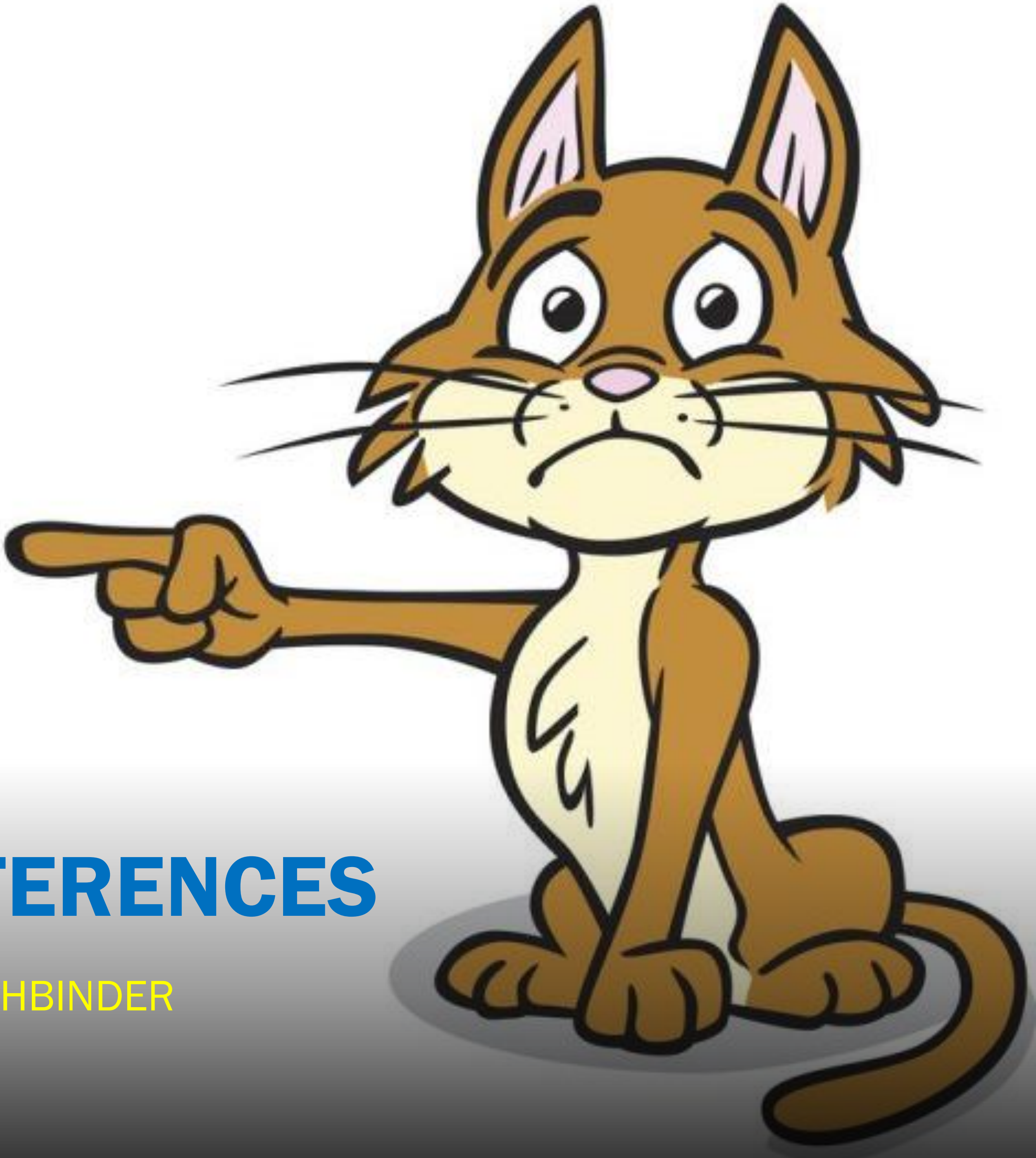


FIRST DIFFERENCES

FELIPE BUCHBINDER

DOES HIGHER INCOME MEAN HIGHER SPENDING?

COULD WE USE POOLED REGRESSION?

$$\text{Spending}_{it} = \beta_0 + \beta_1 \cdot \text{Income}_{it} + U_i + \epsilon_{it}$$

Why or why not?

DOES HIGHER INCOME MEAN HIGHER SPENDING?

