POLS/CSSS 503

Advanced Quantitative Political Methodology

Introduction to Panel Data Analysis

Christopher Adolph*

University of Washington, Seattle

June 2, 2011

*Assistant Professor, Department of Political Science and Center for Statistics and the Social Sciences.

Panel Data Structure

Suppose we observe our response over both time and place:

$$y_{it} = \mathrm{x}_{it}\boldsymbol{\beta} + \varepsilon_{it}$$

We have units $i = 1, \ldots, N$, each observed over periods $t = 1, \ldots, T$, for a total of $N \times T$ observations

Panel Data Structure

Suppose we observe our response over both time and place:

$$y_{it} = \mathrm{x}_{it}\boldsymbol{\beta} + \varepsilon_{it}$$

We have units $i = 1, \ldots, N$, each observed over periods $t = 1, \ldots, T$, for a total of $N \times T$ observations

Balanced data: all units i have the same number of observations T.

Panel Data Structure

Suppose we observe our response over both time and place:

$$y_{it} = \mathrm{x}_{it}\boldsymbol{\beta} + \varepsilon_{it}$$

We have units $i = 1, \ldots, N$, each observed over periods $t = 1, \ldots, T$, for a total of $N \times T$ observations

Balanced data: all units i have the same number of observations T.

Unbalanced data: some units are shorter in T, perhaps due to missing data, perhaps to sample selection

All of our discussion in class will assume balanced panels.

Small adjustments may be needed for unbalanced panels,
unless the imbalance is due to sample selection, which could lead to significant bias.

You say Panel, I say TSCS. . .

Usages of the term *panel data* vary by field and sub-field

1. Data with large $N \approx 1000$ and small $T \approx 5$ (esp. in economics)

You say Panel, I say TSCS. . .

Usages of the term *panel data* vary by field and sub-field

1. Data with large $N \approx 1000$ and small $T \approx 5$ (esp. in economics)

2. Data with any N, T, and repeated observations on units $i = 1, \ldots, N$ (esp. in opinion research)

You say Panel, I say TSCS. . .

Usages of the term *panel data* vary by field and sub-field

1. Data with large $N \approx 1000$ and small $T \approx 5$ (esp. in economics)

2. Data with any N, T, and repeated observations on units $i = 1, \ldots, N$ (esp. in opinion research)

3. Any data with both $N > 1$ and $T > 1$ (sometimes in political science)

You say Panel, I say TSCS. . .

Usages of the term TSCS data vary by field and sub-field

1. Data with small $N \approx 20$ and medium to large $T > 15$ (esp. in political science)

You say Panel, I say TSCS. . .

Usages of the term TSCS data vary by field and sub-field

1. Data with small $N \approx 20$ and medium to large $T > 15$ (esp. in political science)

2. Data with any N, T, but each cross section has new units;
   so i in period t is a different person from i in period $t + 1$ (esp. opinion research)

You say Panel, I say TSCS. . .

Usages of the term TSCS data vary by field and sub-field

1. Data with small $N \approx 20$ and medium to large $T > 15$ (esp. in political science)

2. Data with any N, T, but each cross section has new units;
   so i in period t is a different person from i in period $t+1$ (esp. opinion research)

3. Any data with both $N > 1$ and $T > 1$

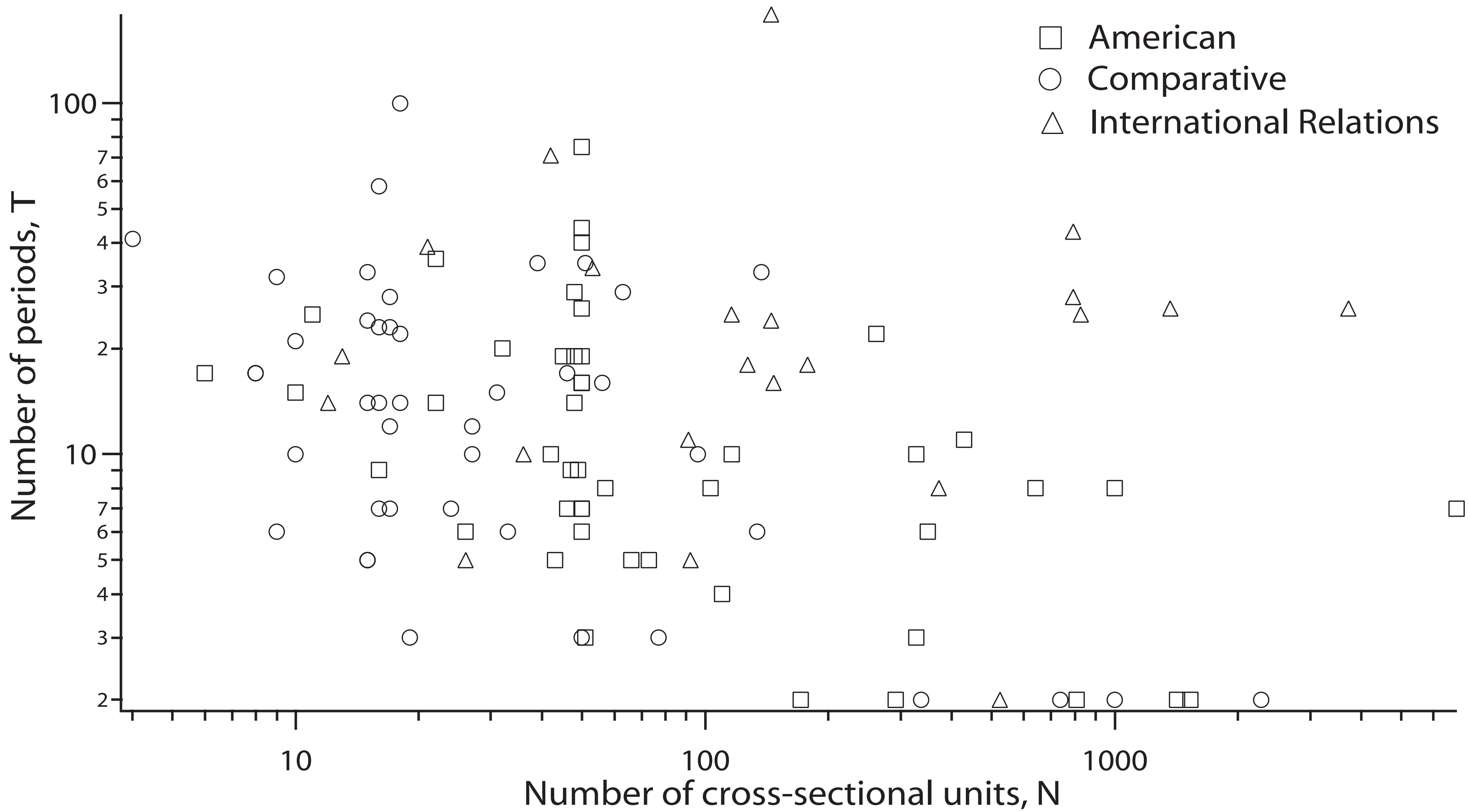
You say Panel, I say TSCS. . .

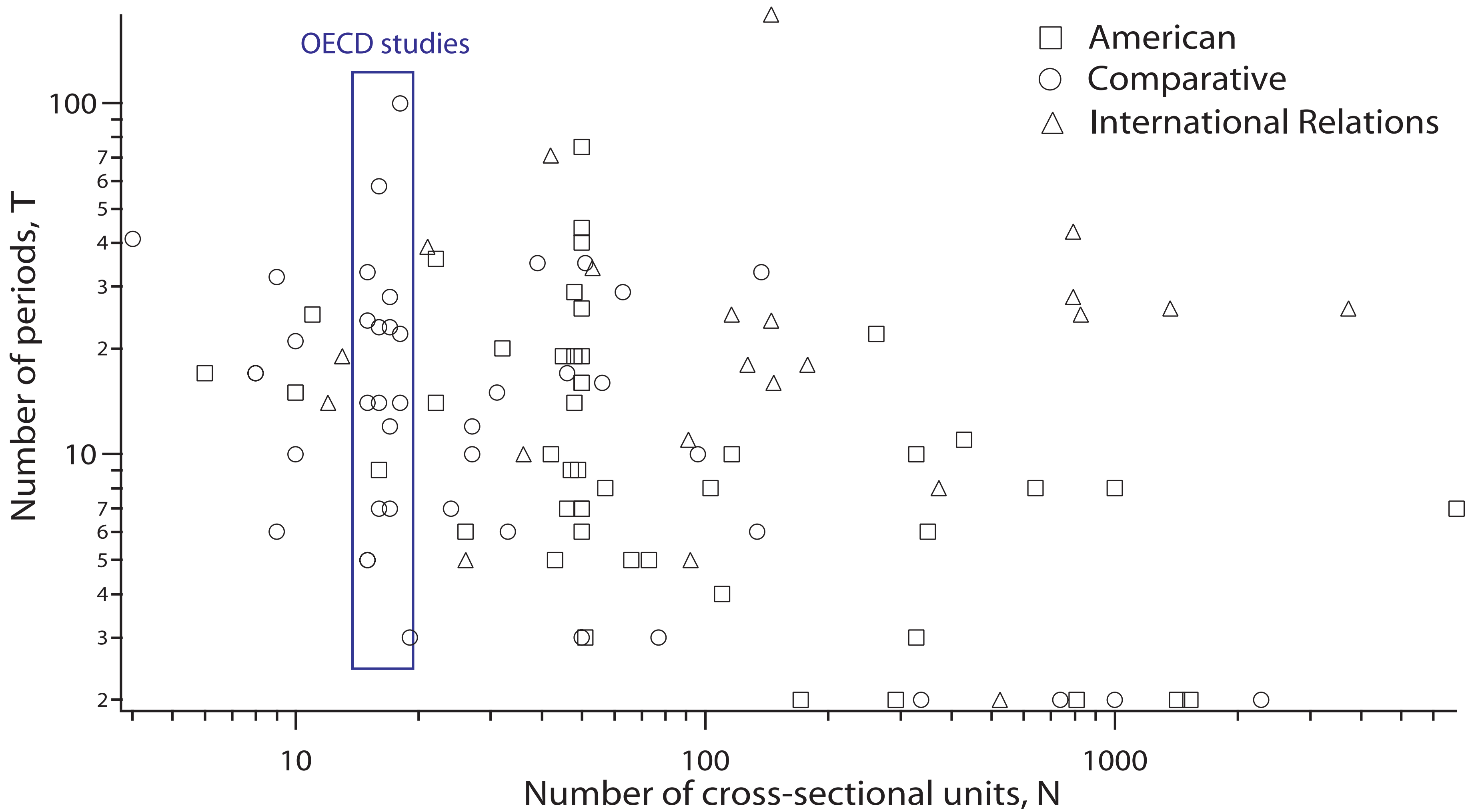
Data with large N and small T offer different problems and opportunities compared to data with small N and medium T

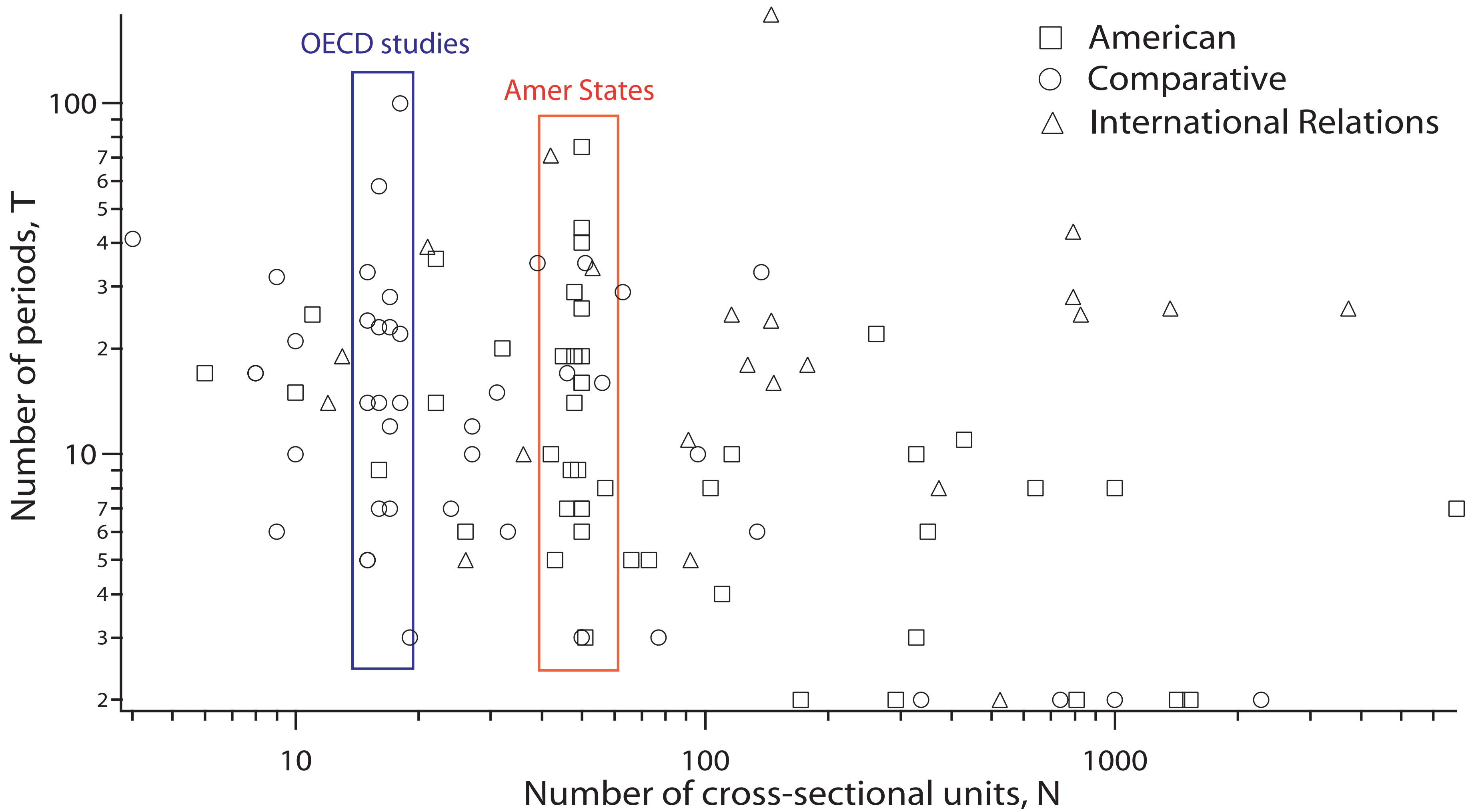
Beware blanket statements about *panel estimators* or *panel data*.

The author—even in a textbook—may be assuming an N and T ubiquitous in his field, but uncommon in yours!

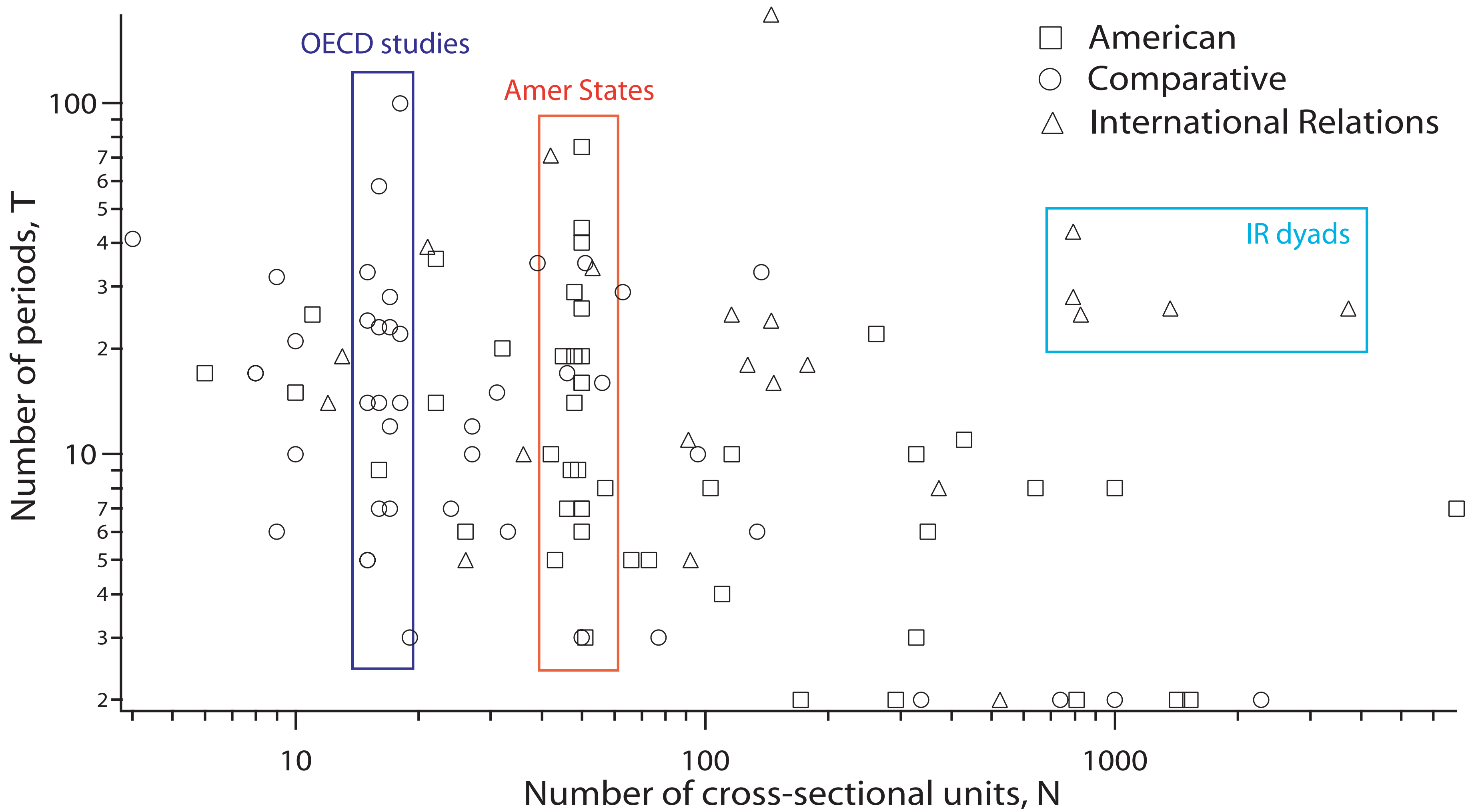
Especially a problem for comparativists learning from econometrics texts

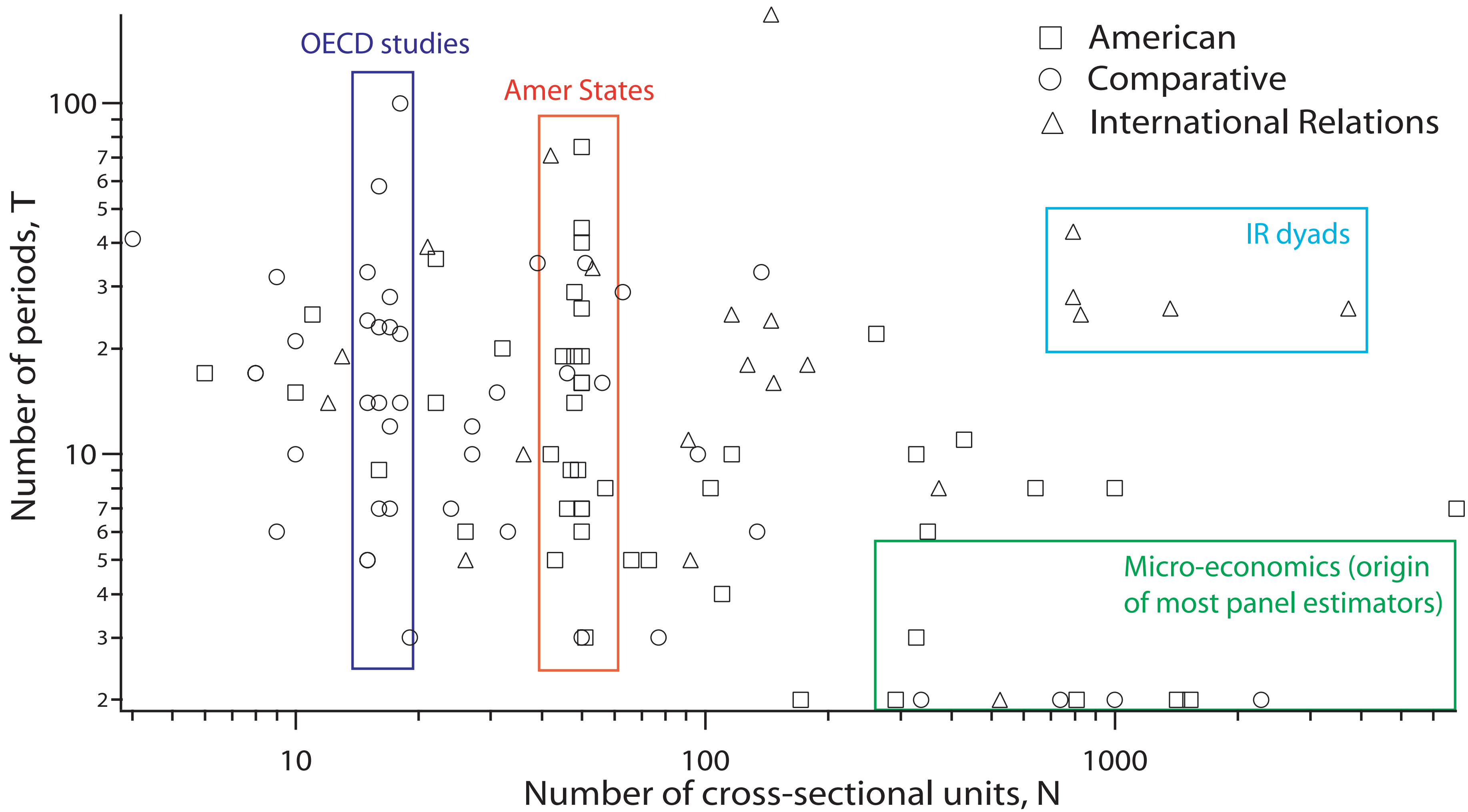

American
Comparative
International Relations
Number of periods, T
100
10
7
6
5
4
3
2
Number of cross-sectional units, N
10
100
1000


OECD studies
American
Comparative
International Relations
Number of periods, T
Number of cross-sectional units, N
100
10
7
6
5
4
3
2
1000



OECD studies
Amer States
American
Comparative
International Relations
Number of periods, T
Number of cross-sectional units, N
100
10
7
6
5
4
3
2
1000


OECD studies
Amer States
IR dyads
American
Comparative
International Relations
Number of periods, T
Number of cross-sectional units, N
100
10
1000



OECD studies
Amer States
IR dyads
Micro-economics (origin of most panel estimators)
American
Comparative
International Relations
Number of periods, T
Number of cross-sectional units, N
100
10
7
6
5
4
3
2
1000

A pooled TSCS model

$$\mathrm{GDP}_{it} = \phi_1 \mathrm{GDP}_{i,t-1} + \beta_0 + \beta_1 \mathrm{Democracy}_{it} + \varepsilon_{it}$$

This model assumes the same effect of Democracy on GDP for all countries i (β_1)

And influence of past GDP on current GDP is the same for all countries i (ϕ_1)

The shared parameters make this a *Pooled* Time Series Cross Section model

Data storage issues

To get panel data ready for analysis, we need it *stacked* by unit and time period, with a time variable and a grouping variable included:

Cty	Year	GDP	lagGDP	Democracy
1	1962	5012	NA	0
1	1963	6083	5012	0
1	1964	6502	6083	0
...				
1	1989	12530	12266	0
1	1990	12176	12530	0
2	1975	1613	NA	NA
2	1976	1438	1613	0
...				
135	1989	6575	6595	0
135	1990	6450	6575	0

Data storage issues

To get panel data ready for analysis, we need it *stacked* by unit and time period, with a time variable and a grouping variable included:

Cty	Year	GDP	lagGDP	Democracy
1	1962	5012	NA	0
1	1963	6083	5012	0
1	1964	6502	6083	0
...				
1	1989	12530	12266	0
1	1990	12176	12530	0
2	1975	1613	NA	NA
2	1976	1438	1613	0
...				
135	1989	6575	6595	0
135	1990	6450	6575	0

Don't use `lag()` to create lags in panel data!

