot be identified a priori
- The loss of degrees of freedom
- More sophisticated approach: using machine learning techniques.

Summary

- Overview of weather and climate change effects on food market
  - Any weather related shocks might be passed through along value chain→ food security→ may aggravate climate change effects
- Different proxies of weather related shocks
  - How to measure weather-related shock is an important task in empirical analysis.

Cost pass through in empirical works

- Short-run price adjustment and long-run price equilibrium
- Asymmetric price adjustment and market performance
- Short-run and long-run effects of temperature variation

1. Analysing the effects of temperature

Group members:

Data: Weekly panel dataset from poultry market of Iran.
Do file:
Goal: estimating long-and short-run equilibrium with interaction terms of temperature between farm and wholesale price, and interpret the results.

2. Analyzing the effects of seasonal temperature volatility

Group members:

Data: Weekly panel dataset from poultry market of Iran.
Do file:
Goal: estimating long- and short-run equilibrium with interaction terms of temperature volatility for each seasons between farm and wholesale price, and interpret the results.

3. Analyzing the effects of temperature volatilities

Group members:

Data: Weekly panel dataset from poultry market of Iran.
Do file:
Goal: estimating long-run equilibrium with interaction terms of temperature volatility between farm and wholesale price, and interpret the results.

4. Analyzing the effects of seasonal dummies

Group members:

Data: Weekly panel dataset from poultry market of Iran.
Do file:
Goal: estimating long- and short-run equilibrium with dummy interaction terms between farm and wholesale price, and interpret the results.

Next session

- We will discuss:

Extreme shock thresholds using Threshold Panel Models.

Different volatility measures: SD, ARCH models

Thank you for your attention!

omid.zamani@ae.uni-kiel.de