POOLED REGRESSION IS NOT A GOOD IDEA

$$\text{Spending}_{it} = \beta_0 + \beta_1 \cdot \text{Income}_{it} + U_i + \epsilon_{it}$$

All these U's correlate with Income:

- Wage
- Investing ability
- Family/Personal fortune
- Ownership of real estate
- Salary negotiation ability
- Ability to "sell" yourself professionally

DOES HIGHER INCOME MEAN HIGHER SPENDING?

POOLED REGRESSION IS NOT A GOOD IDEA

$$\text{Spending}_{it} = \beta_0 + \beta_1 \cdot \text{Income}_{it} + U_i + \epsilon_{it}$$

All these U's correlate with Income:

- Wage
- Investing ability
- Family/Personal fortune
- Ownership of real estate
- Salary negotiation ability
- Ability to "sell" yourself professionally

Flash quiz:

(1) What would happen to the regression estimated coefficients if we *did* run a pooled regression?
(2) Why did we say "all these U's correlate with Income" rather than simply "all these U's affect Spending"?

WHAT IF WE LOOKED AT HOW CHANGES IN INCOME FROM ONE YEAR TO THE NEXT TRANSLATE INTO CHANGES IN SPENDING?

$$\text{Spending}_{it} = \beta_0 + \beta_1 \cdot \text{Income}_{it} + U_i + \epsilon_{it}$$
$$\text{Spending}_{it+1} = \beta_0 + \beta_1 \cdot \text{Income}_{it+1} + U_i + \epsilon_{it+1}$$

WHAT IF WE LOOKED AT HOW CHANGES IN INCOME FROM ONE YEAR TO THE NEXT TRANSLATE INTO CHANGES IN SPENDING?

$$\text{Spending}_{it} = \beta_0 + \beta_1 \cdot \text{Income}_{it} + \cancel{U_i} + \epsilon_{it}$$

$$\text{Spending}_{it+1} = \beta_0 + \beta_1 \cdot \text{Income}_{it+1} + \cancel{U_i} + \epsilon_{it+1}$$

$$\Delta_{\text{t}}\text{Spending}_{it} = \beta_1 \Delta_{\text{t}}\text{Income}_{it} + \epsilon'_{it}$$

Since U_i is doesn't change from one year to the next, a difference in Spending cannot be due to U_i. It must be due to changes in Income.

FIRST DIFFERENCE MODEL

ELIMINATE FIXED UNOBSERVED HETEROGENEITIES BY FOCUSING IN HOW THINGS CHANGE.

Rather than regressing Y on X, regress $\Delta_{\mathrm{t}}Y$ on $\Delta_t X$

- Where $\Delta_{\mathrm{t}}Y \stackrel{\text{def}}{=} Y_{i,t+1} - Y_{it}$ and $\Delta_t X \stackrel{\text{def}}{=} X_{i,t+1} - X_{it}$

A FIRST DIFFERENCES MODEL WORKS JUST LIKE OLS REGRESSION

BUT WITH CHANGES INSTEAD OF ACTUAL VALUES

OLS Regression

- $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \epsilon$
- $\beta = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$
- The assumptions of normality, homoskedasticity and independence must be followed by ϵ

First Differences Regression

- $\boldsymbol{\Delta}_t\mathbf{Y} = \boldsymbol{\Delta}_t\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\Delta}_t\epsilon$
- $\beta = (\boldsymbol{\Delta}_t\mathbf{X}'\boldsymbol{\Delta}_t\mathbf{X})^{-1}\boldsymbol{\Delta}_t\mathbf{X}'\boldsymbol{\Delta}_t\mathbf{Y}$
- The assumptions of normality, homoskedasticity and independence must be followed by $\Delta_t\epsilon$

IS INVESTMENT DETERMINED BY COMPANY VALUE? THE GRUNFELD DATASET

#Pooled regression

pooled <- plm(invest ~ value + capital, index=c(“firm”, “year”), data=Grunfeld, model='pooling’)

#First Differences model

fd <- plm(invest ~ value + capital, index=c(“firm”, “year”), data=Grunfeld, model=‘fd’)