You need a panel lag command that accounts for the breaks where the unit changes, such as `lagpanel()` in the `simcf` package.

Why use Panel Data?

- More data, which might make inference more precise (at least if we believe β is the same or similar across units)

Why use Panel Data?

- More data, which might make inference more precise
  (at least if we believe β is the same or similar across units)

- Can help with omitted variables, especially if they are time invariant

Why use Panel Data?

- More data, which might make inference more precise
  (at least if we believe β is the same or similar across units)

- Can help with omitted variables, especially if they are time invariant

- Some analysis only possible with panel data;
  e.g., if variables don't change much over time, like institutions

Why use Panel Data?

- More data, which might make inference more precise
  (at least if we believe β is the same or similar across units)

- Can help with omitted variables, especially if they are time invariant

- Some analysis only possible with panel data;
  e.g., if variables don't change much over time, like institutions

- Heterogeneity is interesting! As long as we can specify a general DGP for whole panel, can parameterize and estimate more substantively interesting relationships

Why use Panel Data?

If modeled correctly, costs of panel data are born by researcher, not by model or data:

- Differences across the panel would appear the biggest problem,
  but we can relax any homogeneity assumption to get a more flexible panel model

Why use Panel Data?

If modeled correctly, costs of panel data are born by researcher, not by model or data:

- Differences across the panel would appear the biggest problem,
  but we can relax any homogeneity assumption to get a more flexible panel model

- The price of panel data is a more complex structure to conceptualize and model

Why use Panel Data?

If modeled correctly, costs of panel data are born by researcher, not by model or data:

- Differences across the panel would appear the biggest problem,
  but we can relax any homogeneity assumption to get a more flexible panel model

- The price of panel data is a more complex structure to conceptualize and model

- Often need more powerful or flexible estimation tools

Building Time Series into Panel

Consider the ARIMA(p,d,q) model:

$$\Delta^d y_t = \alpha + \mathrm{x}_t \boldsymbol{\beta} + \sum_{p=1}^{P} \Delta^d y_{t-p} \phi_p + \sum_{q=1}^{Q} \varepsilon_{t-q} \rho_q + \varepsilon_t$$

where $\varepsilon \sim N(0, \sigma^2)$ is white noise.

A "mother" specification for all our time series processes.

Includes as special cases:

Building Time Series into Panel

Consider the ARIMA(p,d,q) model:

$$\Delta^d y_t = \alpha + \mathrm{x}_t \boldsymbol{\beta} + \sum_{p=1}^{P} \Delta^d y_{t-p} \phi_p + \sum_{q=1}^{Q} \varepsilon_{t-q} \rho_q + \varepsilon_t$$

where $\varepsilon \sim N(0, \sigma^2)$ is white noise.

A "mother" specification for all our time series processes.

Includes as special cases:

ARMA(p,q) models: Set $d = 0$

Building Time Series into Panel

Consider the ARIMA(p,d,q) model:

$$\Delta^d y_t = \alpha + \mathrm{x}_t\boldsymbol{\beta} + \sum_{p=1}^{P} \Delta^d y_{t-p}\phi_p + \sum_{q=1}^{Q} \varepsilon_{t-q}\rho_q + \varepsilon_t$$

where $\varepsilon \sim N(0, \sigma^2)$ is white noise.

A "mother" specification for all our time series processes.

Includes as special cases:

ARMA(p,q) models: Set $d = 0$

AR(p) models: Set $d = Q = 0$

Building Time Series into Panel

Consider the ARIMA(p,d,q) model:

$$\Delta^d y_t = \alpha + \mathrm{x}_t \boldsymbol{\beta} + \sum_{p=1}^{P} \Delta^d y_{t-p}\phi_p + \sum_{q=1}^{Q} \varepsilon_{t-q}\rho_q + \varepsilon_t$$

where $\varepsilon \sim N(0, \sigma^2)$ is white noise.

A "mother" specification for all our time series processes.

Includes as special cases:

ARMA(p,q) models: Set $d = 0$

AR(p) models: Set $d = Q = 0$

MA(q) models: Set $d = P = 0$

Building Time Series into Panel

Consider the ARIMA(p,d,q) model:

$$\Delta^d y_t = \alpha + \mathrm{x}_t\boldsymbol{\beta} + \sum_{p=1}^{P} \Delta^d y_{t-p}\phi_p + \sum_{q=1}^{Q} \varepsilon_{t-q}\rho_q + \varepsilon_t$$

where $\varepsilon \sim N(0, \sigma^2)$ is white noise.

A "mother" specification for all our time series processes.

Includes as special cases:

ARMA(p,q) models: Set $d = 0$

AR(p) models: Set $d = Q = 0$

MA(q) models: Set $d = P = 0$

Linear regression: Set $d = P = Q = 0$

Could even be re-written as an error correction model

Multiple Time Series

Now notice that if we had several parallel time series $y_{1t}, y_{2t}, \ldots y_{Nt}$, as for N countries, we could estimate a series of regression models:

$$\Delta^{d_1} y_{1t} = \alpha_1 + \mathrm{x}_{1t}\boldsymbol{\beta}_1 + \sum_{p=1}^{P_1} \Delta^{d_1} y_{1,t-p}\phi_{1p} + \sum_{q=1}^{Q_1} \varepsilon_{1,t-q}\rho_{1q} + \varepsilon_{1t}$$

Multiple Time Series

Now notice that if we had several parallel time series $y_{1t}, y_{2t}, \ldots y_{Nt}$, as for N countries, we could estimate a series of regression models:

$$\Delta^{d_1} y_{1t} = \alpha_1 + \mathrm{x}_{1t}\boldsymbol{\beta}_1 + \sum_{p=1}^{P_1} \Delta^{d_1} y_{1,t-p}\phi_{1p} + \sum_{q=1}^{Q_1} \varepsilon_{1,t-q}\rho_{1q} + \varepsilon_{1t}$$

$$\Delta^{d_2} y_{2t} = \alpha_2 + \mathrm{x}_{2t}\boldsymbol{\beta}_2 + \sum_{p=2}^{P_2} \Delta^{d_2} y_{2,t-p}\phi_{2p} + \sum_{q=2}^{Q_2} \varepsilon_{2,t-q}\rho_{2q} + \varepsilon_{2t}$$

Multiple Time Series

Now notice that if we had several parallel time series $y_{1t}, y_{2t}, \ldots y_{Nt}$, as for N countries, we could estimate a series of regression models:

$$
\begin{aligned}
\Delta^{d_1} y_{1t} &= \alpha_1 + \mathrm{x}_{1t}\boldsymbol{\beta}_1 + \sum_{p=1}^{P_1} \Delta^{d_1} y_{1,t-p}\phi_{1p} + \sum_{q=1}^{Q_1} \varepsilon_{1,t-q}\rho_{1q} + \varepsilon_{1t} \\
\Delta^{d_2} y_{2t} &= \alpha_2 + \mathrm{x}_{2t}\boldsymbol{\beta}_2 + \sum_{p=2}^{P_2} \Delta^{d_2} y_{2,t-p}\phi_{2p} + \sum_{q=2}^{Q_2} \varepsilon_{2,t-q}\rho_{2q} + \varepsilon_{2t} \\
&\ldots \\
\Delta^{d_N} y_{Nt} &= \alpha_N + \mathrm{x}_{Nt}\boldsymbol{\beta}_N + \sum_{p=N}^{P_N} \Delta^{d_N} y_{N,t-p}\phi_{Np} + \sum_{q=N}^{Q_N} \varepsilon_{N,t-q}\rho_{Nq} + \varepsilon_{Nt}
\end{aligned}
$$

Each of these models could be estimated separately

Multiple Time Series

The results would be a panel analysis of a particular kind:

- one with maximum flexibility for heterogeneous data generating processes across units i,
- and no borrowing of strength across units i

Generally, we can write this series of regression models as:

$$\Delta^{d_i} y_{it} = \alpha_i + \mathrm{x}_{it}\boldsymbol{\beta}_i + \sum_{p=1}^{P_i} \Delta^{d_i} y_{i,t-p}\phi_{ip} + \sum_{q=1}^{Q_i} \varepsilon_{i,t-q}\rho_{iq} + \varepsilon_{it}$$

We've just written all our time series equations in a single matrix

But estimation is still *separate* for each equation

Be clear what the subscripts and variables are

$$\Delta^{d_i} y_{it} = \alpha_i + \mathrm{x}_{it}\boldsymbol{\beta}_i + \sum_{p=1}^{P_i} \Delta^{d_i} y_{i,t-p}\phi_{ip} + \sum_{q=1}^{Q_i} \varepsilon_{i,t-q}\rho_{iq} + \varepsilon_{it}$$

Be clear what the subscripts and variables are

$$\Delta^{d_i} y_{it} = \alpha_i + \mathrm{x}_{it}\boldsymbol{\beta}_i + \sum_{p=1}^{P_i} \Delta^{d_i} y_{i,t-p}\phi_{ip} + \sum_{q=1}^{Q_i} \varepsilon_{i,t-q}\rho_{iq} + \varepsilon_{it}$$

- x_{it} is the *vector* of covariates for unit i, time t. Not just a scalar.

Be clear what the subscripts and variables are

$$\Delta^{d_i} y_{it} = \alpha_i + \mathrm{x}_{it}\boldsymbol{\beta}_i + \sum_{p=1}^{P_i} \Delta^{d_i} y_{i,t-p}\phi_{ip} + \sum_{q=1}^{Q_i} \varepsilon_{i,t-q}\rho_{iq} + \varepsilon_{it}$$

- x_{it} is the *vector* of covariates for unit i, time t. Not just a scalar.
- $\boldsymbol{\beta}_i$ is the *vector* of parameters applied to $\mathbf{x}_{it}$ *just* for a particular unit i, for all periods

Be clear what the subscripts and variables are

$$\Delta^{d_i} y_{it} = \alpha_i + \mathrm{x}_{it}\boldsymbol{\beta}_i + \sum_{p=1}^{P_i} \Delta^{d_i} y_{i,t-p}\phi_{ip} + \sum_{q=1}^{Q_i} \varepsilon_{i,t-q}\rho_{iq} + \varepsilon_{it}$$

- x_{it} is the *vector* of covariates for unit i, time t. Not just a scalar.
- $\boldsymbol{\beta}_i$ is the *vector* of parameters applied to $\mathbf{x}_{it}$ *just* for a particular unit i, for all periods
- P_i is the number of lags of the response used for unit i. Could vary by unit.

Be clear what the subscripts and variables are

$$\Delta^{d_i} y_{it} = \alpha_i + \mathrm{x}_{it}\boldsymbol{\beta}_i + \sum_{p=1}^{P_i} \Delta^{d_i} y_{i,t-p}\phi_{ip} + \sum_{q=1}^{Q_i} \varepsilon_{i,t-q}\rho_{iq} + \varepsilon_{it}$$

- x_{it} is the *vector* of covariates for unit i, time t. Not just a scalar.
- $\boldsymbol{\beta}_i$ is the *vector* of parameters applied to $\mathbf{x}_{it}$ *just* for a particular unit i, for all periods
- P_i is the number of lags of the response used for unit i. Could vary by unit.
- ϕ_{ip} is the AR parameter applied to the pth lag, $\Delta^{d_i} y_{i,t-p}$, for unit i.

Pooling and Partial Pooling

Alternative: we could "borrow strength" across units in estimating parameters

This involves imposing restrictions on (at least some of) the parameters
to assume they are either related or identical across units

Trade-off between flexibility to measure heterogenity,
and pooling data to estimate shared parameters more precisely

Same kind of trade-off is at work in *all* modeling decisions,
and all modeling involves weighing these trade-offs

All models are oversimplifications

Same trade-off is at work in *all* modeling decisions

For example, why can't we estimate, for a standard cross-sectional dataset with a Normally distributed y_i, this inarguably correct linear model:

$$y_i = \alpha_i + \mathrm{x}_i \beta_i + \varepsilon_i$$

All models are oversimplifications

Same trade-off is at work in *all* modeling decisions

For example, why can't we estimate, for a standard cross-sectional dataset with a Normally distributed y_i, this inarguably correct linear model:

$$y_i = \alpha_i + \mathrm{x}_i \beta_i + \varepsilon_i$$

To do any inference,

to learn anything non-obvious from data,

to reduce any data to a simpler model,

we must impose restrictions on parameters which are arguably false

Panel data simply offers a wider range of choices on which parameters to "pool" and which to separate out

The range of models available for panel data

Full flexibility:

$$\Delta^{d_i} y_{it} = \alpha_i + \mathrm{x}_{it}\boldsymbol{\beta}_i + \sum_{p=1}^{P_i} \Delta^{d_i} y_{i,t-p}\phi_{ip} + \sum_{q=1}^{Q_i} \varepsilon_{i,t-q}\rho_{iq} + \varepsilon_{it}$$

$$\varepsilon_{it} \sim \mathrm{N}(0, \sigma_i^2)$$

For each i, we need to choose p_i, d_i, q_i and estimate $\alpha_i, \boldsymbol{\beta_i}, \boldsymbol{\phi_i}, \boldsymbol{\rho_i}, \sigma_i^2$

The range of models available for panel data

Full flexibility:

$$\Delta^{d_i} y_{it} = \alpha_i + \mathrm{x}_{it}\boldsymbol{\beta}_i + \sum_{p=1}^{P_i} \Delta^{d_i} y_{i,t-p}\phi_{ip} + \sum_{q=1}^{Q_i} \varepsilon_{i,t-q}\rho_{iq} + \varepsilon_{it}$$

$$\varepsilon_{it} \sim \mathrm{N}(0, \sigma_i^2)$$

For each i, we need to choose p_i, d_i, q_i and estimate $\alpha_i, \boldsymbol{\beta_i}, \boldsymbol{\phi_i}, \boldsymbol{\rho_i}, \sigma_i^2$

Full pooling:

$$\Delta^{d} y_{it} = \alpha + \mathrm{x}_{it}\boldsymbol{\beta} + \sum_{p=1}^{P} \Delta^{d} y_{i,t-p}\phi_{p} + \sum_{q=1}^{Q} \varepsilon_{i,t-q}\rho_{q} + \varepsilon_{it}$$

$$\varepsilon_{it} \sim \mathrm{N}(0, \sigma^2)$$

We choose common p, d, q across all i, and estimate common $\alpha, \boldsymbol{\beta}, \boldsymbol{\rho}, \boldsymbol{\phi}, \sigma^2$

Popular panel specifications

Variable intercepts

$$\Delta^d y_{it} = \alpha_i + \mathrm{x}_{it}\boldsymbol{\beta} + \sum_{p=1}^{P} \Delta^d y_{i,t-p}\phi_p + \sum_{q=1}^{Q} \varepsilon_{i,t-q}\rho_q + \varepsilon_{it}$$

$$\varepsilon_{it} \sim \mathrm{N}(0, \sigma^2)$$

Popular panel specifications

Variable intercepts

$$\Delta^d y_{it} = \alpha_i + \mathrm{x}_{it}\boldsymbol{\beta} + \sum_{p=1}^{P} \Delta^d y_{i,t-p}\phi_p + \sum_{q=1}^{Q} \varepsilon_{i,t-q}\rho_q + \varepsilon_{it}$$

$$\varepsilon_{it} \sim \mathrm{N}(0, \sigma^2)$$

Variable slopes and intercepts

$$\Delta^d y_{it} = \alpha_i + \mathrm{x}_{it}\boldsymbol{\beta}_i + \sum_{p=1}^{P} \Delta^d y_{i,t-p}\phi_p + \sum_{q=1}^{Q} \varepsilon_{i,t-q}\rho_q + \varepsilon_{it}$$

$$\varepsilon_{it} \sim \mathrm{N}(0, \sigma^2)$$

Popular panel specifications

Variable lag structures

$$\Delta^{d_i} y_{it} = \alpha + \mathrm{x}_{it}\boldsymbol{\beta} + \sum_{p=1}^{P_i} \Delta^{d_i} y_{i,t-p}\phi_{ip} + \sum_{q=1}^{Q_i} \varepsilon_{i,t-q}\rho_{iq} + \varepsilon_{it}$$

$$\varepsilon_{it} \sim \mathrm{N}(0, \sigma^2)$$

Popular panel specifications

Variable lag structures

$$\Delta^{d_i} y_{it} = \alpha + \mathrm{x}_{it}\boldsymbol{\beta} + \sum_{p=1}^{P_i} \Delta^{d_i} y_{i,t-p}\phi_{ip} + \sum_{q=1}^{Q_i} \varepsilon_{i,t-q}\rho_{iq} + \varepsilon_{it}$$

$$\varepsilon_{it} \sim \mathrm{N}(0, \sigma^2)$$

Panel heteroskedasticity

$$\Delta^{d} y_{it} = \alpha + \mathrm{x}_{it}\boldsymbol{\beta} + \sum_{p=1}^{P} \Delta^{d} y_{i,t-p}\phi_{p} + \sum_{q=1}^{Q} \varepsilon_{i,t-q}\rho_{q} + \varepsilon_{it}$$

$$\varepsilon_{it} \sim \mathrm{N}(0, {\sigma_i}^2)$$

Models of variable intercepts

$$\Delta^d y_{it} = \alpha_i + \mathrm{x}_{it}\boldsymbol{\beta} + \sum_{p=1}^{P} \Delta^d y_{i,t-p}\phi_p + \sum_{q=1}^{Q} \varepsilon_{i,t-q}\rho_q + \varepsilon_{it}$$

$$\varepsilon_{it} \sim \mathrm{N}(0, \sigma^2)$$

How do we model α_i?

Let the mean of α_i be α_i^*.

Models of variable intercepts

$$\Delta^d y_{it} = \alpha_i + \mathrm{x}_{it}\boldsymbol{\beta} + \sum_{p=1}^{P} \Delta^d y_{i,t-p}\phi_p + \sum_{q=1}^{Q} \varepsilon_{i,t-q}\rho_q + \varepsilon_{it}$$

Then there are a range of possibilities:

Let α_i be a random variable with no systemic component
(this type of α_i known as a *random effect*)

$$\alpha_i \sim \mathrm{N}\left(0, \sigma_\alpha^2\right)$$

Models of variable intercepts

$$\Delta^d y_{it} = \alpha_i + \mathrm{x}_{it}\boldsymbol{\beta} + \sum_{p=1}^{P} \Delta^d y_{i,t-p}\phi_p + \sum_{q=1}^{Q} \varepsilon_{i,t-q}\rho_q + \varepsilon_{it}$$

Then there are a range of possibilities:

Let α_i be a random variable with no systemic component
(this type of α_i known as a *random effect*)

$$\alpha_i \sim \mathrm{N}\left(0, \sigma_\alpha^2\right)$$

Let α_i be a systematic component with no stochastic component
(this type of α_i is known as a *fixed effect*)

$$\alpha_i = \alpha_i^*$$

Models of variable intercepts

$$\Delta^d y_{it} = \alpha_i + \mathrm{x}_{it}\boldsymbol{\beta} + \sum_{p=1}^{P} \Delta^d y_{i,t-p}\phi_p + \sum_{q=1}^{Q} \varepsilon_{i,t-q}\rho_q + \varepsilon_{it}$$

Then there are a range of possibilities:

Let α_i be a random variable with no systemic component
(this type of α_i known as a *random effect*)

$$\alpha_i \sim \mathrm{N}\left(0, \sigma_\alpha^2\right)$$

Let α_i be a systematic component with no stochastic component
(this type of α_i is known as a *fixed effect*)

$$\alpha_i = \alpha_i^*$$

Let α_i be a random variable with a unit-specific systematic component
(this type of α_i known as a *mixed effect)*

$$\alpha_i \sim \mathrm{N}\left(\alpha_i^*, \sigma_\alpha^2\right)$$

Random effects

$$\alpha_i \sim \mathrm{N}\left(0, \sigma_\alpha^2\right)$$

Intuitive from a maximum likelihood modeling perspective

A unit specific error term

Assumes the units come from a common population,
with an unknown (estimated) variance, σ_α^2

In likelihood inference, estimation focuses on this variance, not on particular α_i's

Uncorrelated with $\mathrm{x_{it}}$ by design

Need MLE to estimate

Random effects example

A (contrived) example may help clarify what random effects are.