	Dependent variable:	
	invest	
	Pooled	First Differences
	(1)	(2)
value	0.115***	0.090***
	(0.006)	(0.008)
capital	0.228***	0.291***
	(0.024)	(0.051)
Constant	-38.410***	-1.654
	(8.413)	(3.200)
Observations	220	209
R2	0.818	0.411
Adjusted R2	0.816	0.405
F Statistic	487.284*** (df = 2; 217)	71.756*** (df = 2; 206)
Note:		*p<0.1; **p<0.05; ***p<0.01

IS INVESTMENT DETERMINED BY COMPANY VALUE?
THE GRUNFELD DATASET

#Pooled regression

```
pooled <- plm(invest ~ value + capital, index=c("firm", "year"), data=Grunfeld, model='pooling')
```

#First Differences model

```
fd <- plm(invest ~ value + capital, index=c("firm", "year"), data=Grunfeld, model='fd')
```

	Dependent variable: invest	
	Pooled	First Differences
	(1)	(2)
value	0.115***	0.090***
	(0.006)	(0.008)
capital	0.2[illegible]	[illegible]
	([illegible]	[illegible]
Constant	-3[illegible]	[illegible]4
	(8.41[illegible]	([illegible].200)
Observations	220	209
R2	0.818	0.411
Adjusted R2	0.816	0.405
F Statistic	487.284*** (df = 2; 217)	71.756*** (df = 2; 206)
Note:		*p<0.1; **p<0.05; ***p<0.01

Why do we have less observations in FD than in PR?

IS INVESTMENT DETERMINED BY COMPANY VALUE? THE GRUNFELD DATASET

#Pooled regression

pooled <- plm(invest ~ value + capital, index=c(“firm”, “year”), data=Grunfeld, model='pooling’)

#First Differences model

fd <- plm(invest ~ value + capital, index=c(“firm”, “year”), data=Grunfeld, model=‘fd’)

	Dependent variable: invest	
	Pooled (1)	First Differences (2)
value	0.115***	0.090***
	$\boxed{(0.006)}$	$\boxed{(0.008)}$
capital	0.228***	0.291***
	$\boxed{(0.024)}$	$\boxed{(0.051)}$
Constant	-38.41[illegible]	[illegible]654
	[illegible]	[illegible]
Observations	[illegible]	[illegible]09
R2	[illegible]	0.411
Adjusted R2	0.816	0.405
F Statistic	487.284*** (df = 2; 217)	71.756*** (df = 2; 206)
Note:		*p<0.1; **p<0.05; ***p<0.01

Why are the coefficient’s standard errors larger in FD than in PR?

A TRICK FIRST DIFFERENCES REGRESSION CAN'T DO:

ACCOUNT FOR THINGS THAT DO NOT CHANGE IN TIME

(WHY?)

FIRST DIFFERENCE MODELS CANNOT ACCOUNT FOR TIME-INVARIANT COVARIATES

$$\text{Spending}_{it} = \beta_0 + \beta_1 \cdot \text{Income}_{it} + \cancel{\text{Gender}_i} + U_i + \epsilon_{it}$$

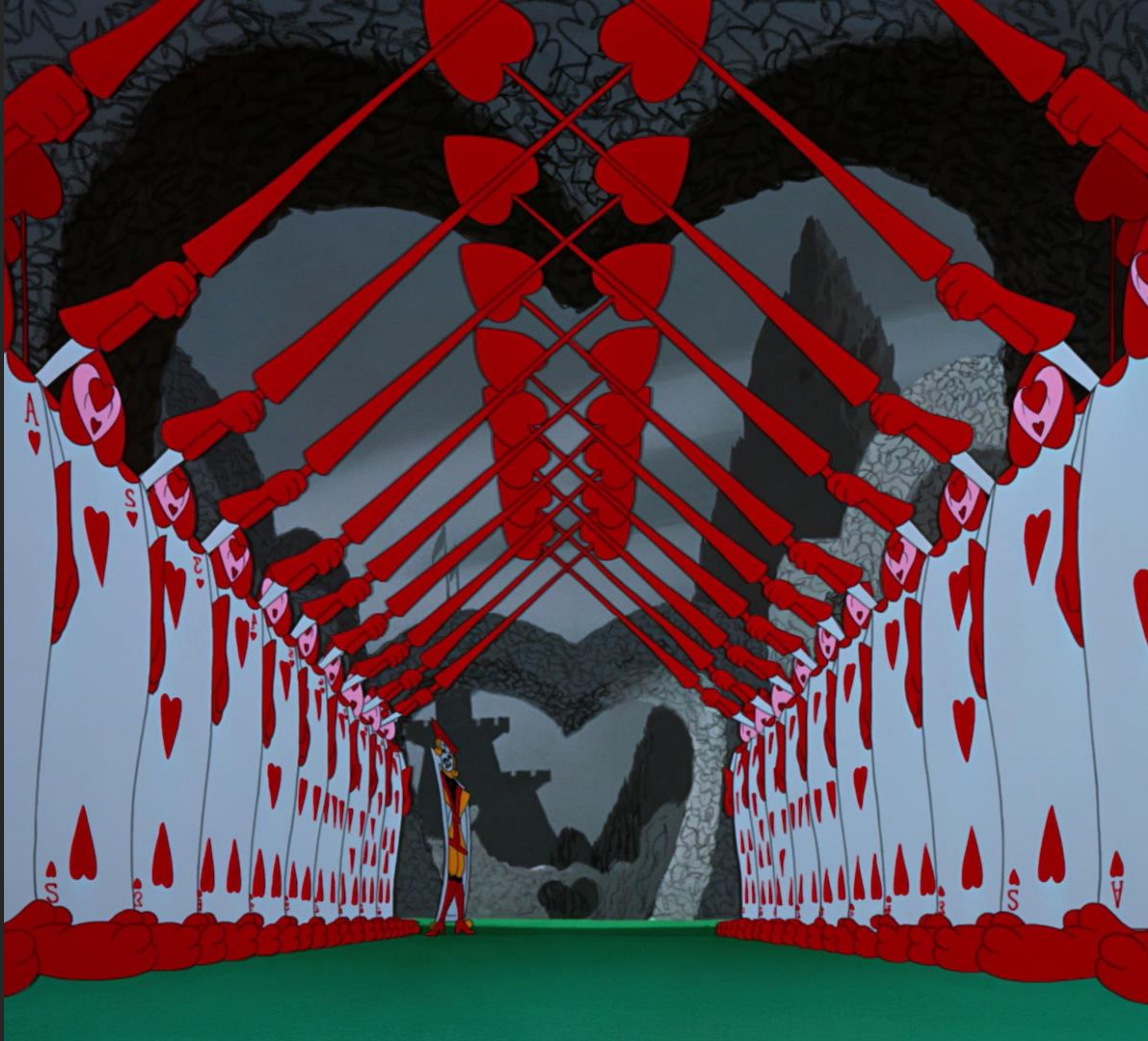
$$\text{Spending}_{it+1} = \beta_0 + \beta_1 \cdot \text{Income}_{it+1} + \cancel{\text{Gender}_i} + U_i + \epsilon_{it+1}$$

$$\Delta_{\text{t}}\text{Spending}_{it} = \beta_1 \Delta_{\text{t}}\text{Income}_{it} + 0 + \epsilon'_{it}$$

Since Gender_i doesn't change from one year to the next, it cancels out when we calculate $\Delta_t\text{Spending}_{it}$.
As a result, we cannot estimate the effect that Gender has on Spending.

FIRST DIFFERENCES REGRESSION REQUIRES THE $\Delta_t \epsilon_{it}$ TO BE SERIALLY UNCORRELATED

(WHY?)




THIS *DOES NOT HAPPEN* IF THE ERRORS THEMSELVES
ARE SERIALLY UNCORRELATED

If errors are uncorrelated in the true relationship, then a First Differences model does not satisfy the assumption of serially uncorrelated residuals. The First Difference model is not a good model when the errors in the true relationship are serially uncorrelated.

PROOF

Problem Statement:

The true relationship is

$$y_{it} = \mathrm{X}_{\mathrm{it}}\beta + U_i + \epsilon_{it}$$

Suppose idiosyncratic errors are serially uncorrelated and homoeskedastic. This implies, for all t:

$$\mathrm{Cov}(\epsilon_{t+1}, \epsilon_t) = 0 \text{ (No serial correlation)}$$
$$\mathrm{Cov}(\epsilon_t, \epsilon_t) = \sigma^2 \text{ (Homoskedasticity)}$$

In a First Differences model, the erros will be $\Delta\epsilon_{it}$. Are these$\Delta\epsilon_{it}$ also serially uncorrelated? In other words,

Is $\mathrm{Cov}(\Delta\epsilon_{t+1}, \Delta\epsilon_t)$ also zero?