Suppose that we have data following this true model:

$$
\begin{aligned}
y_{it} &= \beta_0 + \beta_1 x_{it} + \alpha_i + \varepsilon_{it} \\
\alpha_i &\sim \mathcal{N}(0, \sigma_\alpha^2) \\
\varepsilon_{it} &\sim \mathcal{N}(0, \sigma^2)
\end{aligned}
$$

with $i \in \{1, \ldots N\}$ and $t \in \{1, \ldots T\}$

Note that we are ignoring time series dynamics for now

It may help to pretend that these data have a real world meaning
though remember throughout we have created them out of thin air and `rnorm()`

So let's pretend these data reflect undergraduate student assignment scores over a term for $N = 100$ students and $T = 5$ assignments

Random effects example: Student aptitude & effort

Let's pretend these data reflect undergraduate student assignment scores

over a term for $N = 100$ students and $T = 5$ assignments:

$$\begin{aligned} \text{score}_{it} &= \beta_0 + \beta_1 \text{hours}_{it} + \alpha_i + \varepsilon_{it} \\ \alpha_i &\sim \mathcal{N}(0, \sigma_\alpha^2) \\ \varepsilon_{it} &\sim \mathcal{N}(0, \sigma^2) \end{aligned}$$

with $i \in \{1, \ldots N\}$ and $t \in \{1, \ldots T\}$

The reponse is the assignment score, score_{it}

and the covariate is the hours studied, hours_{it}

and each student has an unobservable aptitude α_i which is Normally distributed

Aptitude has the same (random) effect on each assignment by a given student

Random effects example: Student aptitude & effort

Let's pretend these data reflect undergraduate student assignment scores

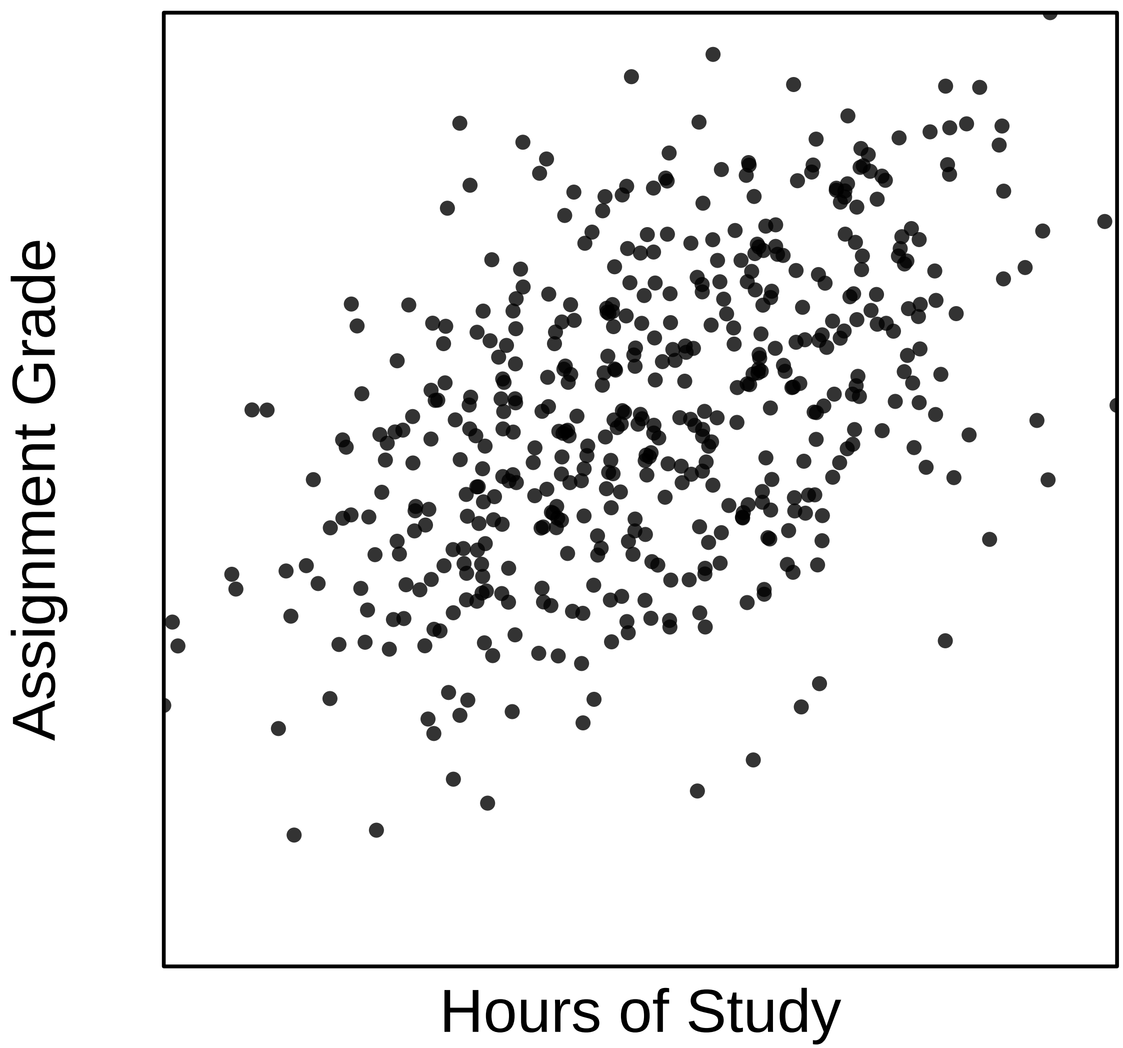
over a term for $N = 100$ students and $T = 5$ assignments:

$$\text{score}_{it} = 0 + 0.75 \times \text{hours}_{it} + \alpha_i + \varepsilon_{it}$$
$$\alpha_i \sim \mathcal{N}(0, 0.7^2)$$
$$\varepsilon_{it} \sim \mathcal{N}(0, 0.2^2)$$

with $i \in \{1, \ldots 100\}$ and $t \in \{1, \ldots 5\}$

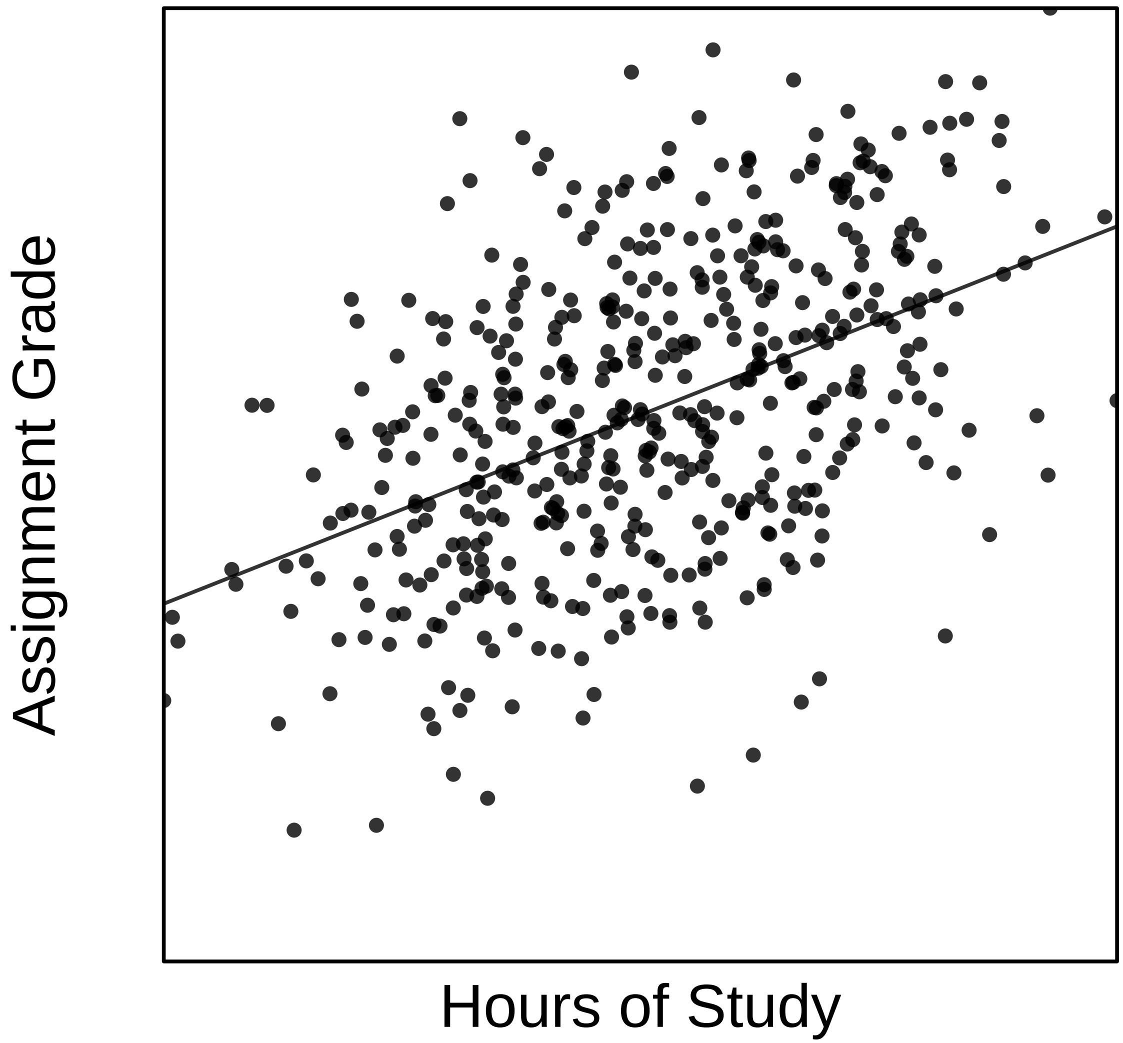
the above are the true values of the parameters I used to generate the data

let's see what role the random effect α_i plays here



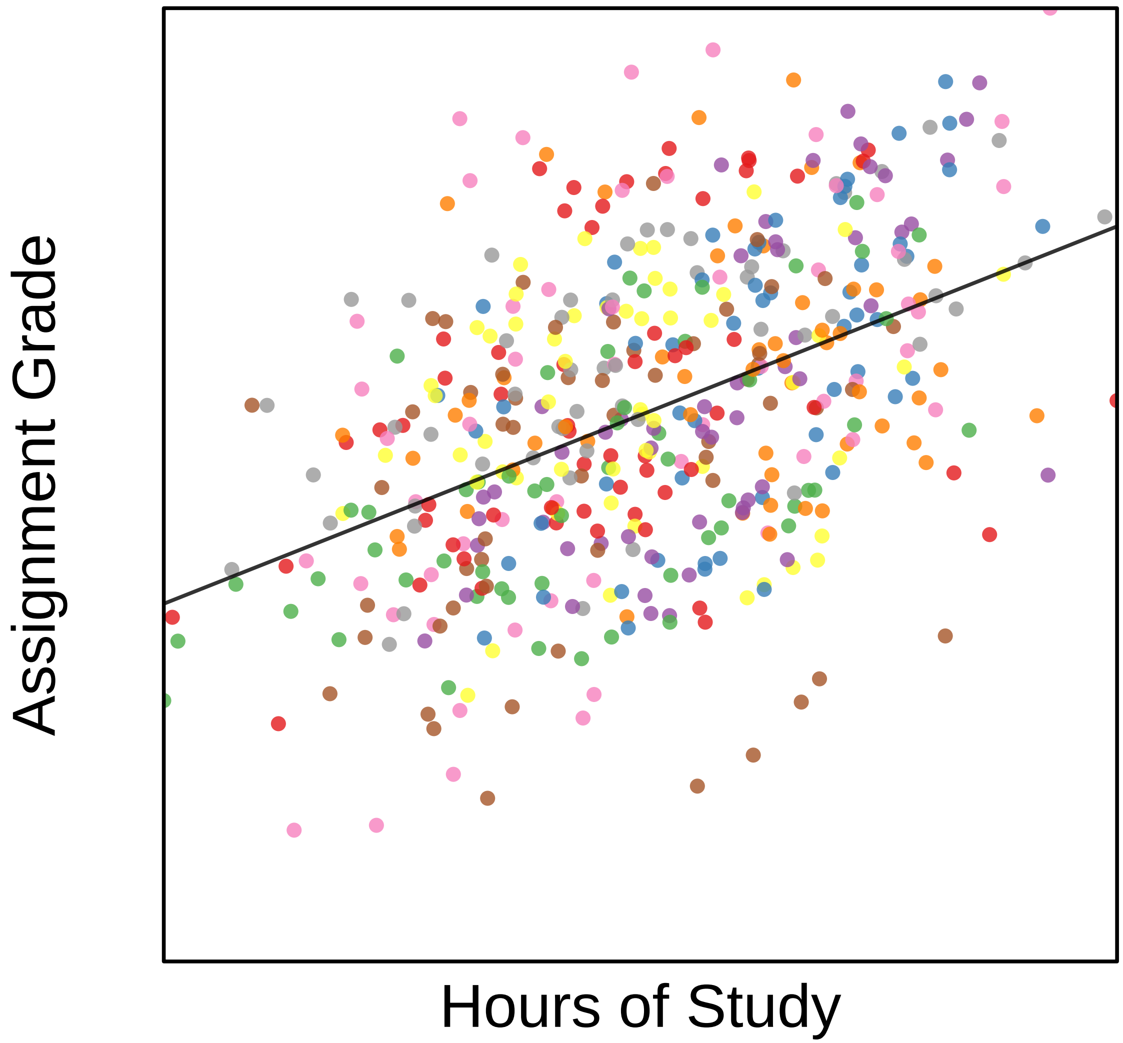
Here are the 500 obervations.

A relationship between effort and grades seems evident.



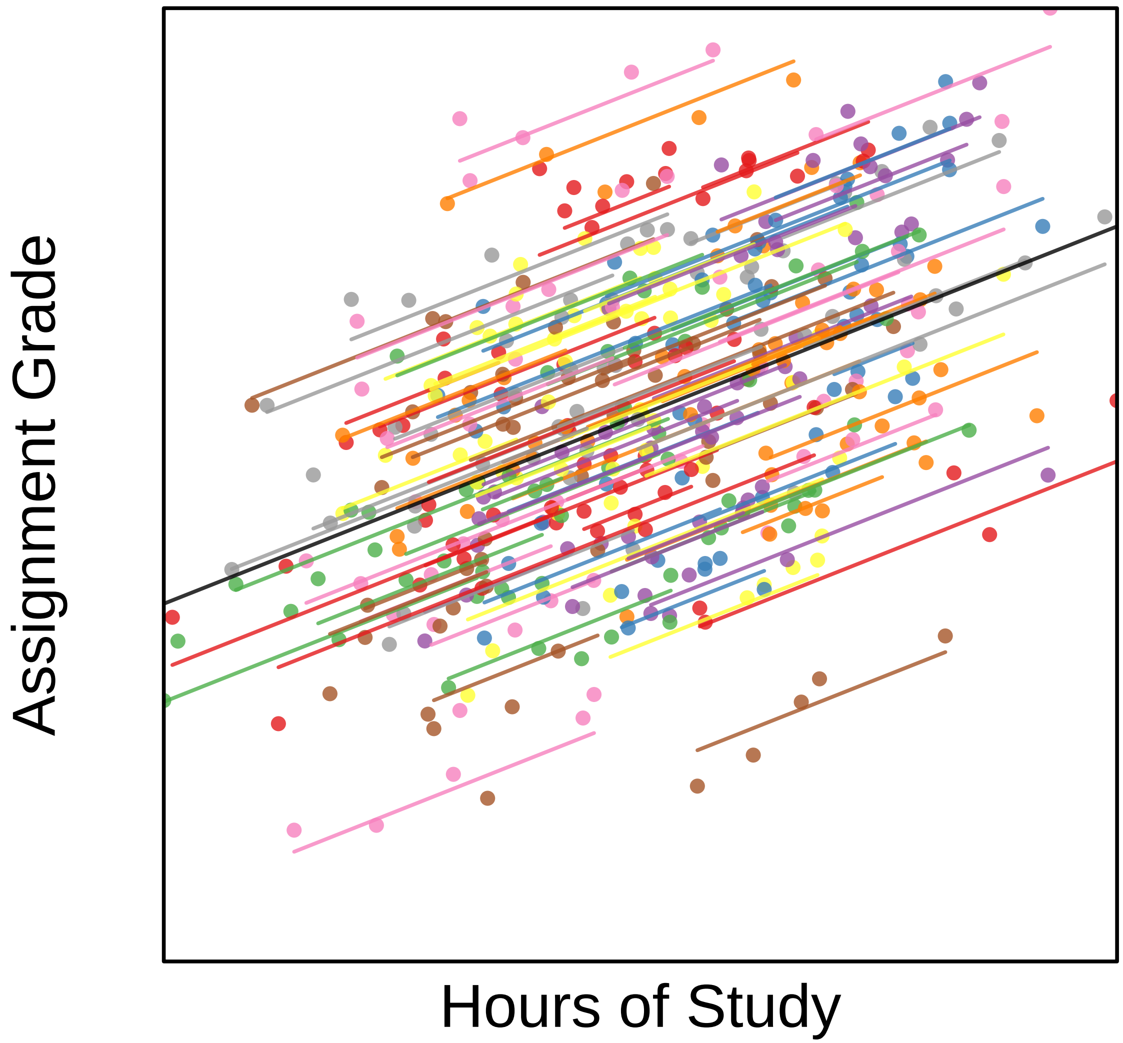
We can summarize that relationship using the least squares estimate of $\hat{\beta}_1$, which is approximately equal to the true $\beta_1 = 0.75$

We haven't discussed, used, or estimated the random effects yet. Do we "need" them?



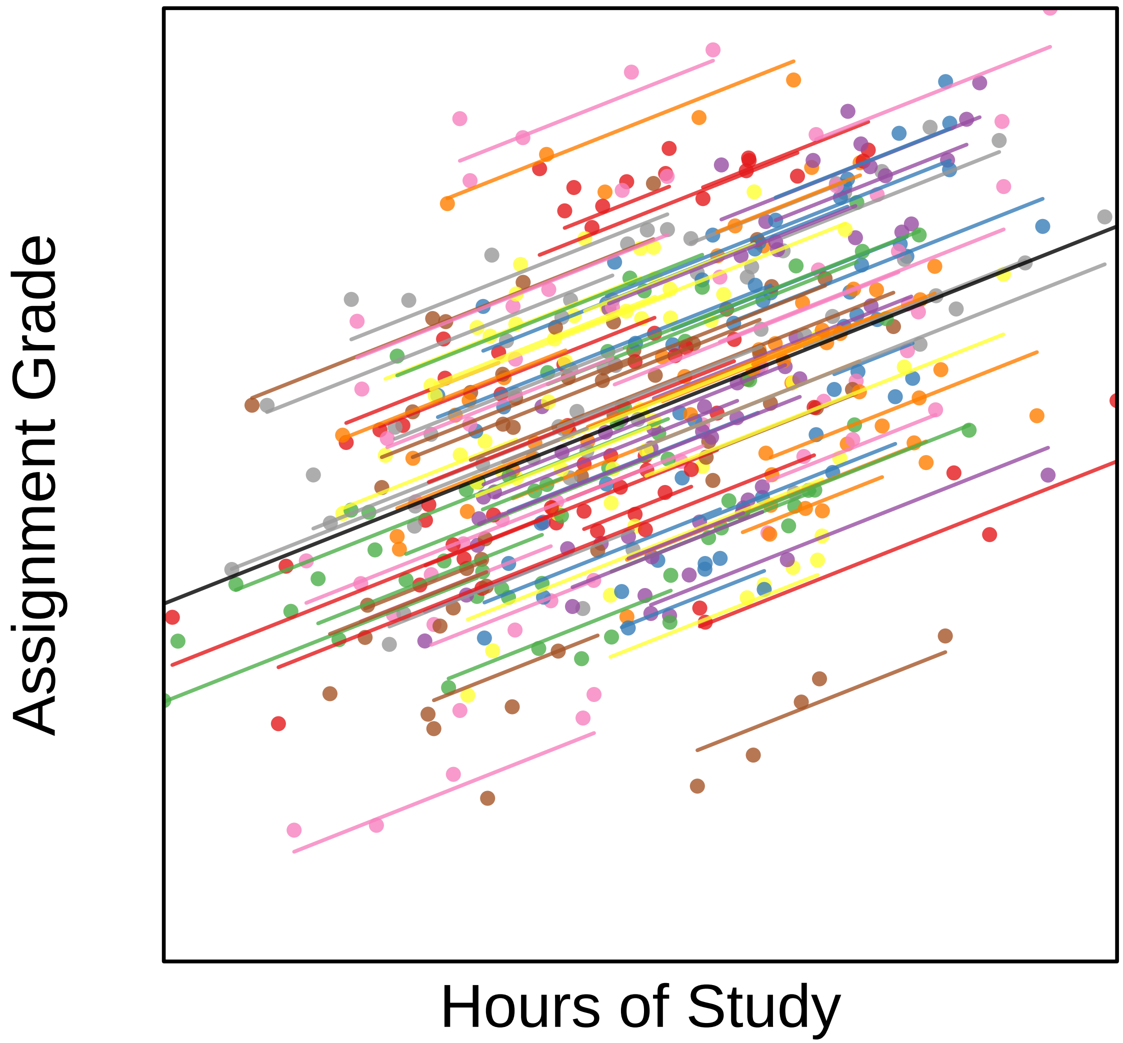
I've identified each of the 100 student using colored dots

Colors repeat, but each student's scores are tightly clustered.
Note the student-level pattern



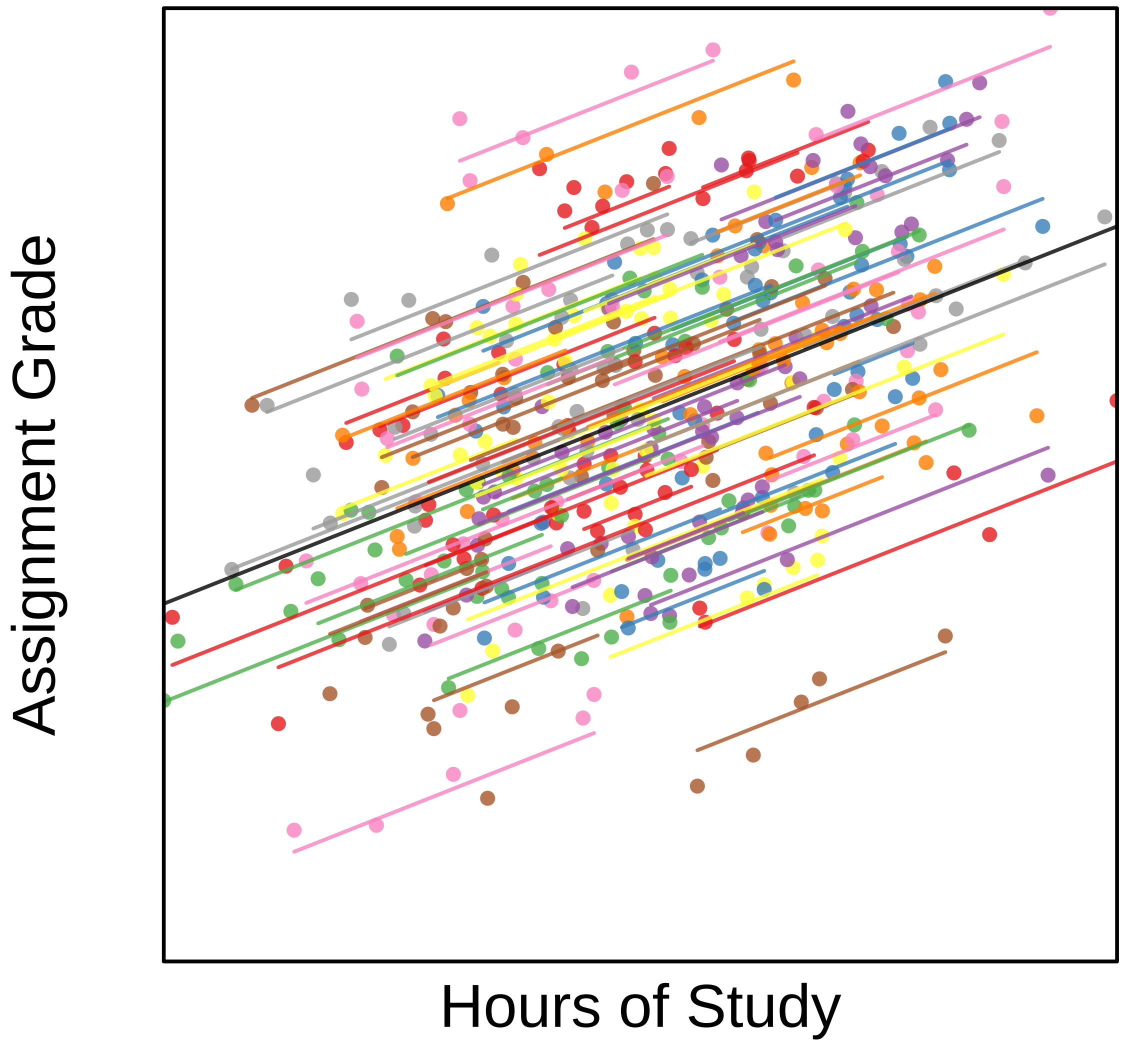
It is clear that each student is following the same regression line as the whole class, but with a unique intercept

That intercept is the random effect. It is the average difference between that student's scores and the class-level regression line