PROOF

$$\mathrm{Cov}(\Delta\epsilon_{t+1}, \Delta\epsilon_t) = \mathrm{Cov}(\epsilon_{t+1} - \epsilon_t\ ;\ \epsilon_t - \epsilon_{t-1})$$

$$= \underbrace{\mathrm{Cov}(\epsilon_{t+1}; \epsilon_t)}_{0} + \underbrace{\mathrm{Cov}(\epsilon_{t+1}; \epsilon_{t-1})}_{0} - \underbrace{\mathrm{Cov}(\epsilon_t; \epsilon_t)}_{\sigma^2} + \underbrace{\mathrm{Cov}(\epsilon_{t+1}; \epsilon_{t-1})}_{0} = \sigma^2$$

So

$$\mathrm{Cov}(\Delta\epsilon_{t+1}, \Delta\epsilon_t) \neq 0$$

Meaning that if errors are uncorrelated in the true relationship, then a First Differences model does not satisfy the assumption of serially uncorrelated residuals. The First Difference model is *not* a good model when the errors in the true relationship are serially uncorrelated.

If errors are uncorrelated in the true relationship, then a First Differences model does not satisfy the assumption of serially uncorrelated residuals. The First Difference model is not a good model when the errors in the true relationship are serially uncorrelated.

The First Difference model works well when errors in the true relationship obey a random walk.

ERRORS ARE SAID TO FOLLOW A RANDOM WALK WHEN

$$\epsilon_{t+1} = \epsilon_t + \nu_t$$

$$\nu_t \sim \mathrm{N}(0; \sigma^2)$$

$$\mathrm{COV}(\nu_t; \nu_s) = 0 \; \forall \, t \neq s$$

THIS WORKS BECAUSE…

$$\epsilon_{t+1} = \epsilon_t + \nu_t$$
$$\Downarrow$$
$$\Delta_t \epsilon_t = \nu_t$$

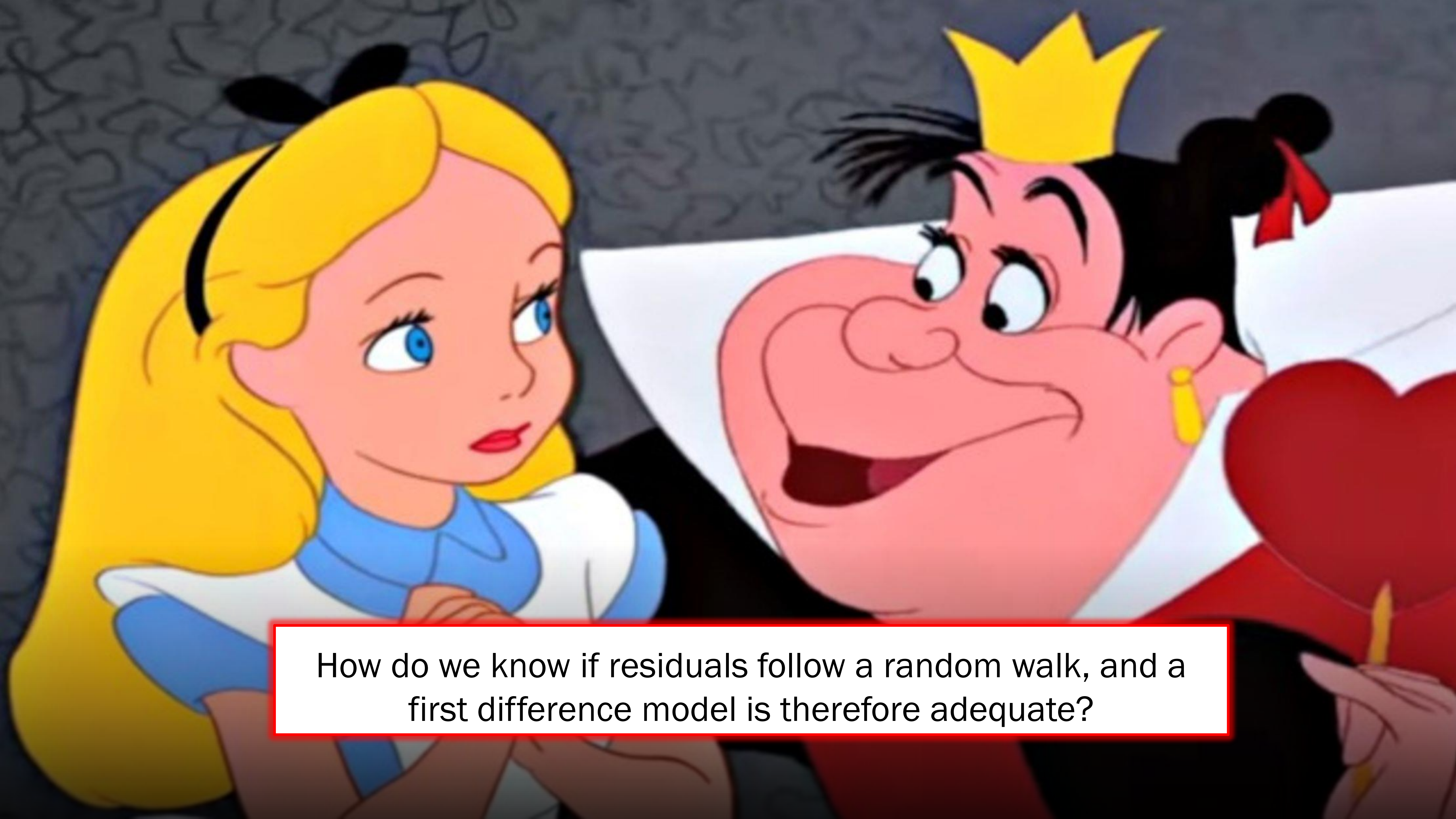
Which are homoscedastic and serially uncorrelated by hypothesis, thus satisfying the requirements of OLS regression.

ONE THING YOU SHOULD KNOW BEFORE TAKING A RANDOM WALK…

$$\epsilon_{t+1} = \epsilon_t + \nu_t$$

Random Walks exhibit long-term dependencies. Because the effect of a past ϵ_t never vanishes…

Does this make theoretical sense in your research problem?



How do we know if residuals follow a random walk, and a first difference model is therefore adequate?

HOW DO WE KNOW IF ERRORS FOLLOW A RANDOM WALK?

INTUITION

1. Regress $\Delta\epsilon_t$ on$\Delta\epsilon_{t-1}$ i.e.

$$\Delta\epsilon_t = \rho\Delta\epsilon_{t-1} + \nu_t$$
$$\nu_t \sim N(0;\sigma^2)$$

2. Test if linear coefficient is statistically significant i.e.

$$\begin{cases} H_0\colon \rho = 0 \\ H_a\colon \rho \neq 0 \end{cases}$$

HOW DO WE KNOW IF ERRORS FOLLOW A RANDOM WALK?

INTUITION

1. Regress $\Delta\epsilon_t$ on$\Delta\epsilon_{t-1}$ i.e.

$$\Delta\epsilon_t = \rho\Delta\epsilon_{t-1} + \nu_t$$
$$\nu_t \sim N(0;\sigma^2)$$

2. Test if linear coefficient is statistically significant i.e.

$$\begin{cases} H_0: \rho = 0 \\ H_a: \rho \neq 0 \end{cases}$$

Tricky question:

Why do we test $\Delta\epsilon_t$, rather than ϵ_t itself?

HOW DO WE KNOW IF ERRORS FOLLOW A RANDOM WALK?

THE DURBIN-WHATSON STATISTIC

1. Obtain $\Delta\epsilon_t$ and$\Delta\epsilon_{t-1}$
2. Calculate the Durbin-Whatson Statistic

$$DW = \frac{\sum_t (\Delta\epsilon_t - \Delta\epsilon_{t-1})^2}{\sum_t (\Delta\epsilon_t)^2}$$

3. Compare with critical values proper from this Statistic. As a rule of thumb,

$$DW \approx 2\left(1 - \rho_{\Delta\epsilon_t, \Delta\epsilon_{t-1}}\right)$$

So if $\Delta\epsilon_t$ and $\Delta\epsilon_{t-1}$ are serially uncorrelated,

$$DW \approx 2.$$

HOW DO WE KNOW IF ERRORS FOLLOW A RANDOM WALK?