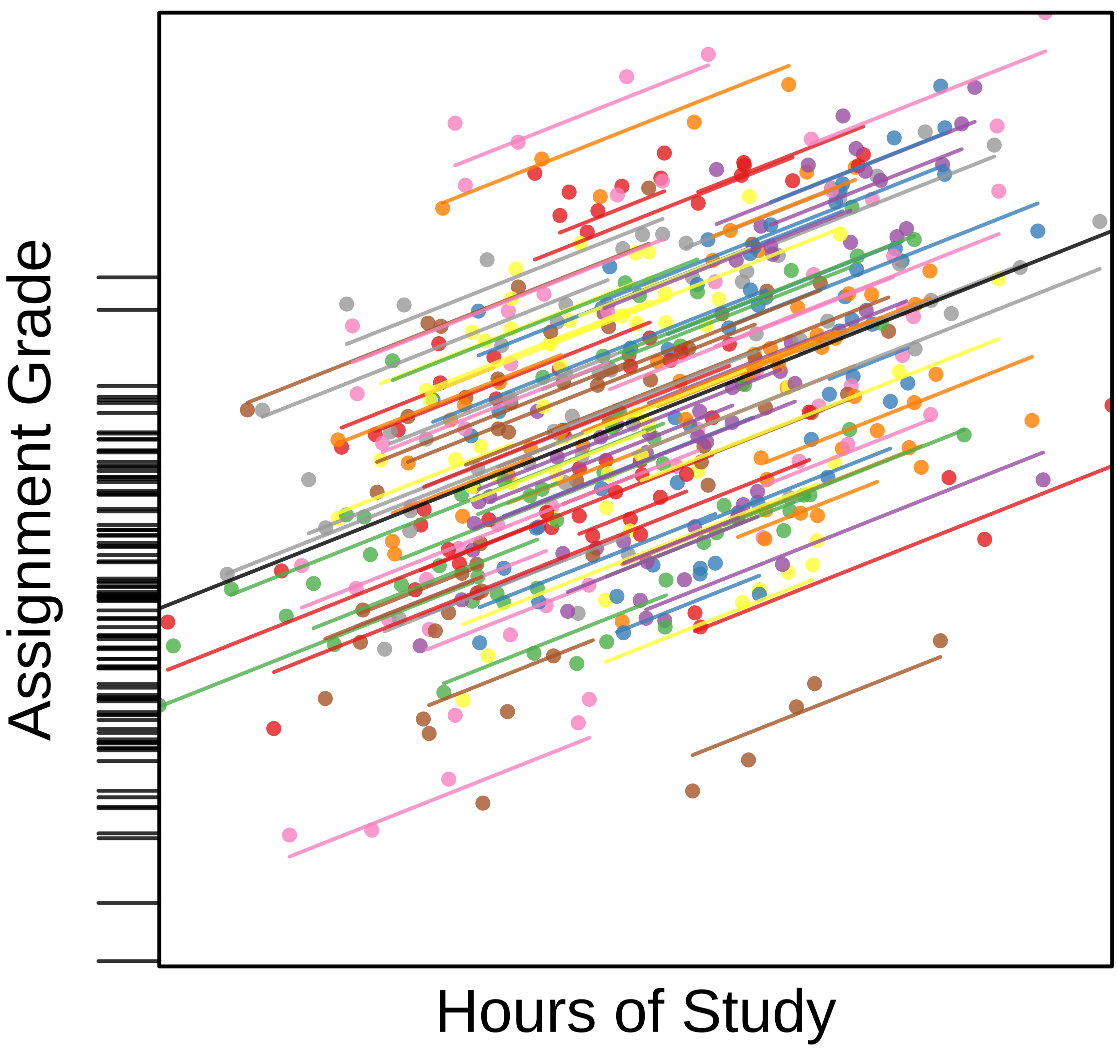
The student random effect is the student-specific component of the error term

After we remove it, the student scores exhibit white noise variation around a student-specific version of the overall regression line



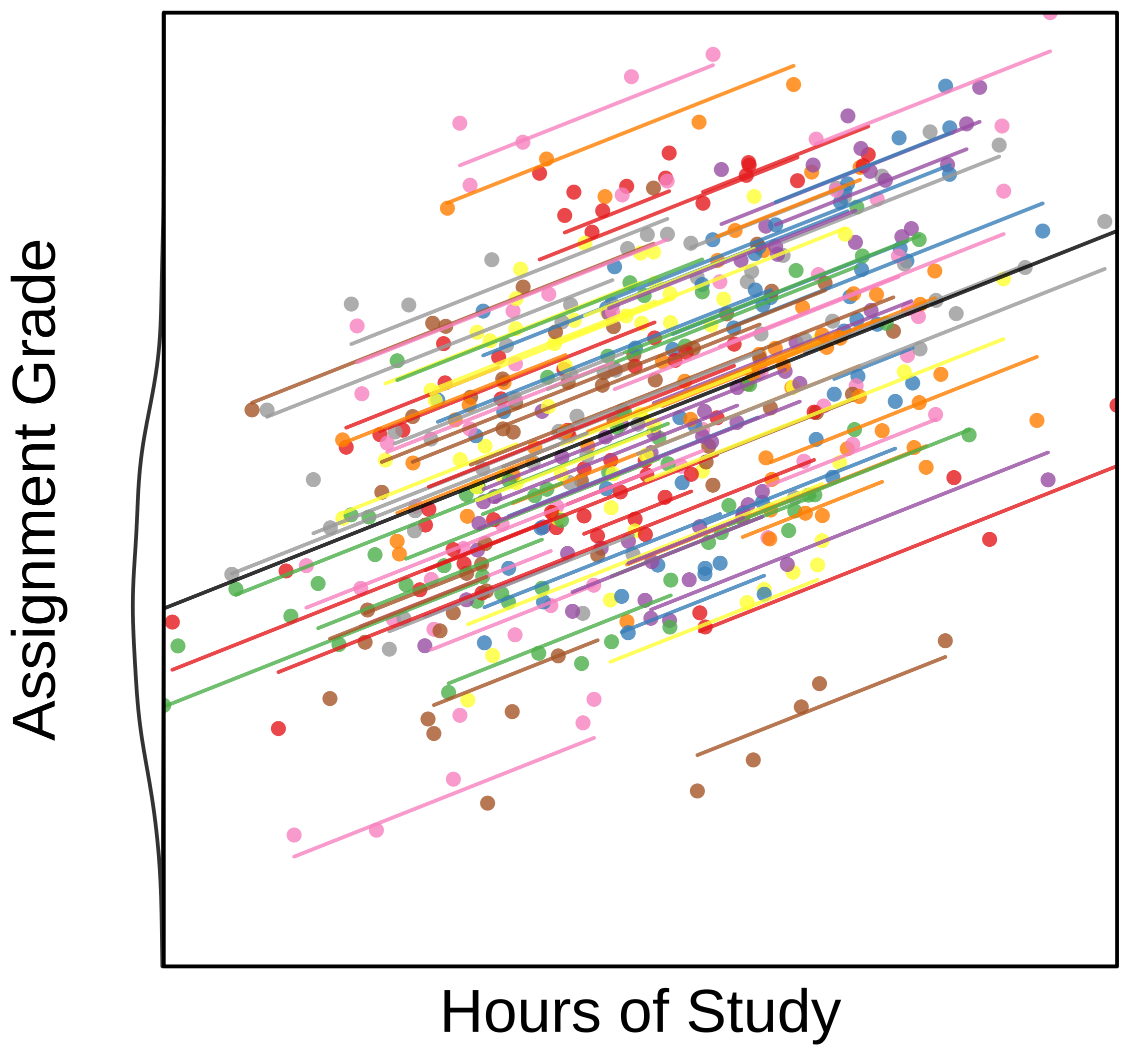
Conceptually, we can think of the random effects as displaying that portion of the error term which reflects unmeasured student characteristics

I've labelled this "aptitude", which is just a word for everything fixed about a student's ability

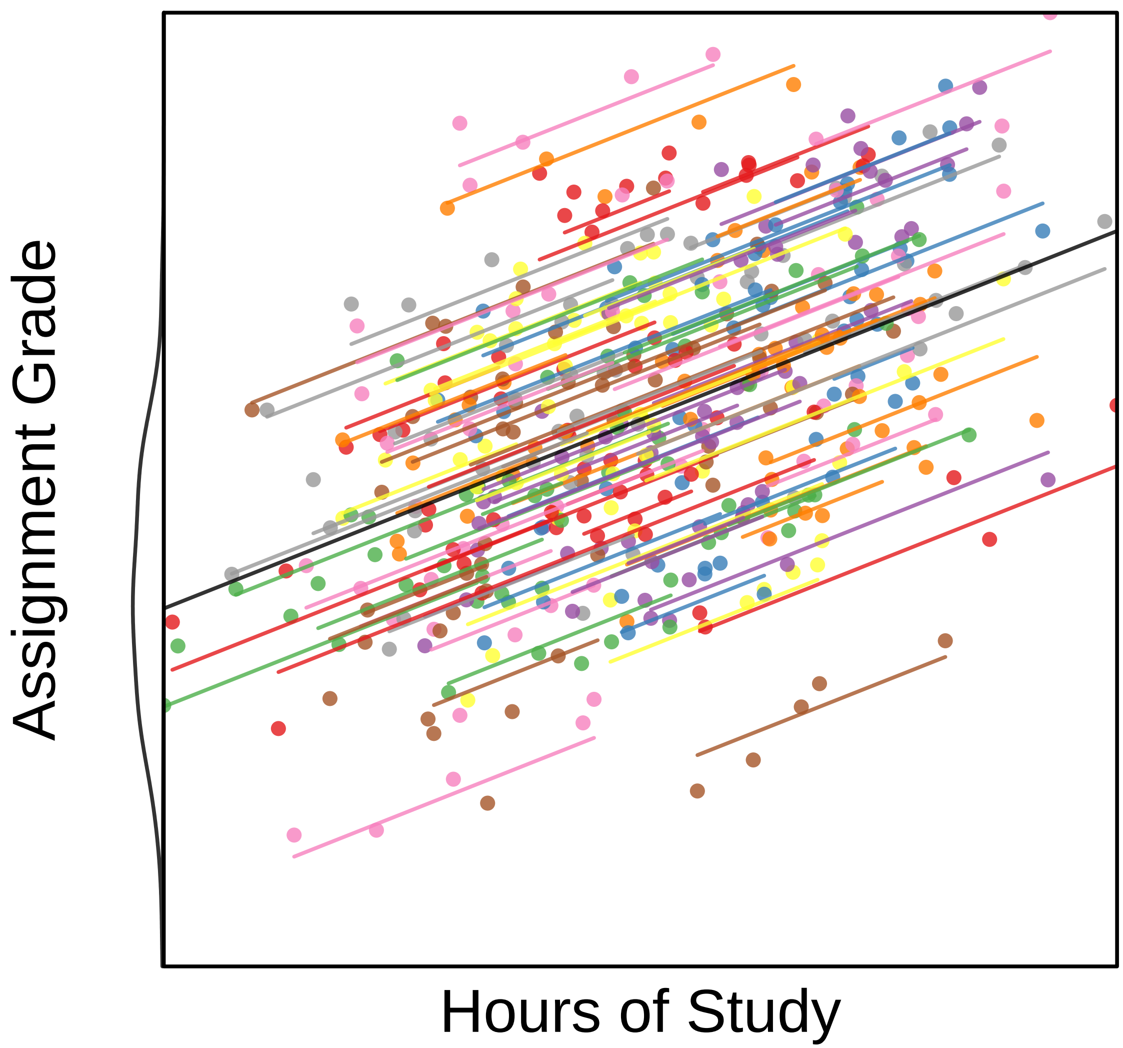


The distribution of the random effects is shown at the left

A plot of a marginal distribution on the side of a scatterplot is called a “rug”

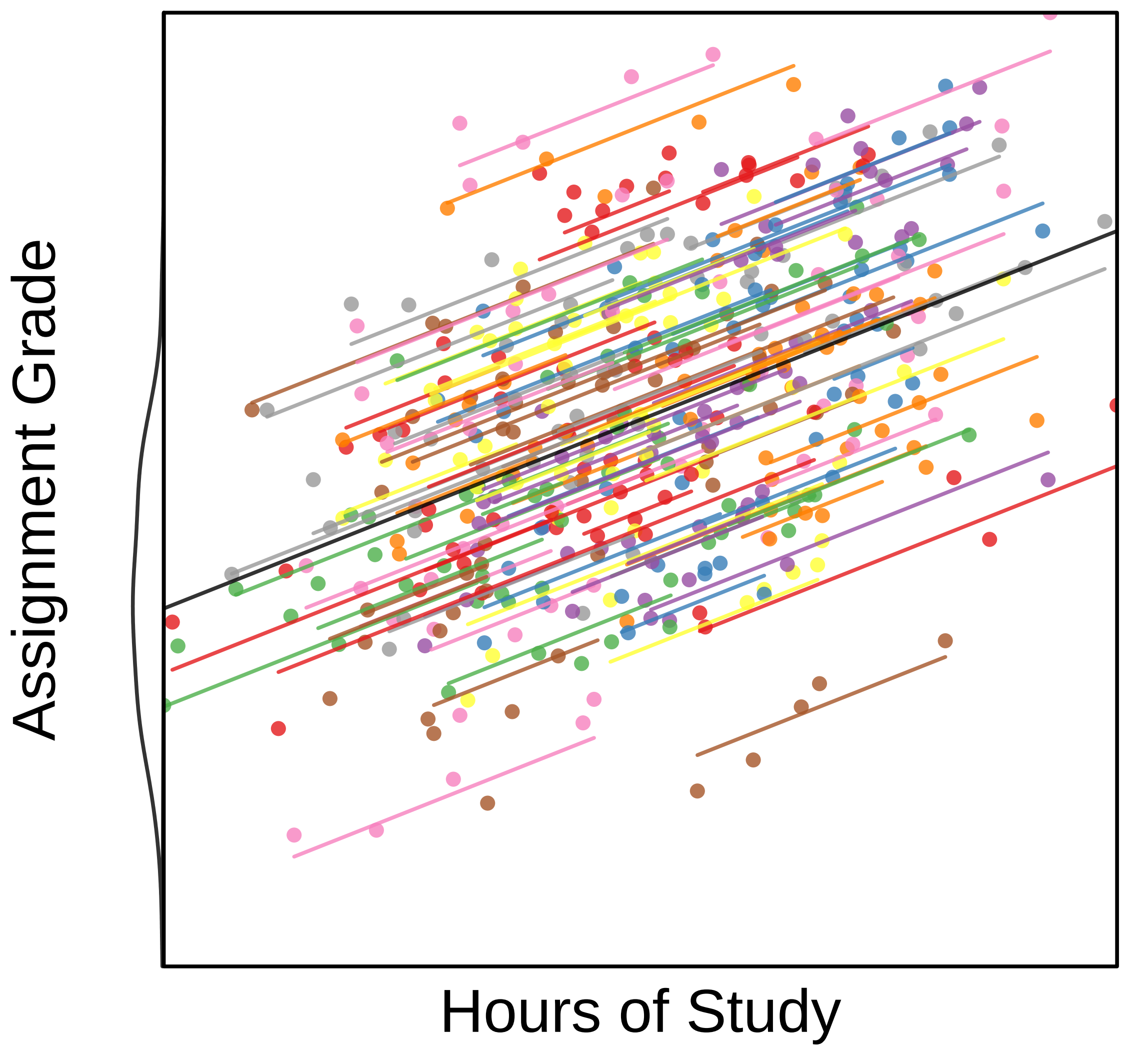


A density version of the distribution of random effects confirms they are approximately Normal



Random effects are a decomposition of the error term into a unit-specific part and an idiosyncratic part

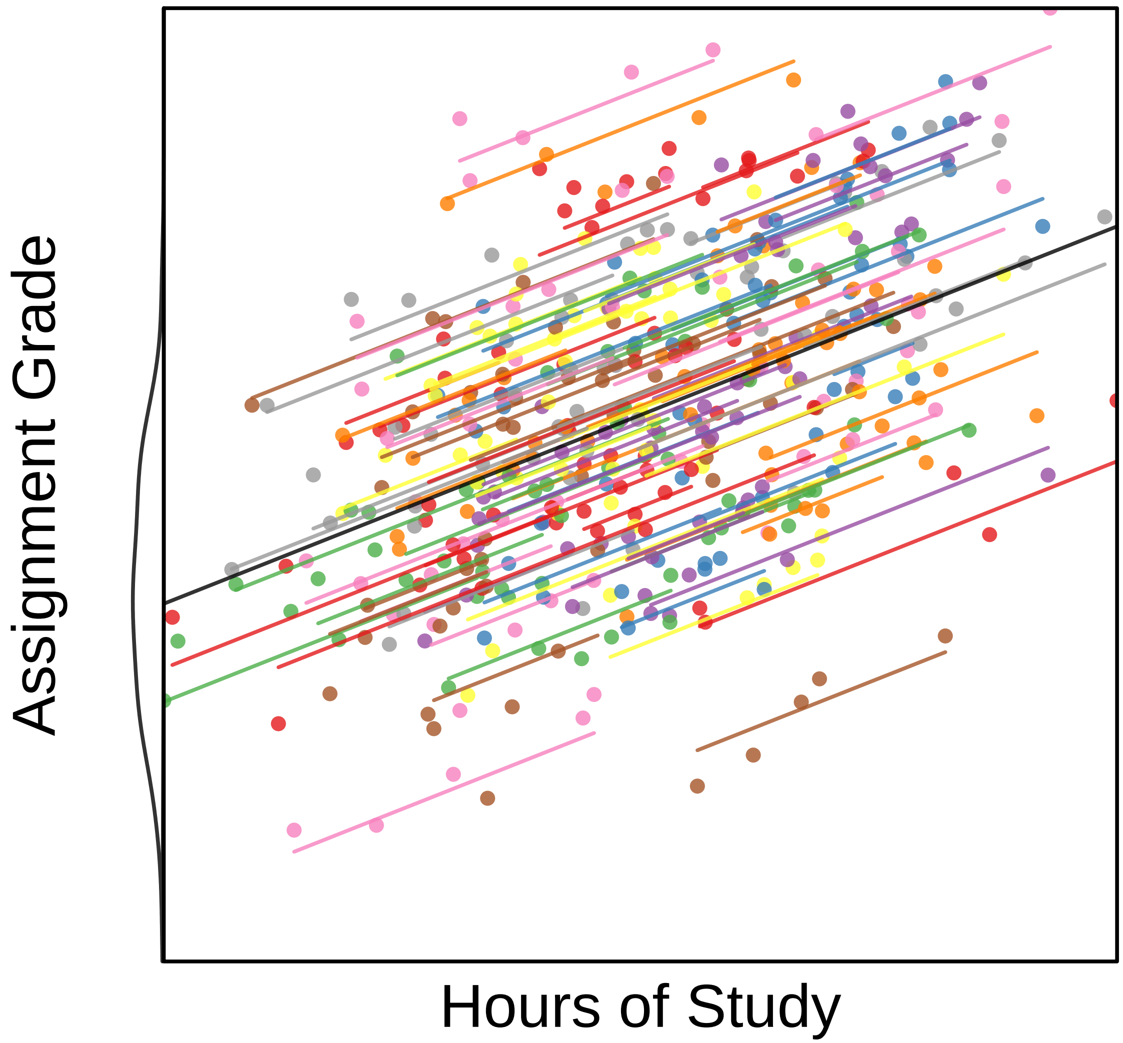
The random effects are determined after we have the overall regression slope, and cannot change that slope



This model is now hierarchical or multi-level

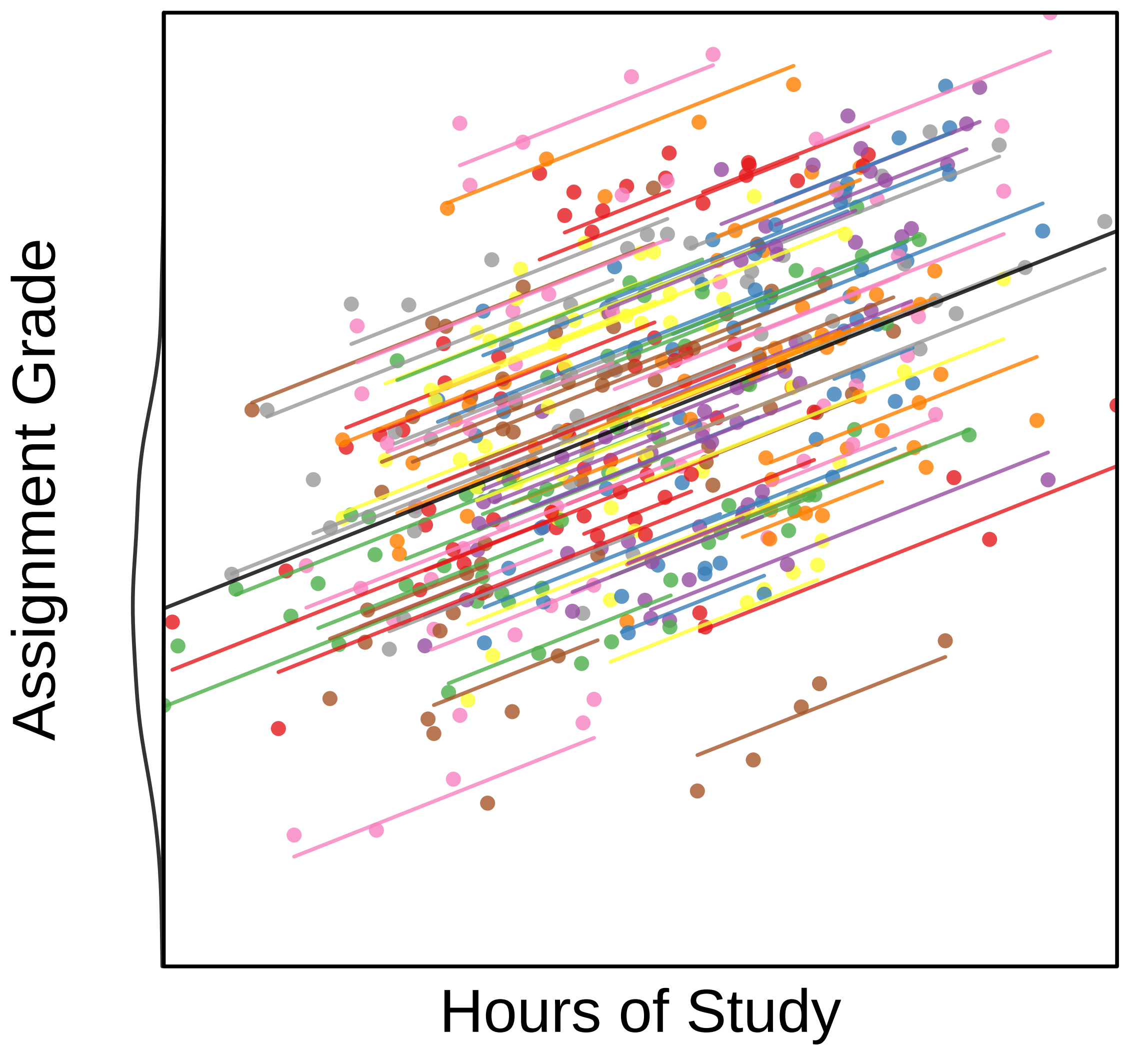
Level 1: student level sits above Level 2: Student $\times$ assignment level

There is random variation at both levels, but mainly at the student level



Students randomly vary a lot: $\sigma^{\alpha} = 0.7$,
but assignments for a given student vary little: $\sigma = 0.2$

Student level random effects comprise
$100\% \times \sqrt{0.7^2/(0.7^2 + 0.2^2)} = 96\%$ of the total error variance



We haven't controlled for any omitted confounders

If unmeasured ability is correlated with study effort, then our $\hat{\beta}_1$ estimate will be biased even if we include random effects

Fixed effects

$$\alpha_i = \alpha_i^*$$

Easiest to conceptualize in a linear regression framework

Easiest to estimate: just add dummies for each unit, and drop the intercept

Can be correlated with $\mathbf{x}_{it}$: FEs control for *all* omitted time-invariant variables

Indeed, that's usually the point.
FEs usually included to capture unobserved variance potentially correlated with $\mathbf{x}_{it}$.

Comes at a large cost:
we're actually pruging the cross-sectional variation from the analysis

Then assuming a change in $\mathbf{x}$ would yielf the same response in each time series

Fixed effects models use over-time variation in covariates to estimate parameters

More on fixed effects

$$\alpha_i = \alpha_i^*$$

Cannot be added to models with perfectly time invariant covariates

Fixed effects specifications incur an incidental parameters problem:
MLE is consistent as $T \rightarrow \infty$, but *not* as $N \rightarrow \infty$.

More on fixed effects

$$\alpha_i = \alpha_i^*$$

Cannot be added to models with perfectly time invariant covariates

Fixed effects specifications incur an incidental parameters problem:
MLE is consistent as $T \to \infty$, but *not* as $N \to \infty$.

Of concern in microeconomics, where panels are sampled on N with T fixed.
Not of concern in CPE/IPE, where N is fixed, and T could expand

Monte Carlo experiments indicate small sample properties of fixed effects pretty good if $t > 15$ or so.

Fixed effects are common in studies where N is not a random sample, but a (small) universe (e.g., the industrialized countries).

More on fixed effects

$$\alpha_i = \alpha_i^*$$

Cannot be added to models with perfectly time invariant covariates

Fixed effects specifications incur an incidental parameters problem:
MLE is consistent as $T \rightarrow \infty$, but *not* as $N \rightarrow \infty$.

Of concern in microeconomics, where panels are sampled on N with T fixed.
Not of concern in CPE/IPE, where N is fixed, and T could expand

Monte Carlo experiments indicate small sample properties of fixed effects pretty good if $t > 15$ or so.

Fixed effects are common in studies where N is not a random sample, but a (small) universe (e.g., the industrialized countries).

Sui generis: Fixed effects basically say "France is different because it's France", "America is different because it's America", etc.

Fixed effects example

Another example may help clarify what fixed effects are.

Suppose that we have data following this true model:

$$y_{ij} = \beta_0 + \beta_1 x_{ij} + \beta_2 z_i + \varepsilon_{ij}$$
$$\varepsilon_{ij} \sim \mathcal{N}(0, \sigma^2)$$

with $i \in \{1, \ldots N\}$ and $j \in \{1, \ldots M_i\}$

j indexes a set of M_i counties drawn from state i

There are $N = 15$ states total, and we drew $M_j = M = 15$ counties from each state

Note that we are ignoring time series dynamics completely now

(We could add them back in if j were ordered in time)

Fixed effects example

Suppose the data represent county level voting patterns for the US

(I.e., let's illustrate Gelman *et al*, *Red State, Blue State, Rich State, Poor State* w/ contrived data)

$$
\begin{aligned}
\mathrm{RVS}_{ij} &= \beta_0 + \beta_1 \mathrm{Income}_{ij} + \beta_2 \mathrm{ConservativeCulture}_i + \varepsilon_{ij} \\
\varepsilon_{ij} &\sim \mathcal{N}(0, \sigma^2)
\end{aligned}
$$

with $i \in \{1, \ldots N\}$ and $j \in \{1, \ldots M_i\}$

j indexes a set of M_i counties drawn from state i

Remember: the data I'm using are fake, and contrived to illustrate a concept simply

Gelman et al investigate this in detail with real data and get similar but more nuanced findings

Fixed effects example: What's the matter with Kansas?

Suppose the data represent county level voting patterns for the US

(I.e., let's illustrate Gelman *et al*, *Red State, Blue State, Rich State, Poor State* using similar but contrived data)

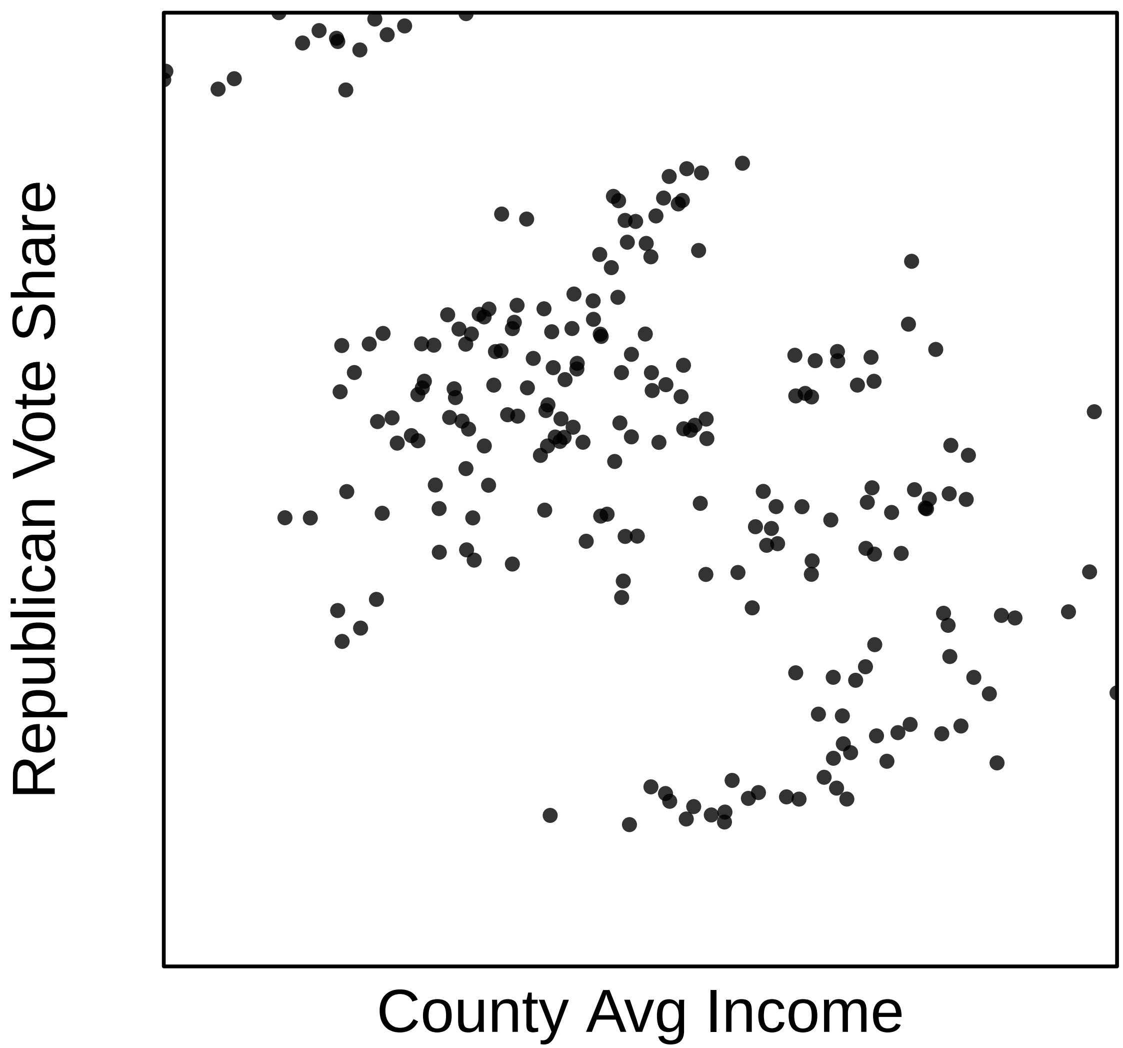
$$\begin{aligned} \mathrm{RVS}_{ij} &= \beta_0 + \beta_1 \mathrm{Income}_{ij} + \beta_2 \mathrm{Conservatism}_i + \varepsilon_{ij} \\ \varepsilon_{ij} &\sim \mathcal{N}(0, \sigma^2) \end{aligned}$$

with $i \in \{1, \ldots N\}$ and $j \in \{1, \ldots M_i\}$

A problem:
suppose we don't have (or don't trust) a measure of state-level Conservatism

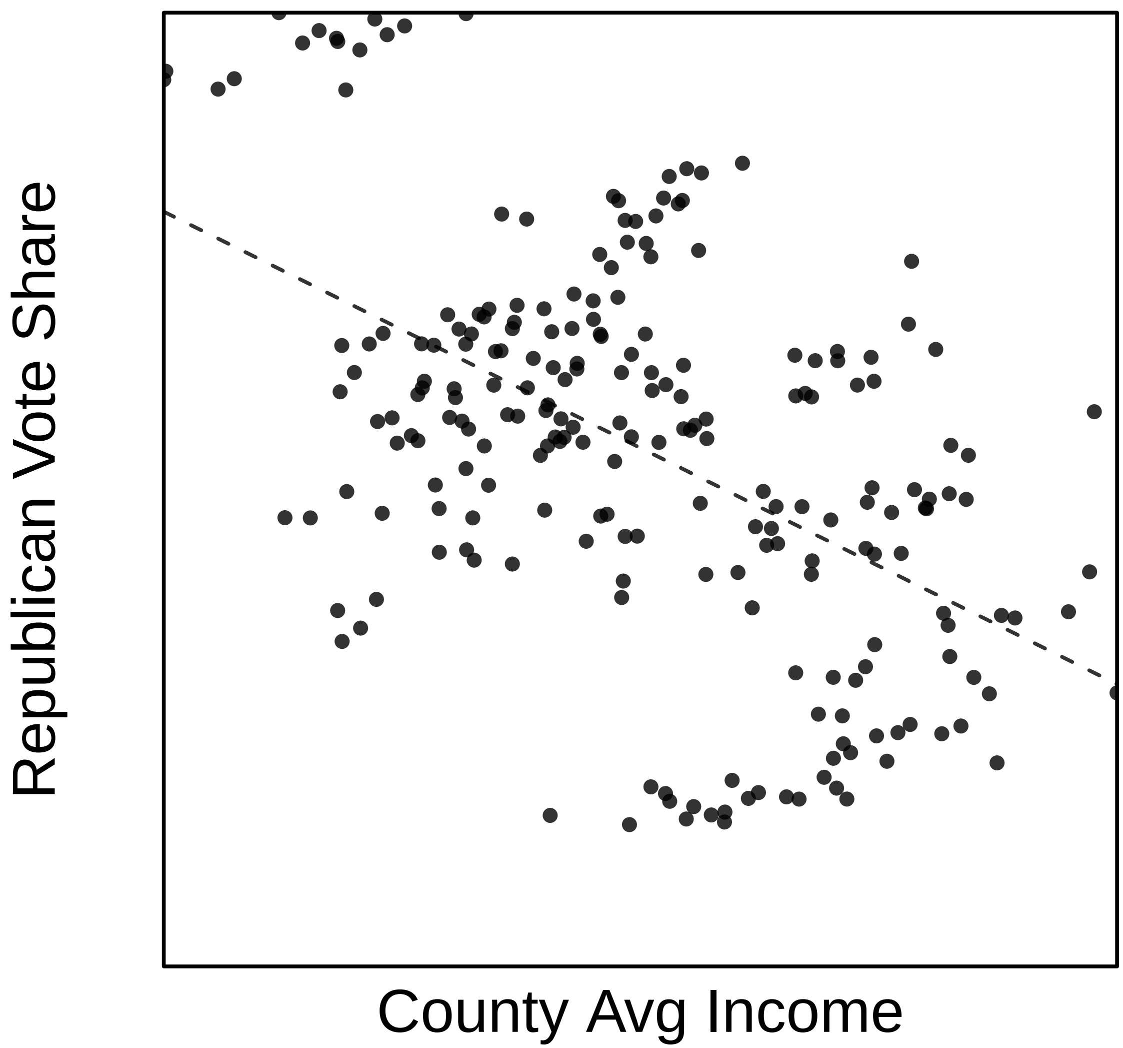
If we exclude it, or mismeasure it, we could get omitted variable bias in $\hat{\beta}_1$

This leads to potentially large misconceptions. . .



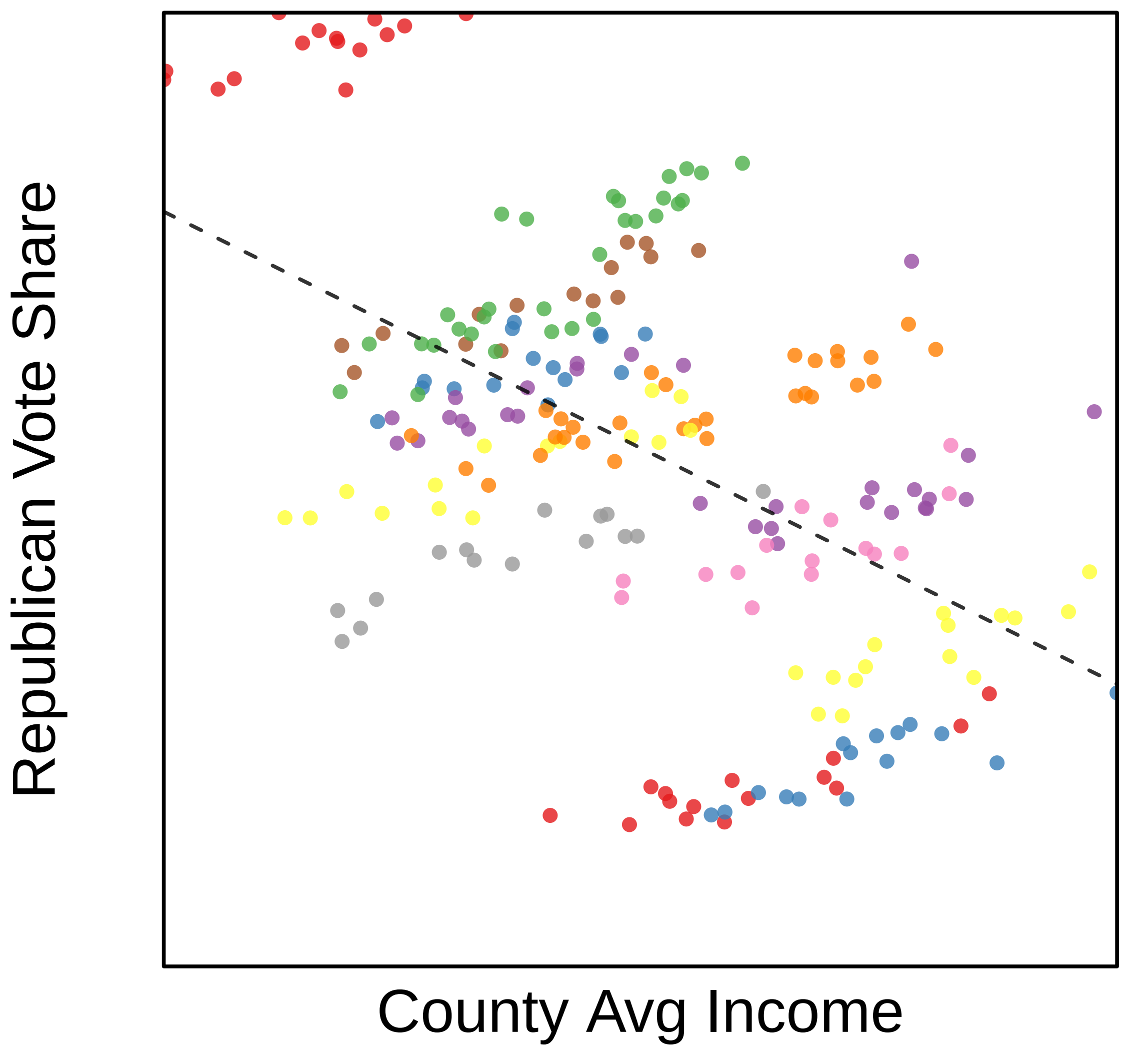
Suppose we observed the above data, drawn from 15 counties from each of 15 states (for a total of 225 observations)

Our first cut is to estimate this simple linear regression: $y_{ij} = \beta_0 + \beta_1 \text{Income}_{ij} + \varepsilon_{ij}$



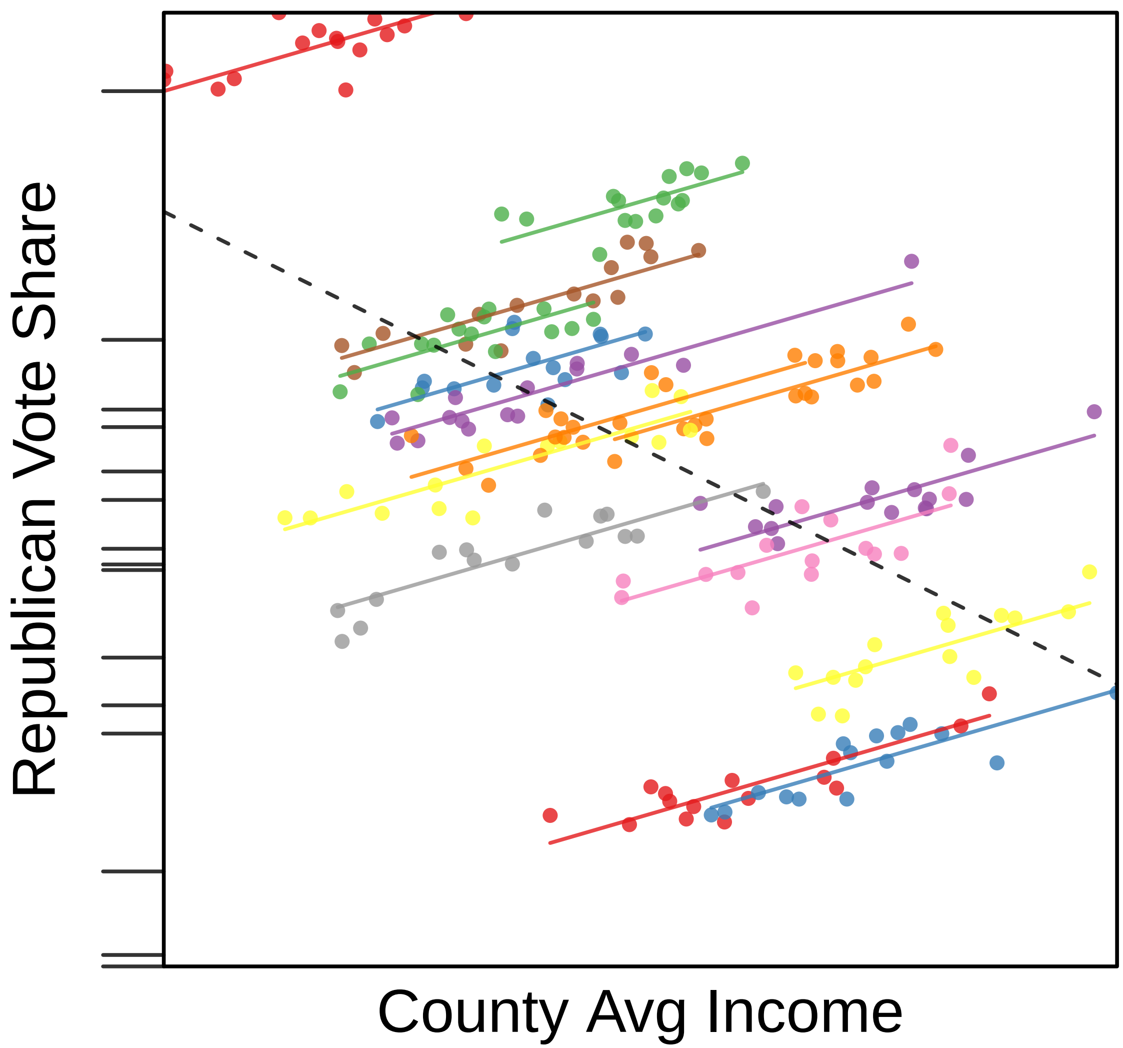
We find that $\hat{\beta}_1$ is negative:
poor counties seem to vote more Republican than rich counties!

But Republican elected officials attempt to represent the affluent.
What's the matter with (poor counties in) Kansas, as Thomas Frank asked?



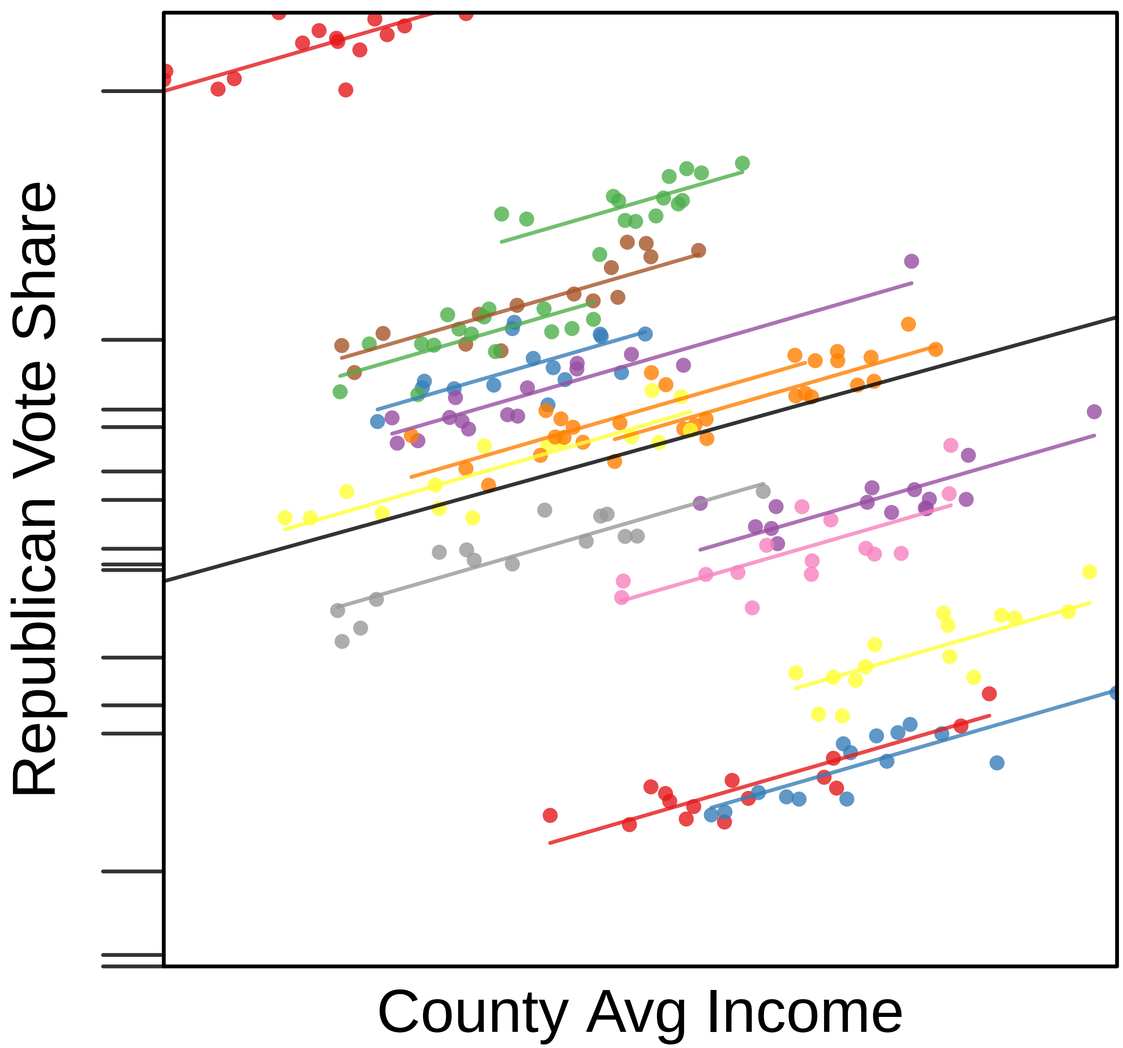
Let's look at which observations come from which states