THE DURBIN-WHATSON STATISTIC

1. Obtain $\Delta\epsilon_t$ and$\Delta\epsilon_{t-1}$
2. Calculare the Durbin-Whatson Statistic

$$DW = \frac{\sum_t(\Delta\epsilon_t - \Delta\epsilon_{t-1})^2}{\sum_t(\Delta\epsilon_t)^2}$$

3. Compare with critical values proper from this Statistic. As a rule of thumb,

$$DW \approx 2\left(1 - \rho_{\Delta\epsilon_t,\Delta\epsilon_{t-1}}\right)$$

So if $\Delta\epsilon_t$ and $\Delta\epsilon_{t-1}$ are serially uncorrelated,

$$DW \approx 2.$$

Flash quiz:

What are the maximum and minimum values for the DW statistic?

HOW DO WE KNOW IF ERRORS FOLLOW A RANDOM WALK?

THE DURBIN-WHATSON STATISTIC

1. Obtain $\Delta\epsilon_t$ and$\Delta\epsilon_{t-1}$
2. Calculare the Durbin-Whatson Statistic

$$DW = \frac{\sum_t (\Delta\epsilon_t - \Delta\epsilon_{t-1})^2}{\sum_t (\Delta\epsilon_t)^2}$$

3. Compare with critical values proper from this Statistic. As a rule of thumb,

$$DW \approx 2\left(1 - \rho_{\Delta\epsilon_t, \Delta\epsilon_{t-1}}\right)$$

So if $\Delta\epsilon_t$ and $\Delta\epsilon_{t-1}$ are serially uncorrelated,

$$DW \approx 2.$$

Flash quiz:

A rule of thumb says that $DW < 1$ means trouble.
To what value of ρ does this correspond?

PROOF THAT $D \approx 2\left(1 - \rho_{\Delta\epsilon_t, \Delta\epsilon_{t-1}}\right)$

$$DW = \frac{\sum_t (\Delta\epsilon_t - \Delta\epsilon_{t-1})^2}{\sum_t (\Delta\epsilon_t)^2} = \frac{\sum_t (\Delta\epsilon_t)^2}{\sum_t (\Delta\epsilon_t)^2} - 2\frac{\sum_t \Delta\epsilon_t \Delta\epsilon_{t-1}}{\sum_t (\Delta\epsilon_t)^2} + \frac{\sum_t (\Delta\epsilon_{t-1})^2}{\sum_t (\Delta\epsilon_t)^2}$$

Note that $\sum_t (\Delta\epsilon_t)^2 \approx \sum_t (\Delta\epsilon_{t-1})^2 \approx \sqrt{\sum_t \Delta\epsilon_t^2 \sum \Delta\epsilon_{t-1}^2}$. So

$$DW \approx \underbrace{\frac{\sum_t (\Delta\epsilon_t)^2}{\sum_t (\Delta\epsilon_t)^2}}_{1} - 2\underbrace{\frac{\sum_t \Delta\epsilon_t \Delta\epsilon_{t-1}}{\sqrt{\sum_t \Delta\epsilon_t^2 \sum \Delta\epsilon_{t-1}^2}}}_{\rho_{\Delta\epsilon_t, \Delta\epsilon_{t-1}}} + \underbrace{\frac{\sum_t (\Delta\epsilon_{t-1})^2}{\sum_t (\Delta\epsilon_{t-1})^2}}_{1}$$

$$DW \approx 2 - 2\rho_{\Delta\epsilon_t, \Delta\epsilon_{t-1}}$$

$$DW \approx 2\left(1 - \rho_{\Delta\epsilon_t, \Delta\epsilon_{t-1}}\right)$$

KEY TAKEAWAYS

1. First Difference Models are a way to eliminate the unobserved effect U_i in panel data regression
2. In First Difference models, rather than using X_{it} to explain y_{it}, we use changes in X_{it} to explain changes in y_{it}.
3. In other words, we regress $\Delta_t y_{it}$ on $\Delta_t X_{it}$
4. First Difference models work well when the idiosyncratic error of y_{it} follows a Random Walk. Unfortunately, we don't see the values of the idiosyncratic errors. We only see their *deltas*
5. One way to test if ϵ_{it} follows a random walk is to regress $\Delta_t \epsilon_{it}$ on $\Delta_t \epsilon_{it-1}$ (no intercept needed) and see if the coefficient is statistically indistinguishable from zero.
6. Another way is to use the Durbin-Whatson Statistic. As a rule of thumb, it should ideally be close to 2.