Clearly, counties from the same state are clustered



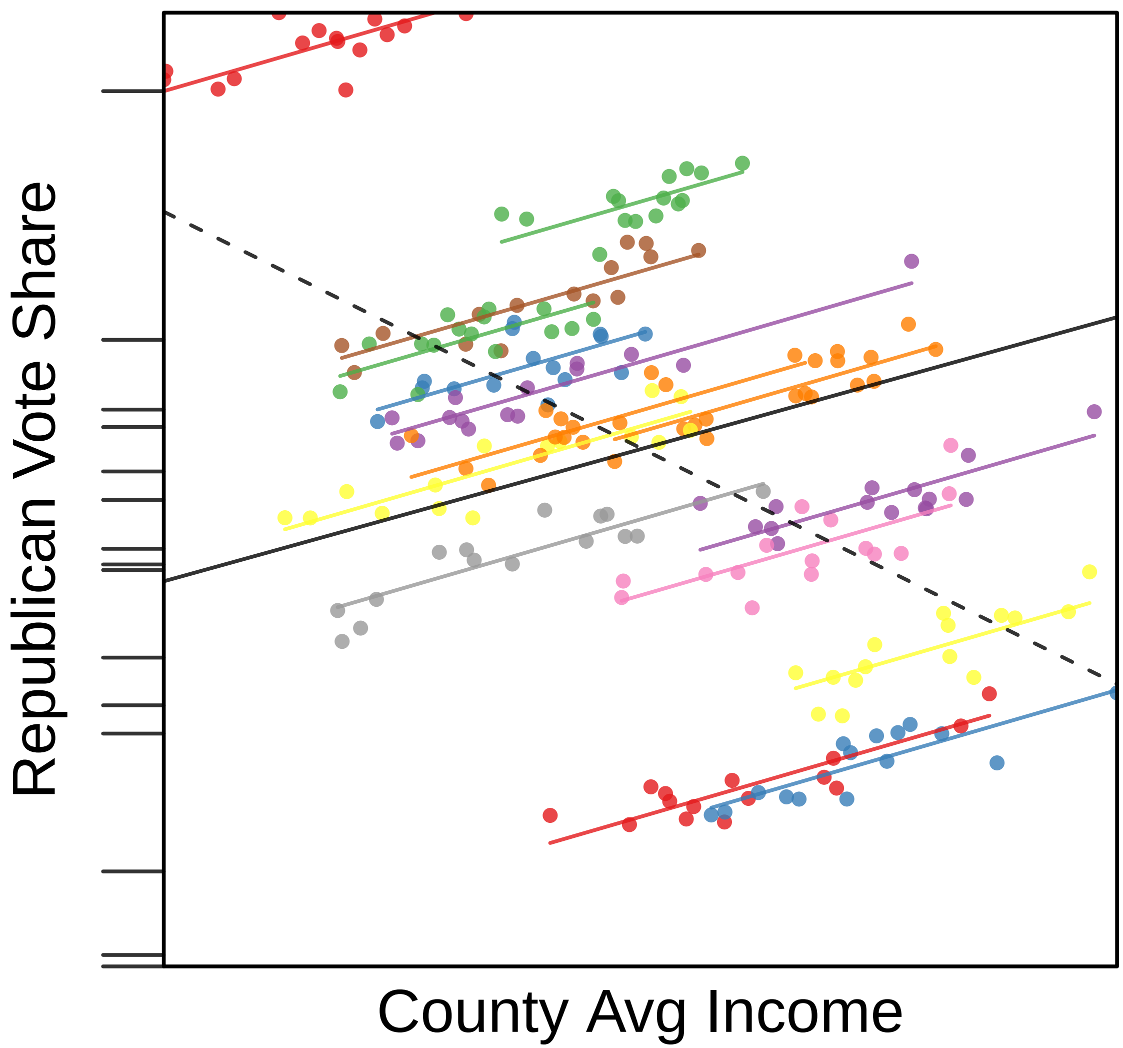
Within each state, there appears to be a *positive* relationship between income and Republican voting

This suggests that we need to control for variation at the state level, either by collecting the state level variables causing the variation. . .



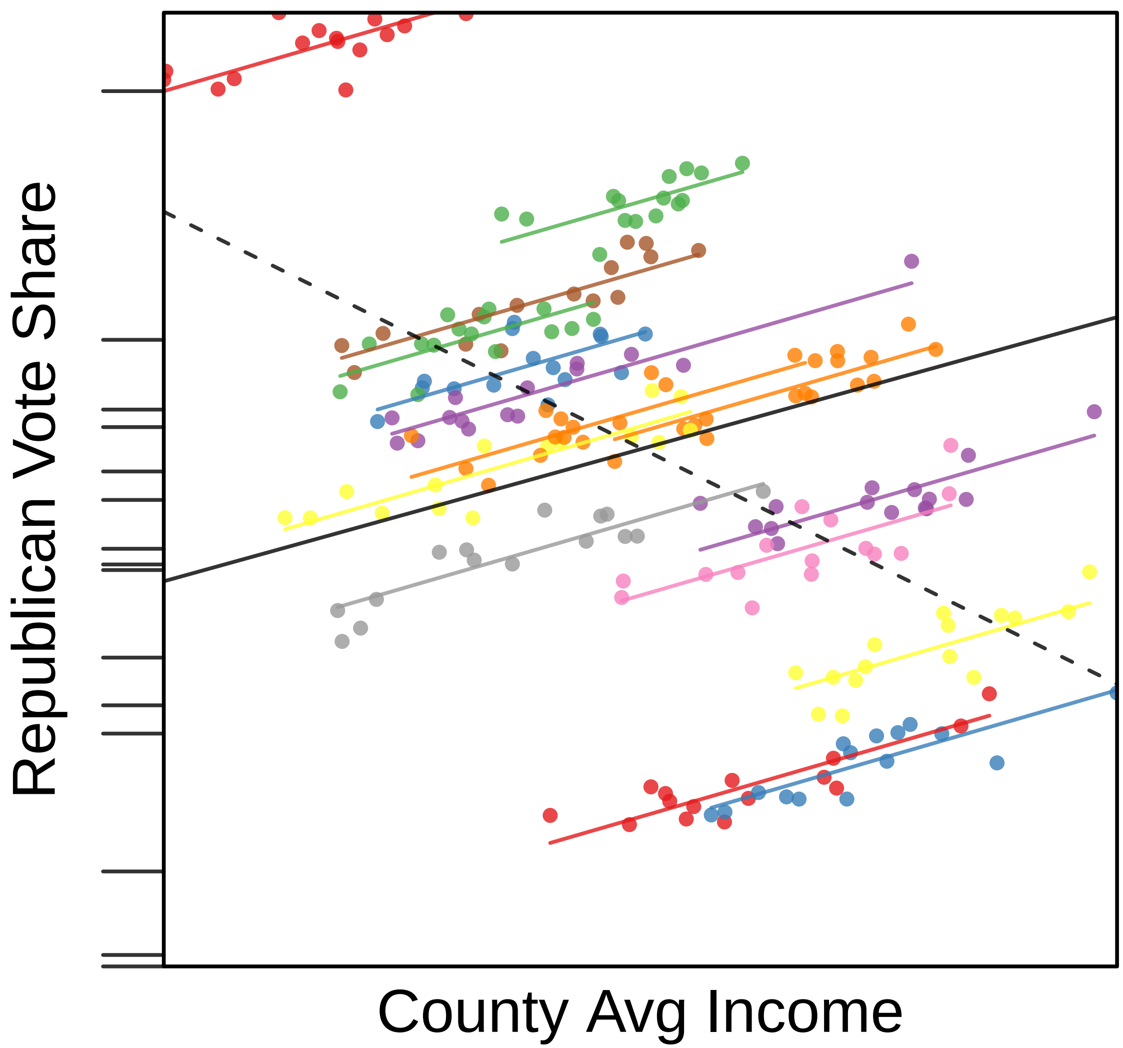
or we could use brute force: include a dummy for each state in the matrix of covariates to purge the omitted variable bias

If we controlled for state fixed effects, our estimate of $\hat{\beta}_1$ would flip signs!



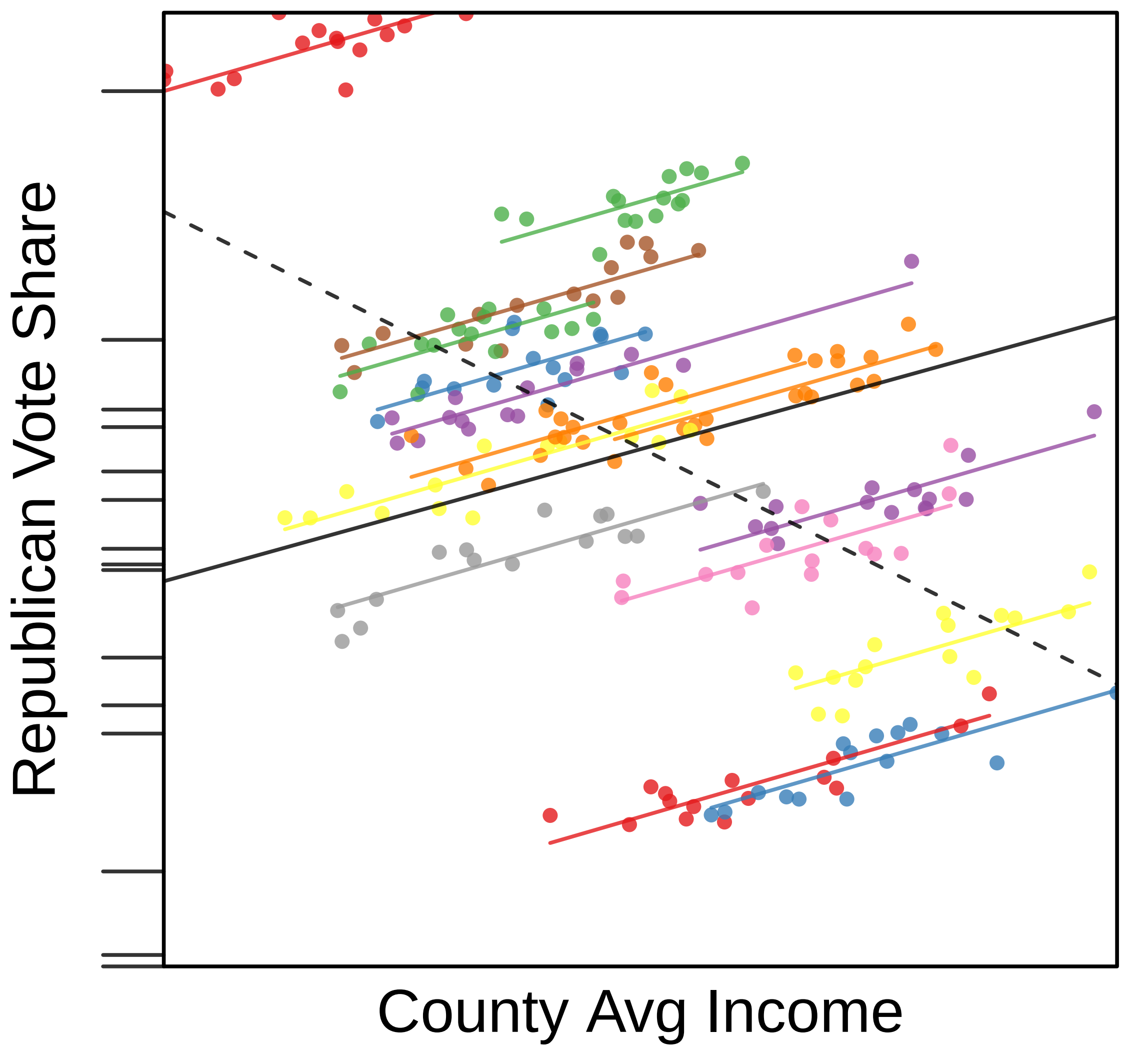
Including fixed effects for each state removes state-level omitted variable bias, and now estimates the correct $\hat{\beta}_1$

What's the matter with Kansas? On average, Kansans are more conservative than other Americans, but within Kansas, the same divide between rich and poor holds



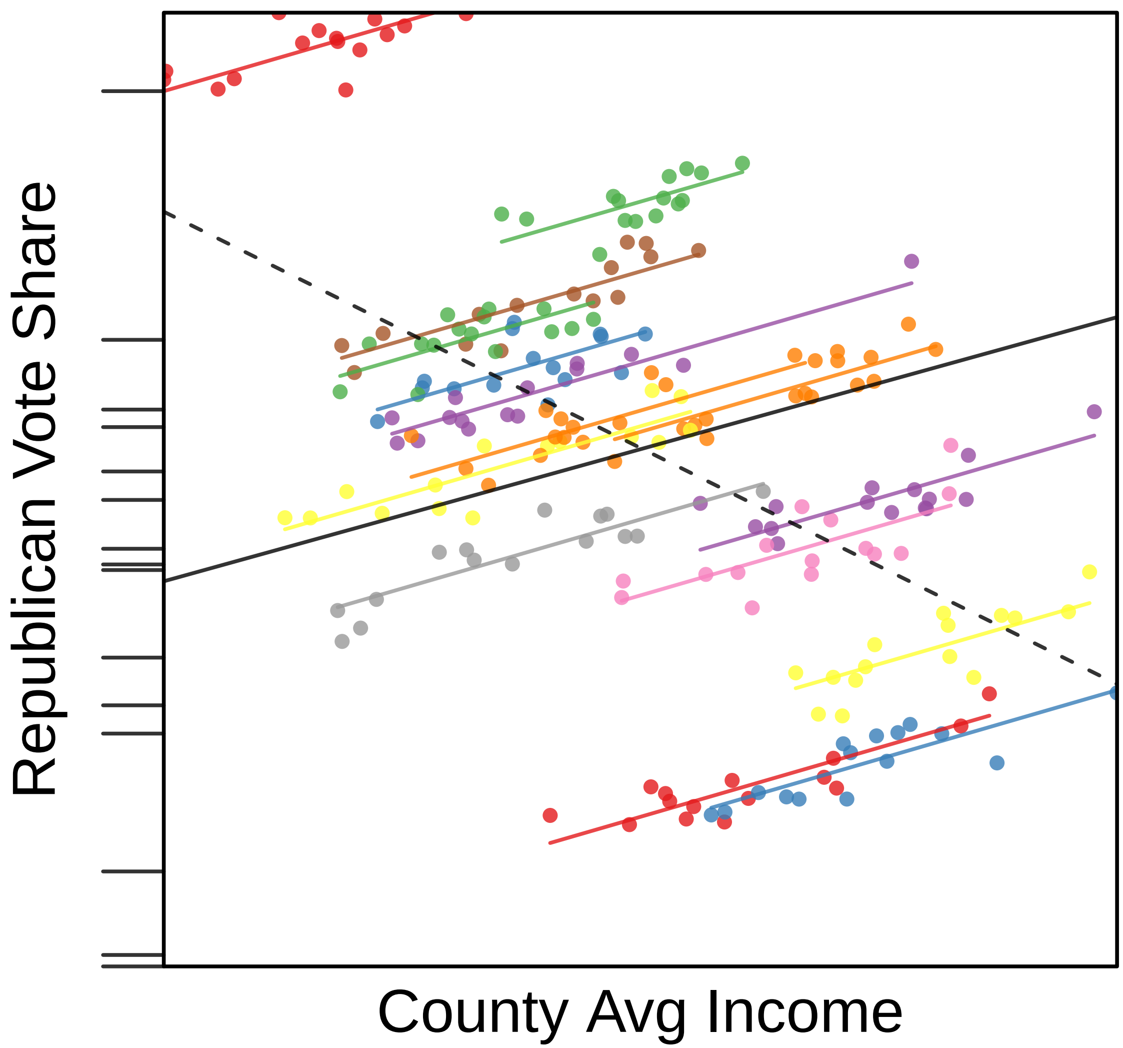
How are fixed effects different from random effects?

Fixed effects control for omitted variables (random effects don't)
Fixed effects don't follow any particular distribution (random effects do)



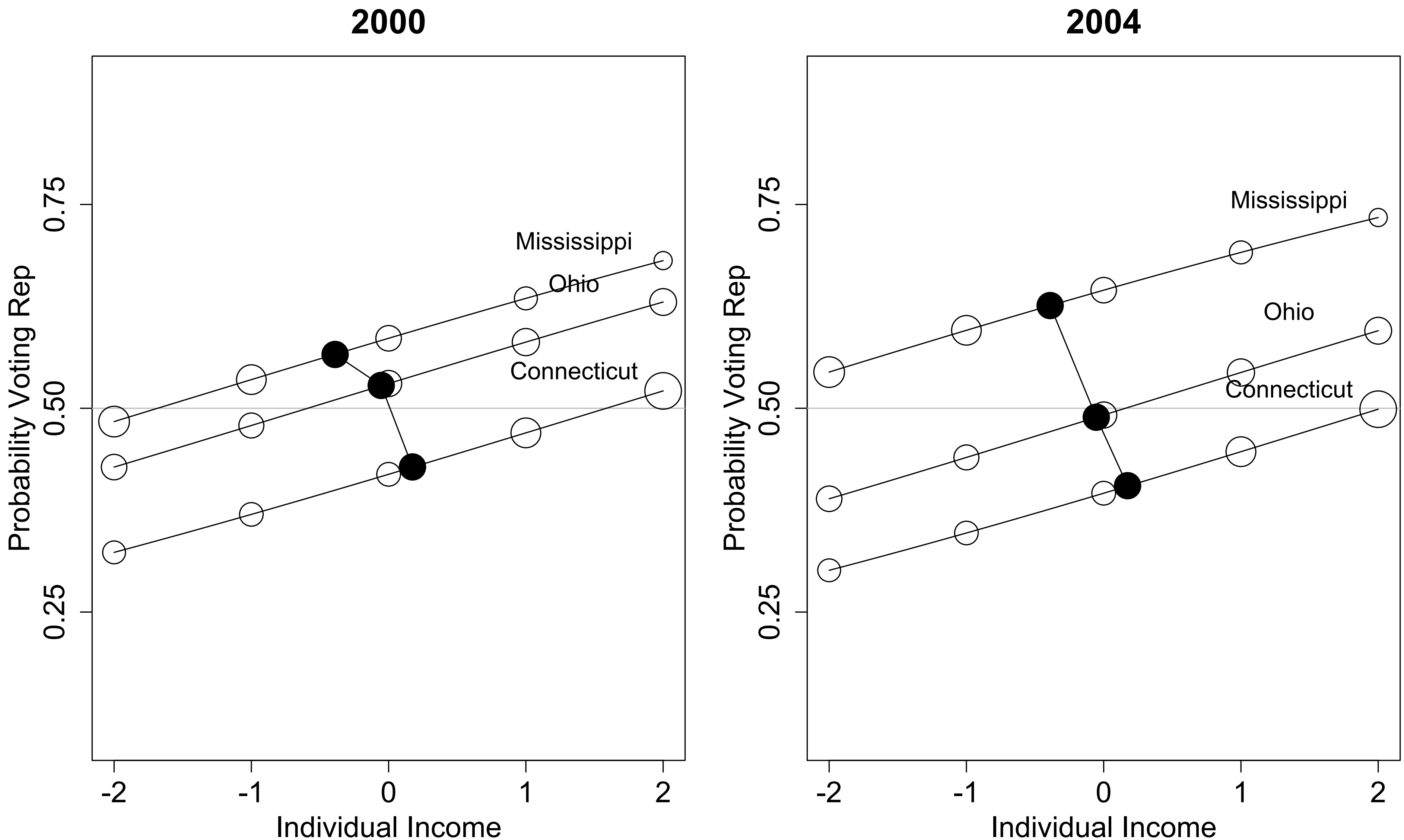
Aside 1: the above reversal is an example of the *ecological fallacy*, which says that aggregate data can mislead us about individual level relationships

Here, the pattern across states mislead us as to the pattern within states



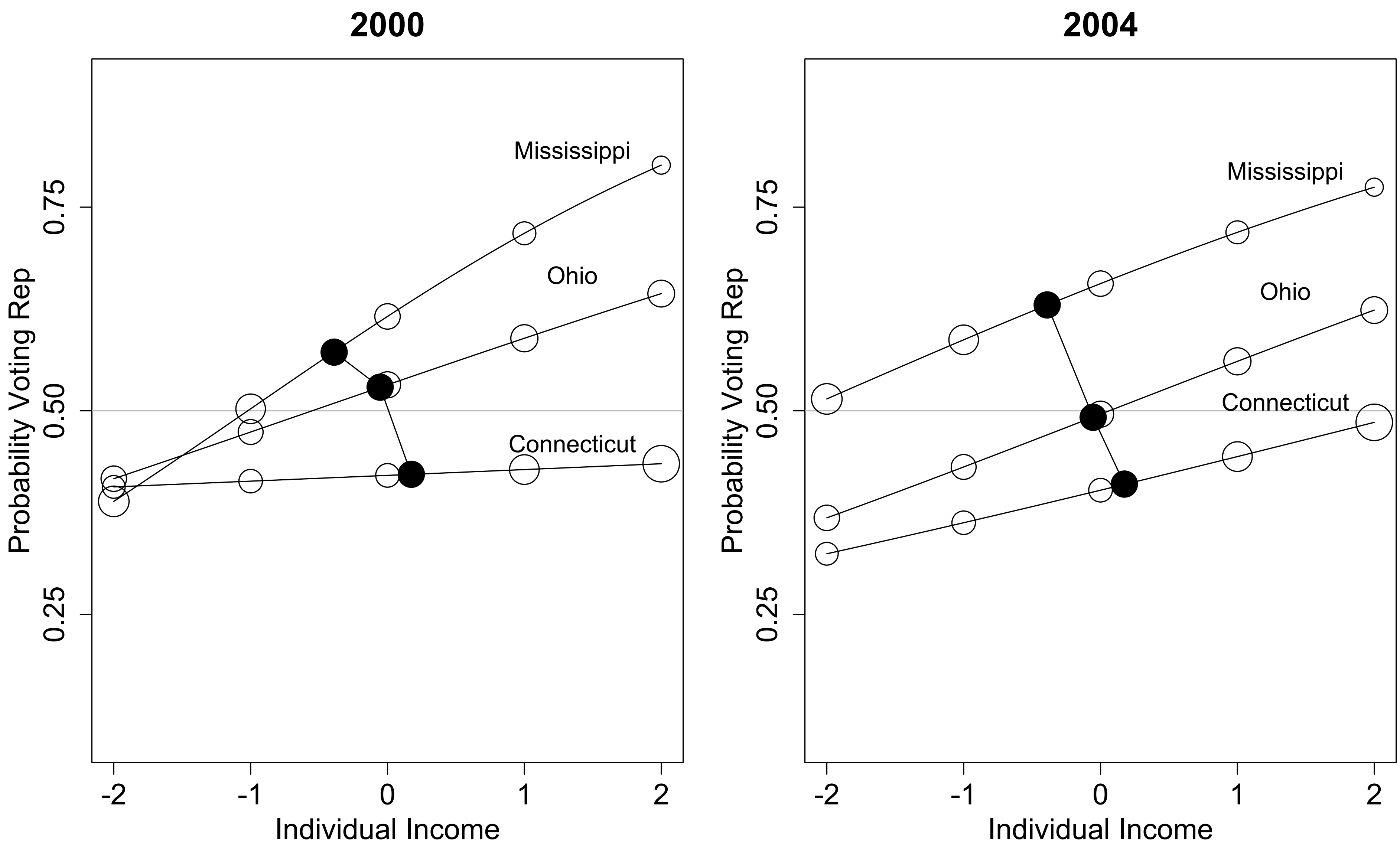
Aside 2: Gelman et al take one more step,
and allow the slopes $\hat{\beta}_{1i}$ of the state level regression lines to vary

They find that the rich-poor divide is actually *steeper* in poor states!



Aside 2: The above are results on actual data from Gelman et al

These results assume the intercepts (but not slopes) vary by state



Aside 2: Gelman et al take one more step,
and allow the slopes $\hat{\beta}_{1i}$ of the state level regression lines to vary

They find that the rich-poor divide is actually *steeper* in poor states!

Mixed effects

$$\alpha_i \sim \mathrm{N}\left(\alpha_i^*, \sigma_\alpha^2\right)$$

Mixed effects give us the best of both worlds

Random effects and fixed effects are just special cases of mixed effects

- Mixed Effects turns into pure Fixed Effects as $\sigma_\alpha^2 \longrightarrow 0$
- Mixed Effects turns into pure Random Effects as $\alpha_i^* \longrightarrow 0$ for all i

If anything, pure FE or RE seem like unreasonable knife-edge cases compared to ME

ME are a natural fit with Bayesian models

More on estimating these models next time

Variable slopes and intercepts

$$\Delta^d y_{it} = \alpha_i + \mathrm{x}_{it}\boldsymbol{\beta}_i + \sum_{p=1}^{P} \Delta^d y_{i,t-p}\phi_p + \sum_{q=1}^{Q} \varepsilon_{i,t-q}\rho_q + \varepsilon_{it}$$

How do we let $\boldsymbol{\beta}_i$ vary over the units?

For the kth covariate x_{kit}, let $\boldsymbol{\beta}_{ki}$ be random, with a multivariate Normal distribution

$$\begin{aligned}
\boldsymbol{\beta}_{ki} &\sim \mathrm{MVN}(\boldsymbol{\beta}^*_{ki}, \Sigma_{\boldsymbol{\beta}_{ki}}) \\
\boldsymbol{\beta}^*_{ki} &= \mathbf{w}_i \boldsymbol{\zeta}
\end{aligned}$$

That is, the β_{ki}'s are now a function of *unit-level covariates* $\mathbf{w}_i$
and their associated *hyperparameters* $\boldsymbol{\zeta}$

Variable slopes and intercepts

$$\text{GDP}_{it} = \phi_1 \text{GDP}_{i,t-1} + \alpha_i + \beta_1 \text{Democracy}_{it} + \varepsilon_{it}$$

Variable slopes and intercepts

$$\text{GDP}_{it} = \phi_1 \text{GDP}_{i,t-1} + \alpha_i + \beta_1 \text{Democracy}_{it} + \varepsilon_{it}$$

$$\alpha_i \sim \text{N}(0, \sigma^2_\alpha)$$

Variable slopes and intercepts

$$
\begin{aligned}
\mathrm{GDP}_{it} &= \phi_1 \mathrm{GDP}_{i,t-1} + \alpha_i + \beta_1 \mathrm{Democracy}_{it} + \varepsilon_{it} \\
\alpha_i &\sim \mathrm{N}(0, \sigma^2_{\alpha}) \\
\beta_1 &\sim \mathrm{N}(\boldsymbol{\beta}^*_{1i}, \sigma^2_{\boldsymbol{\beta}_{1i}})
\end{aligned}
$$

Variable slopes and intercepts

$$\text{GDP}_{it} = \phi_1 \text{GDP}_{i,t-1} + \alpha_i + \beta_1 \text{Democracy}_{it} + \varepsilon_{it}$$
$$\alpha_i \sim \text{N}(0, \sigma^2_{\alpha})$$
$$\beta_1 \sim \text{N}(\boldsymbol{\beta}^*_{1i}, \sigma^2_{\boldsymbol{\beta}_{1i}})$$
$$\beta^*_{1i} = \zeta_0 + \zeta_1 \text{Education}_i$$

Now the effect of Democracy on GDP varies across countries, as a function of their level of Education *and* a country random effect with variance $\sigma^2_{\boldsymbol{\beta}_{1i}}$

This is now a *multilevel* or *hierarchical* model

See Gelman & Hill for a nice textbook on these models

Easiest to accomplish using Bayesian inference
(place priors on each parameter and estimate by MCMC)

Variable slopes and intercepts: Poor man's version

$$\begin{aligned}
\text{GDP}_{it} = \phi_1 \text{GDP}_{i,t-1} + \alpha_i + \beta_1 \text{Democracy}_{it} + \beta_2 \text{Education} \\
+ \beta_3 \text{Democracy} \times \text{Education} + \varepsilon_{it}
\end{aligned}$$

α_i is a matrix of country dummies

This version omits the random effects for α_i and β_i; instead, we have fixed country effects

and a fixed, interactive effect that makes the relation between Democracy and GDP conditional on Education

Should have approximately similar results

Estimating Panel Models

Last time, we discussed how including random and/or fixed effects changes the properties of our estimators of β

Today, we'll talk about how to estimate and interpet panel models using fixed and/or random effects

And how to decide if we need (or even can use) fixed effects

We can always add random effects, but in some cases FEs either be too costly to estimate (in terms of dfs), or simply impossible to estimate

Estimating Fixed Effects Models

Option 1: Fixed effects or "within" estimator:

$$y_{it} - \bar{y}_i = (\mathrm{x}_{it} - \bar{\mathrm{x}}_i)\boldsymbol{\beta} + (u_{it} - \bar{u}_{it})$$

- estimating the fixed effects by differencing them out
- including time-invariant variables directly in x_{it} impossible here
- (there are workarounds, e.g., if we have an instrument for the time-invariant variable that is uncorrelated with the fixed effects; see Hausman-Taylor)
- suggests a complementary "between" estimator of $\bar{y}_i$ on $\bar{x}_i$ which could include time-invariant x_i; together these models explain the variance in y_{it}
- does not actually provide estimates of the fixed effects themselves; just purges them from the model to remove omitted time-invariant variables

Estimating Fixed Effects Models

Option 2: Dummy variable estimator (sometimes called LSDV)

$$y_{it} = \mathrm{x}_{it}\boldsymbol{\beta} + \alpha_i + u_{it}$$

- yields estimates of α_i fixed effects (may be useful in quest for omitted variables; see if the α_i look like a variable you know)
- for large T, should be very similar to FE estimator
- not a good idea for very small T: estimates of α_i will be poor

Time-Invariant Covariates & Fixed Effects

We can't include time-invariant variables in fixed effects models

If we do, we will have perfect collinearity, and can't get estimates

That is, we will get some parameter estimates equal to `NA`

Never report a regression with `NA` parameters

The regression you tried to run was impossible. Start over with a possible one.

Time-Invariant Covariates & Fixed Effects

If we can't include time-invariant variables in a fixed effects model,
does that mean time-invariant variables can never explain changes over time?

Time-Invariant Covariates & Fixed Effects

If we can't include time-invariant variables in a fixed effects model,
does that mean time-invariant variables can never explain changes over time?

You might think so: how can a constant explain a variable?

But time-invariant variables could still effect time-varying outcomes in a special way. . .

Time-Invariant Covariates & Fixed Effects

If we can't include time-invariant variables in a fixed effects model,
does that mean time-invariant variables can never explain changes over time?

You might think so: how can a constant explain a variable?

But time-invariant variables could still effect time-varying outcomes in a special way. . .

Time-invariant variables can influence how a unit weathers time-varying *shocks* in some other variable

Example: labor market regulations (e.g. employment protection) don't change much over time

Blanchard & Wolfers found that when a negative economic shock hits,
unemployment may rebound more slowly where such protections are stronger

Time-Invariant Covariates & Fixed Effects

We can model how a slow moving or time-invariant covariate conditions the effect of a quickly changing covariate on y_{it}

To estimate how a time-invariant covariate x_{it} mediates the effect of a shock, s_{it}, include on the RHS $x_{it} \times s_{it}$ and s_{it}, while omitting x_{it} itself

(It's okay *and necessary* to omit the x_{it} base term in this special case, because α_i already captures the effect of x_{it})

Many theories about institutions can be tested this way

Time-Invariant Covariates & Fixed Effects

What if we want to "include" time-invariant covariates' effect on the long term average level of y?

We might partition the fixed effect into:

1. the portion "explained" by known time-invariant variables and

2. the portion still unexplained

Plümper & Troeger have methods to do this.

In this case, our estimates of the time-invariant effects are vulnerable to omitted variable bias from unmeasured time-invariant variables, even though time varying variables in the model are not

Thus you now need to control for *lots* of time-invariant variables directly, even hard to measure ones like culture

Estimating Random & Mixed Effects Models

Estimation of random effects is by maximum likelihood (ML) or generalized least squares (GLS)

Technically we're just adding one parameter to estimate: the variance of the random effects, σ^2_α

This is partitioned out of the overall variance, σ^2

Can understand this most easily by abstracting away from time series for a moment

Estimating Random & Mixed Effects Models

Recall that for linear regression, we assume homoskedastic, serially uncorrelated errors, and thus a variance-covariance matrix like this:

$$\Omega = \begin{bmatrix} \sigma^2 & 0 & 0 & 0 \\ 0 & \sigma^2 & 0 & 0 \\ 0 & 0 & \sigma^2 & 0 \\ 0 & 0 & 0 & \sigma^2 \end{bmatrix}$$

Estimating Random & Mixed Effects Models

And recall that heteroskedastic (but serially uncorrelated) errors have this variance-covariance matrix

$$\Omega = \begin{bmatrix} \sigma_1^2 & 0 & 0 & 0 \\ 0 & \sigma_2^2 & 0 & 0 \\ 0 & 0 & \sigma_3^2 & 0 \\ 0 & 0 & 0 & \sigma_4^2 \end{bmatrix}$$

Estimating Random & Mixed Effects Models

And finally, remember heteroskedastic, serially correlated errors follow this general form of variance-covariance

$$\Omega = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \sigma_{13} & \sigma_{14} \\ \sigma_{21} & \sigma_2^2 & \sigma_{23} & \sigma_{24} \\ \sigma_{31} & \sigma_{32} & \sigma_3^2 & \sigma_{34} \\ \sigma_{41} & \sigma_{42} & \sigma_{43} & \sigma_4^2 \end{bmatrix}$$

What does this matrix look like for random effects with no serial correlation?

Estimating Random & Mixed Effects Models

Define the variance of the random effect as

$$\mathrm{E}(\alpha_i^2) = \sigma_\alpha^2 = \mathrm{var}(\alpha_i)$$

Estimating Random & Mixed Effects Models

Define the variance of the random effect as

$$\mathrm{E}(\alpha_i^2) = \sigma_\alpha^2 = \mathrm{var}(\alpha_i)$$

Define the expected value of the squared white noise term as σ_ε^2

$$\mathrm{E}(\varepsilon_{it}^2) = \sigma_\varepsilon^2 = \mathrm{var}(\varepsilon_{it})$$

Estimating Random & Mixed Effects Models

Define the variance of the random effect as

$$\mathrm{E}(\alpha_i^2) = \sigma_\alpha^2 = \mathrm{var}(\alpha_i)$$

Define the expected value of the squared white noise term as σ_ε^2

$$\mathrm{E}(\varepsilon_{it}^2) = \sigma_\varepsilon^2 = \mathrm{var}(\varepsilon_{it})$$

White noise is serially uncorrelated, so has covariance 0 for $t \neq s$:

$$\mathrm{E}(\varepsilon_{it}\varepsilon_{is}) = 0 = \mathrm{cov}(\varepsilon_{it}, \varepsilon_{is})$$

Estimating Random & Mixed Effects Models

Define the variance of the random effect as

$$\mathrm{E}(\alpha_i^2) = \sigma_\alpha^2 = \mathrm{var}(\alpha_i)$$

Define the expected value of the squared white noise term as σ_ε^2

$$\mathrm{E}(\varepsilon_{it}^2) = \sigma_\varepsilon^2 = \mathrm{var}(\varepsilon_{it})$$

White noise is serially uncorrelated, so has covariance 0 for $t \neq s$:

$$\mathrm{E}(\varepsilon_{it}\varepsilon_{is}) = 0 = \mathrm{cov}(\varepsilon_{it}, \varepsilon_{is})$$

Finally, note that we assumed the white noise error and random effect are uncorrelated,

$$\mathrm{E}(\alpha_i\varepsilon_{it}) = 0 = \mathrm{cov}(\alpha_i, \varepsilon_{it})$$

Estimating Random & Mixed Effects Models

Thus the variance of the whole random component of the model is

$$\mathrm{E}((\alpha_i + \varepsilon_{it})(\alpha_i + \varepsilon_{it})) = \mathrm{E}(\alpha_i^2) + 2\mathrm{E}(\alpha_i \varepsilon_{it}) + \mathrm{E}(\varepsilon_{it}^2)$$

Estimating Random & Mixed Effects Models

Thus the variance of the whole random component of the model is

$$\begin{aligned}
\mathrm{E}((\alpha_i + \varepsilon_{it})(\alpha_i + \varepsilon_{it})) &= \mathrm{E}(\alpha_i^2) + 2\mathrm{E}(\alpha_i \varepsilon_{it}) + \mathrm{E}(\varepsilon_{it}^2) \\
&= \sigma_\alpha^2 + 0 + \sigma_\varepsilon^2
\end{aligned}$$

Estimating Random & Mixed Effects Models

Thus the variance of the whole random component of the model is

$$\begin{aligned}
\mathrm{E}((\alpha_i + \varepsilon_{it})(\alpha_i + \varepsilon_{it})) &= \mathrm{E}(\alpha_i^2) + 2\mathrm{E}(\alpha_i \varepsilon_{it}) + \mathrm{E}(\varepsilon_{it}^2) \\
&= \sigma_\alpha^2 + 0 + \sigma_\varepsilon^2 \\
&= \sigma_\alpha^2 + \sigma_\varepsilon^2
\end{aligned}$$

Estimating Random & Mixed Effects Models

Thus the variance of the whole random component of the model is

$$\begin{aligned}
\mathrm{E}((\alpha_i + \varepsilon_{it})(\alpha_i + \varepsilon_{it})) &= \mathrm{E}(\alpha_i^2) + 2\mathrm{E}(\alpha_i \varepsilon_{it}) + \mathrm{E}(\varepsilon_{it}^2) \\
&= \sigma_\alpha^2 + 0 + \sigma_\varepsilon^2 \\
&= \sigma_\alpha^2 + \sigma_\varepsilon^2
\end{aligned}$$

And the covariance of the whole random component is:

$$\mathrm{E}((\alpha_i + \varepsilon_{it})(\alpha_i + \varepsilon_{is})) = \mathrm{E}(\alpha_i^2) + \mathrm{E}(\alpha_i \varepsilon_{is}) + \mathrm{E}(\alpha_i \varepsilon_{it}) + \mathrm{E}(\varepsilon_{it} \varepsilon_{is})$$

Estimating Random & Mixed Effects Models

Thus the variance of the whole random component of the model is

$$\begin{aligned}
\mathrm{E}((\alpha_i + \varepsilon_{it})(\alpha_i + \varepsilon_{it})) &= \mathrm{E}(\alpha_i^2) + 2\mathrm{E}(\alpha_i \varepsilon_{it}) + \mathrm{E}(\varepsilon_{it}^2) \\
&= \sigma_\alpha^2 + 0 + \sigma_\varepsilon^2 \\
&= \sigma_\alpha^2 + \sigma_\varepsilon^2
\end{aligned}$$

And the covariance of the whole random component is:

$$\begin{aligned}
\mathrm{E}((\alpha_i + \varepsilon_{it})(\alpha_i + \varepsilon_{is})) &= \mathrm{E}(\alpha_i^2) + \mathrm{E}(\alpha_i \varepsilon_{is}) + \mathrm{E}(\alpha_i \varepsilon_{it}) + \mathrm{E}(\varepsilon_{it} \varepsilon_{is}) \\
&= \sigma_\alpha^2 + 0 + 0 + 0
\end{aligned}$$

Estimating Random & Mixed Effects Models

Thus the variance of the whole random component of the model is

$$\begin{aligned} \mathrm{E}((\alpha_i + \varepsilon_{it})(\alpha_i + \varepsilon_{it})) &= \mathrm{E}(\alpha_i^2) + 2\mathrm{E}(\alpha_i \varepsilon_{it}) + \mathrm{E}(\varepsilon_{it}^2) \\ &= \sigma_\alpha^2 + 0 + \sigma_\varepsilon^2 \\ &= \sigma_\alpha^2 + \sigma_\varepsilon^2 \end{aligned}$$

And the covariance of the whole random component is:

$$\begin{aligned} \mathrm{E}((\alpha_i + \varepsilon_{it})(\alpha_i + \varepsilon_{is})) &= \mathrm{E}(\alpha_i^2) + \mathrm{E}(\alpha_i \varepsilon_{is}) + \mathrm{E}(\alpha_i \varepsilon_{it}) + \mathrm{E}(\varepsilon_{it} \varepsilon_{is}) \\ &= \sigma_\alpha^2 + 0 + 0 + 0 \\ &= \sigma_\alpha^2 \end{aligned}$$

Estimating Random & Mixed Effects Models

If our data have a single random effect in the mean for each unit
$\longrightarrow$ serially correlated errors, but expressable using only two variances:

- the random effects variance σ_α^2
- the white noise term's variance σ_ε^2

$$\Omega = \begin{bmatrix} \sigma_\alpha^2 + \sigma_\varepsilon^2 & \sigma_\alpha^2 & \sigma_\alpha^2 & \sigma_\alpha^2 \\ \sigma_\alpha^2 & \sigma_\alpha^2 + \sigma_\varepsilon^2 & \sigma_\alpha^2 & \sigma_\alpha^2 \\ \sigma_\alpha^2 & \sigma_\alpha^2 & \sigma_\alpha^2 + \sigma_\varepsilon^2 & \sigma_\alpha^2 \\ \sigma_\alpha^2 & \sigma_\alpha^2 & \sigma_\alpha^2 & \sigma_\alpha^2 + \sigma_\varepsilon^2 \end{bmatrix}$$

Estimating Random & Mixed Effects Models

We have drastically simplified this matrix, and can now use FGLS (Feasible Generalized Least Squares) or ML to estimate it

$$\hat{\boldsymbol{\beta}}_{\mathrm{GLS}} = \left(\sum_{i=1}^{N} \mathrm{X}_i' \Omega^{-1} \mathrm{X}_i \right)^{-1} \left(\sum_{i=1}^{N} \mathrm{X}_i' \hat{\Omega}^{-1} \mathrm{y}_i \right)$$

where X_i is the $T \times K$ matrix of covariates for unit i, all times $t = 1, \ldots T$, and all K covariates

All we need are the estimates $\hat{\sigma}^2_\alpha$ and $\hat{\sigma}^2_\varepsilon$, and we can calculate $\hat{\boldsymbol{\beta}}_{\mathrm{GLS}}$

Estimating Random & Mixed Effects Models

We get $\hat{\sigma}_\varepsilon^2$ from the residuals from a LS regression:

$$\hat{\sigma}_\varepsilon^2 = \frac{1}{NT - K} \sum_{i=1}^{N} \sum_{t=1}^{T} \hat{\varepsilon}_{it\mathrm{LS}}^2$$

(This is the usual estimator, but for NT observations)

Estimating Random & Mixed Effects Models

To get an estimator of $\hat{\sigma}^2_\alpha$, we need to adjust for the fact that we have only so many unique pairs of errors to compare:

$$\hat{\sigma}^2_\alpha = \mathrm{E}\left(\sum_{t=1}^{T-1}\sum_{s=t+1}^{T}(\alpha_i + \varepsilon_{it})(\alpha_i + \varepsilon_{is})\right)$$

Estimating Random & Mixed Effects Models

To get an estimator of $\hat{\sigma}_\alpha^2$, we need to adjust for the fact that we have only so many unique pairs of errors to compare:

$$\hat{\sigma}_\alpha^2 = \mathrm{E}\left(\sum_{t=1}^{T-1}\sum_{s=t+1}^{T}(\alpha_i + \varepsilon_{it})(\alpha_i + \varepsilon_{is})\right)$$

$$= \mathrm{E}\sum_{t=1}^{T-1}\sum_{s=t+1}^{T}((\alpha_i + \varepsilon_{it})(\alpha_i + \varepsilon_{is}))$$

Estimating Random & Mixed Effects Models

To get an estimator of $\hat{\sigma}^2_\alpha$, we need to adjust for the fact that we have only so many unique pairs of errors to compare:

$$\begin{aligned}
\hat{\sigma}^2_\alpha &= \mathrm{E}\left(\sum_{t=1}^{T-1}\sum_{s=t+1}^{T}(\alpha_i+\varepsilon_{it})(\alpha_i+\varepsilon_{is})\right) \\
&= \mathrm{E}\sum_{t=1}^{T-1}\sum_{s=t+1}^{T}((\alpha_i+\varepsilon_{it})(\alpha_i+\varepsilon_{is})) \\
&= \sum_{t=1}^{T-1}\sum_{s=t+1}^{T}\sigma^2_\alpha
\end{aligned}$$

Estimating Random & Mixed Effects Models

To get an estimator of $\hat{\sigma}^2_\alpha$, we need to adjust for the fact that we have only so many unique pairs of errors to compare:

$$\begin{aligned}
\hat{\sigma}^2_\alpha &= \mathrm{E}\left(\sum_{t=1}^{T-1}\sum_{s=t+1}^{T}(\alpha_i+\varepsilon_{it})(\alpha_i+\varepsilon_{is})\right) \\
&= \mathrm{E}\sum_{t=1}^{T-1}\sum_{s=t+1}^{T}\left((\alpha_i+\varepsilon_{it})(\alpha_i+\varepsilon_{is})\right) \\
&= \sum_{t=1}^{T-1}\sum_{s=t+1}^{T}\sigma^2_\alpha \\
&= \sigma^2_\alpha\sum_{t=1}^{T-1}(T-t)
\end{aligned}$$

Estimating Random & Mixed Effects Models

To get an estimator of $\hat{\sigma}_\alpha^2$, we need to adjust for the fact that we have only so many unique pairs of errors to compare:

$$\begin{aligned}
\hat{\sigma}_\alpha^2 &= \mathrm{E}\left(\sum_{t=1}^{T-1}\sum_{s=t+1}^{T}(\alpha_i+\varepsilon_{it})(\alpha_i+\varepsilon_{is})\right)\\
&= \mathrm{E}\sum_{t=1}^{T-1}\sum_{s=t+1}^{T}((\alpha_i+\varepsilon_{it})(\alpha_i+\varepsilon_{is}))\\
&= \sum_{t=1}^{T-1}\sum_{s=t+1}^{T}\sigma_\alpha^2\\
&= \sigma_\alpha^2\sum_{t=1}^{T-1}(T-t)\\
&= \sigma_\alpha^2((T-1)+(T-2)+\ldots+2+1)
\end{aligned}$$

Estimating Random & Mixed Effects Models

$$= \sigma_\alpha^2((T-1)+(T-2)+\ldots+2+1)$$

Estimating Random & Mixed Effects Models

$$
\begin{aligned}
&= \sigma_\alpha^2((T-1) + (T-2) + \ldots + 2 + 1) \\
&= \sigma_\alpha^2 T(T-1)/2
\end{aligned}
$$

Estimating Random & Mixed Effects Models

$$= \sigma_\alpha^2((T-1) + (T-2) + \ldots + 2 + 1)$$
$$= \sigma_\alpha^2 T(T-1)/2$$

$$\hat{\sigma}_\alpha^2 = \frac{1}{NT(T-1)/2 - K} \sum_{i=1}^{N} \sum_{t=1}^{T-1} \sum_{s=t+1}^{T} \hat{\varepsilon}_{it}\hat{\varepsilon}_{is}$$

where in the last step we replace σ_α^2 with its estimator from pooled LS (the average of the products of the unique pairs of residuals)

Estimating Random & Mixed Effects Models

$$
\begin{aligned}
&= \sigma_\alpha^2((T-1) + (T-2) + \ldots + 2 + 1) \\
&= \sigma_\alpha^2 T(T-1)/2
\end{aligned}
$$

$$\hat{\sigma}_\alpha^2 = \frac{1}{NT(T-1)/2 - K} \sum_{i=1}^{N} \sum_{t=1}^{T-1} \sum_{s=t+1}^{T} \hat{\varepsilon}_{it} \hat{\varepsilon}_{is}$$

where in the last step we replace σ_α^2 with its estimator from pooled LS (the average of the products of the unique pairs of residuals)

With some algebra, this approach extends to serial correlaton of other kinds (ARMA)

For complex models, with many levels and/or hyperparameters, best to go Bayesian, set diffuse priors on the parameters, and use MCMC

Selecting Fixed Effects vs Random Effects Models

Choosing random effects when α_i is actually correlated with x_{it}
will lead to omitted variable bias

Choosing fixed effects when α_i is really uncorrelated with x_{it}
will lead to inefficient estimates of β (compared to random effects estimation)
and kick out our time-invariant variables

Often in comparative we are certain there are important omitted time invariant variables (culture, unmeasured institutions, long effects of history)

So choice to include fixed effects requires nothing more than theory

Still could include random effects in addition to the fixed effects

Selecting Fixed Effects vs Random Effects Models

But if we are uncertain, or want to check against estimating unnecessary fixed effects, we can use the Hausman test for (any) fixed effexts versus just having random effecxts

Hausman sets up the null hypothesis of random effects

Attempts to reject it in favor of fixed effects

Selecting Fixed Effects vs Random Effects Models

But if we are uncertain, or want to check against estimating unnecessary fixed effects, we can use the Hausman test for (any) fixed effexts versus just having random effecxts

Hausman sets up the null hypothesis of random effects

Attempts to reject it in favor of fixed effects

Checks whether the random α_i's are correlated with $\mathrm{x_i}$ under the null

Does this by calculating the variance-covariance matrices of regressors under FE and then just RE

Null is no correlation between these covariances

Selecting Fixed Effects vs Random Effects Models

But if we are uncertain, or want to check against estimating unnecessary fixed effects, we can use the Hausman test for (any) fixed effexts versus just having random effecxts

Hausman sets up the null hypothesis of random effects

Attempts to reject it in favor of fixed effects

Checks whether the random α_i's are correlated with $\mathrm{x_i}$ under the null

Does this by calculating the variance-covariance matrices of regressors under FE and then just RE

Null is no correlation between these covariances

If there is no correlation, that means the regressors do not predict the random effects (ie, are uncorrelated)

Rejecting the null suggests you may need fixed effects to deal with omitted variable bias

`phtest` in `plm` library

Interpreting Random Effects Models

Usually, interest focuses on the percentage of variance explained by the random effects

And how this variance compares to that remaining in the model

Reported by your estimation routine

What if T is very small?

If T is very small (< 15 perhaps), estimating panel dynamics efficiently and without bias gets harder

In these cases, we should investigate alternatives:

1. First differencing the series to produce a stationary, hopefully white noise processs

2. Including fixed effects for the time period (time dummies)

3. Checking for serial correlation after estimation (LM test)

4. Using lags of the dependent variable, while removing the bias from including lags with fixed effects by instrumenting with lagged differences (Arellano-Bond)

Example: GDP in a panel

Let's use the Przeworski et al democracy data to try out our variable intercept models

This exercise is for pedagogical purposes only; the models we fit are badly specified

We will investigate the following model:

$$\Delta^d \mathrm{GDP}_{it} = \alpha_i + \beta_1 \mathrm{OIL}_{it} + \beta_2 \mathrm{REG}_{it} + \beta_3 \mathrm{EDT}_{it} + \nu_{it}$$

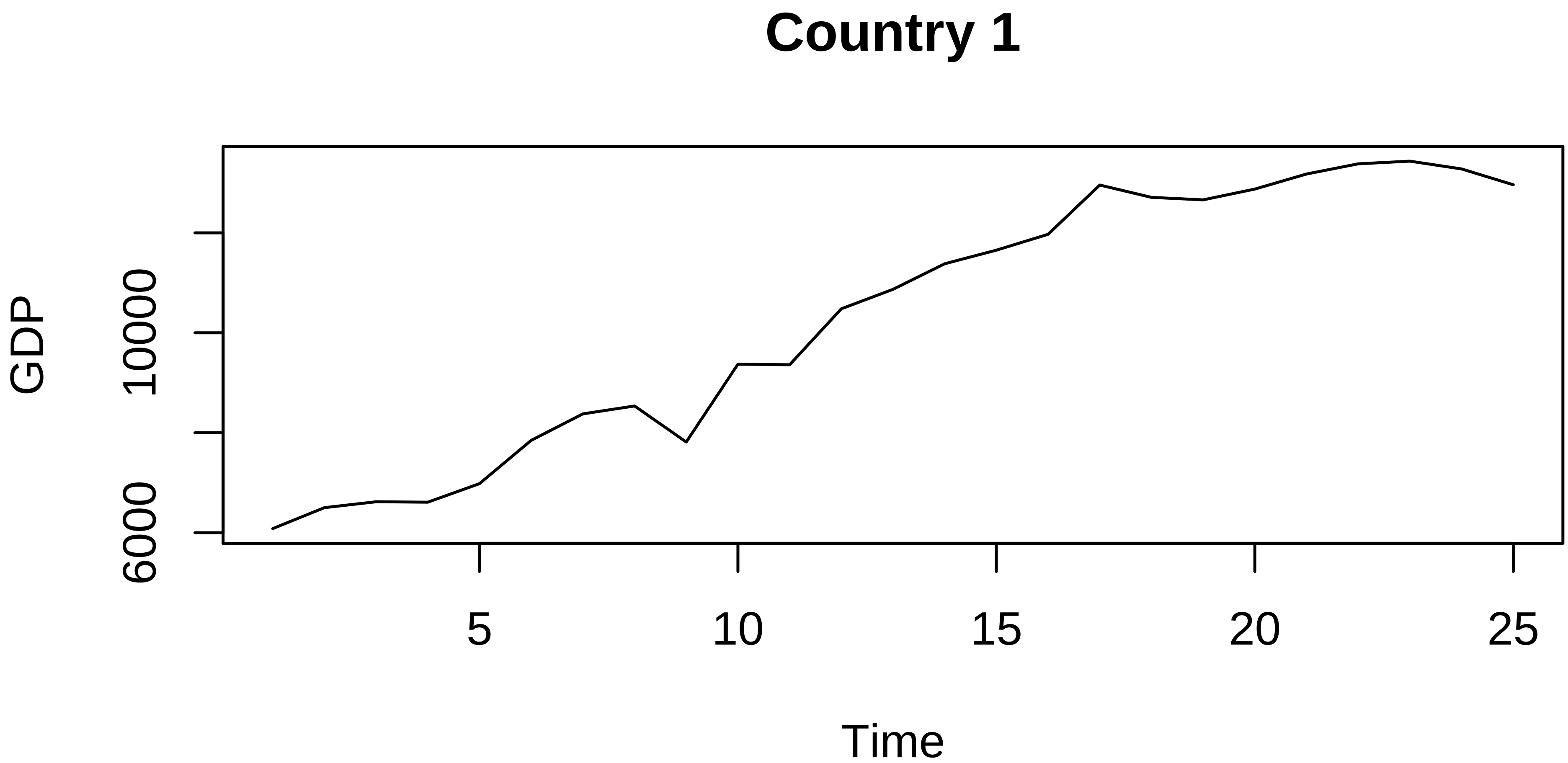
- where $\nu_{it} \sim \mathrm{ARIMA}(p, d, q)$,
- d may be 0 or 1, and
- α_i may be fixed, random, or a mixed

Example: GDP in a panel

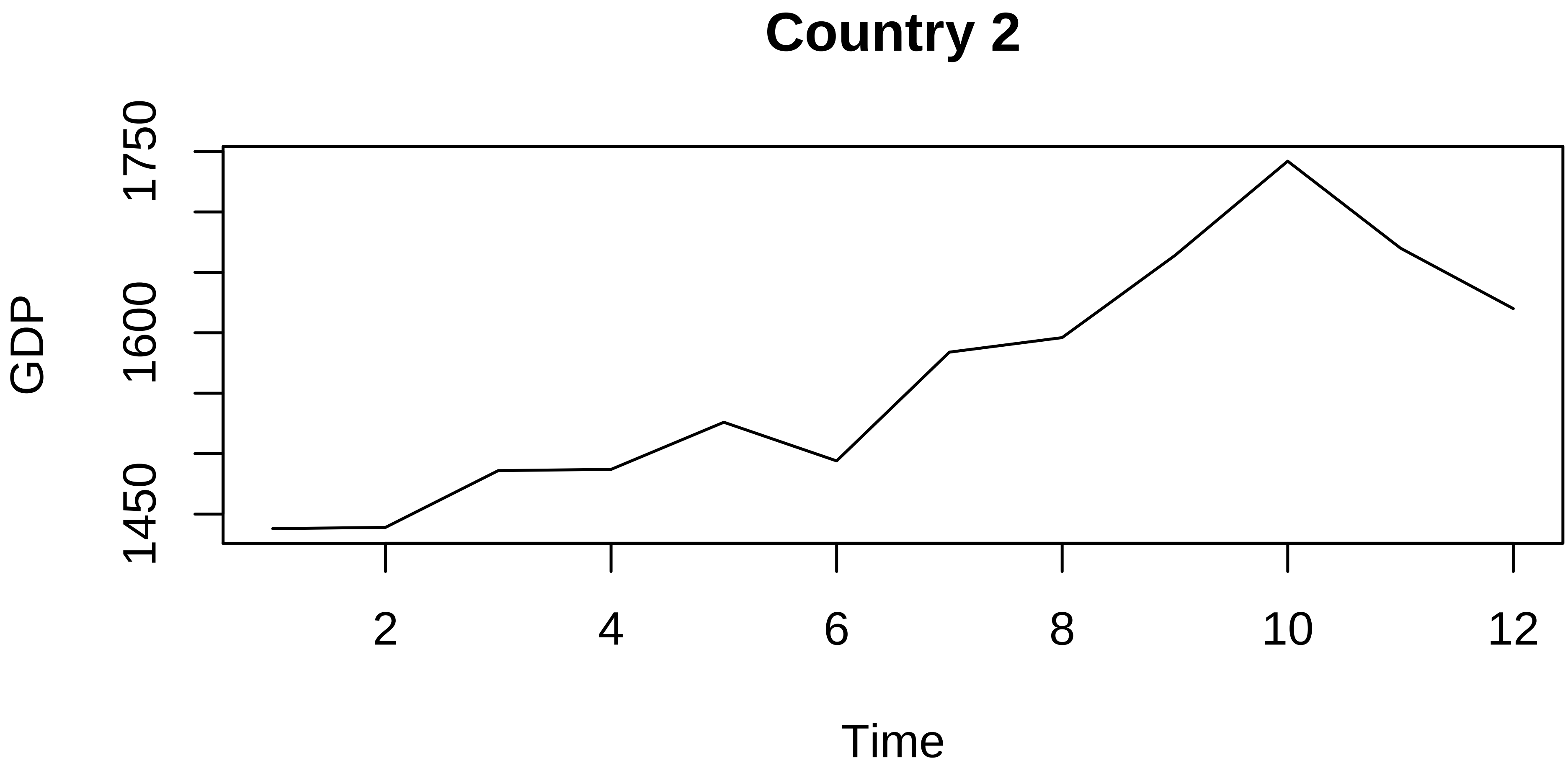
We first investigate the time series properties of GDP

But we have $N = 113$ countries! So we would have to look at 113 time series plots, 113 ACF plots, and 113 PACF plots

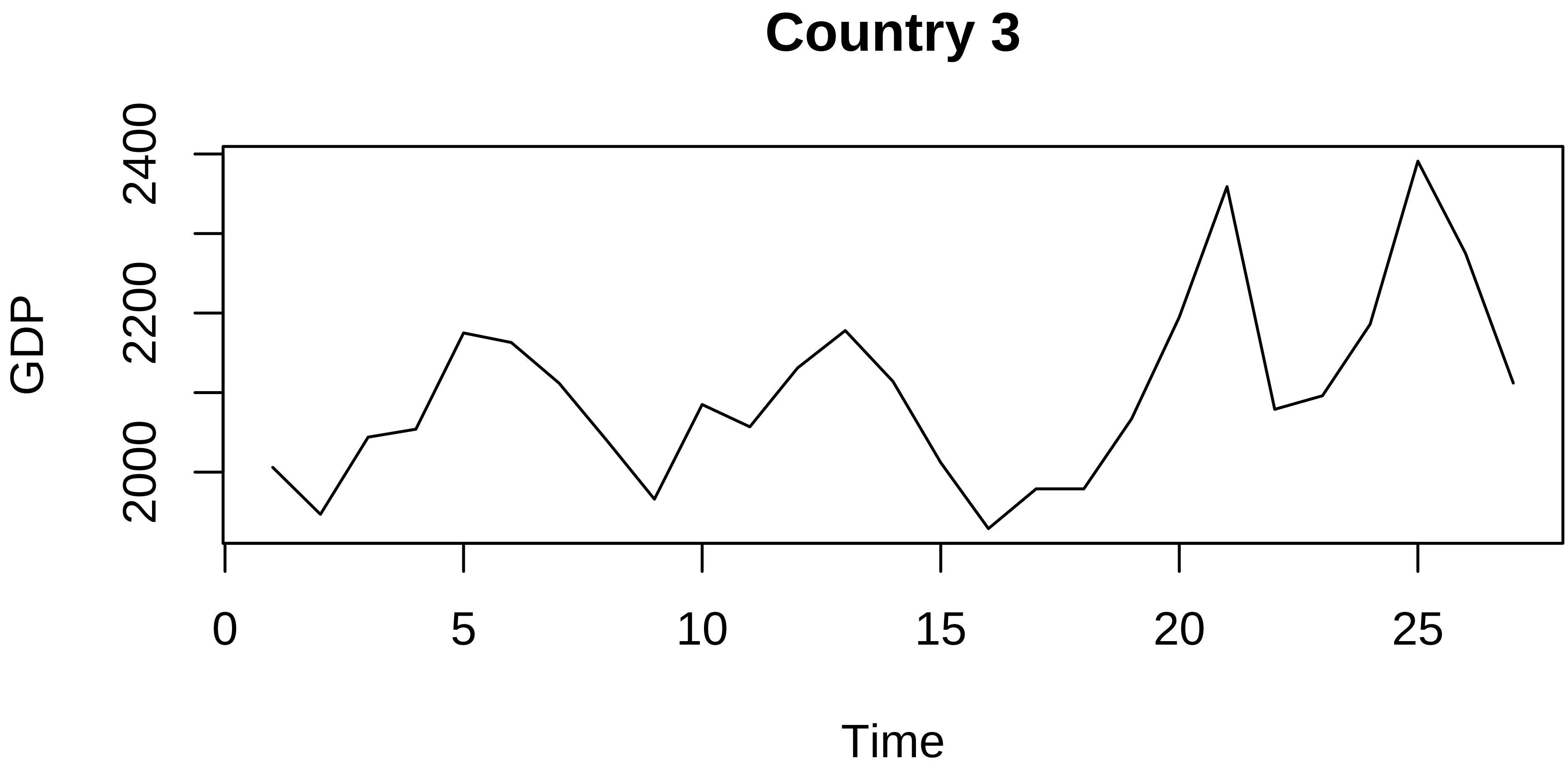
Fortunately, they do look fairly similar. . .



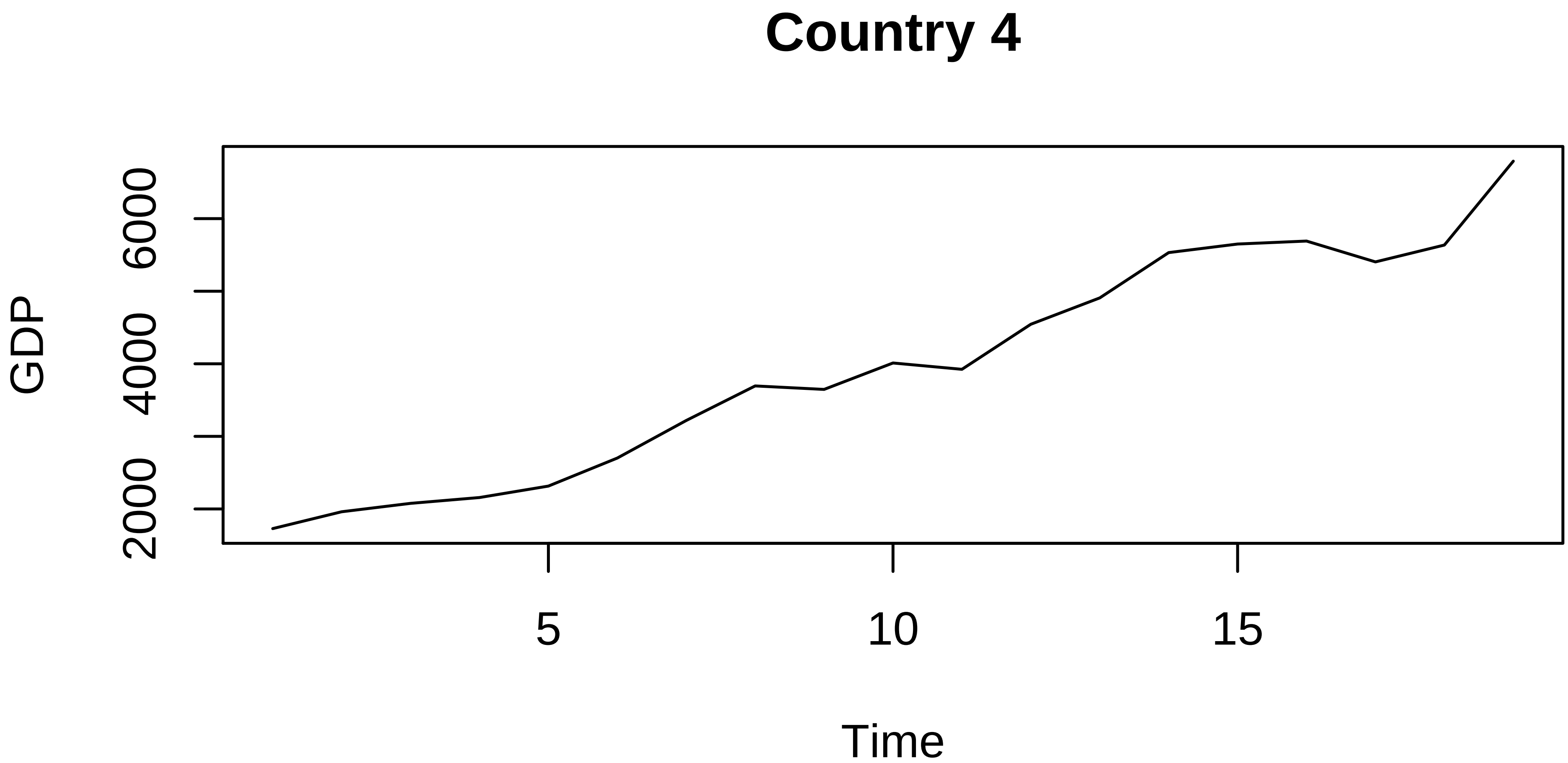
GDP time series for country 1



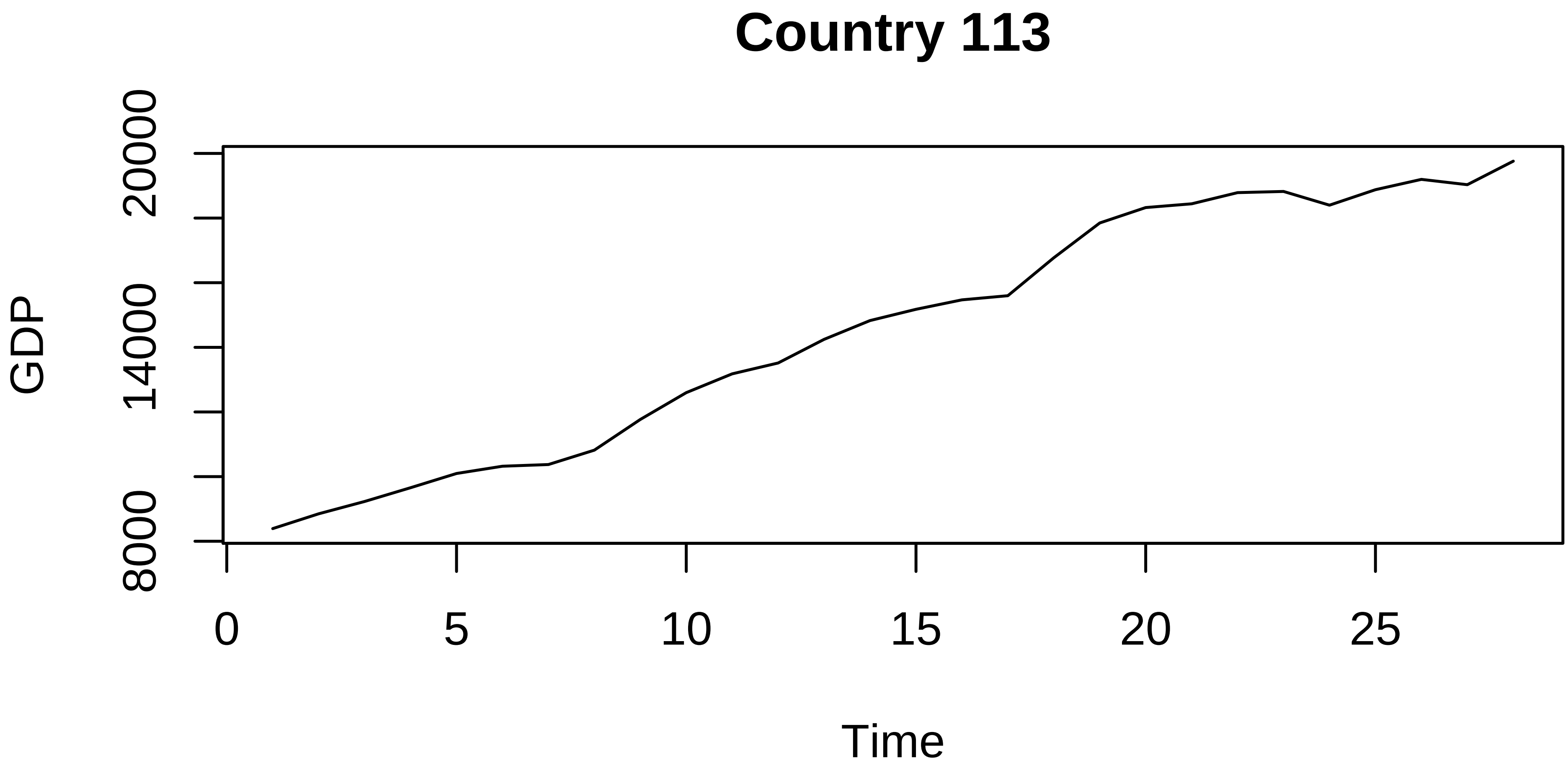
GDP time series for country 2



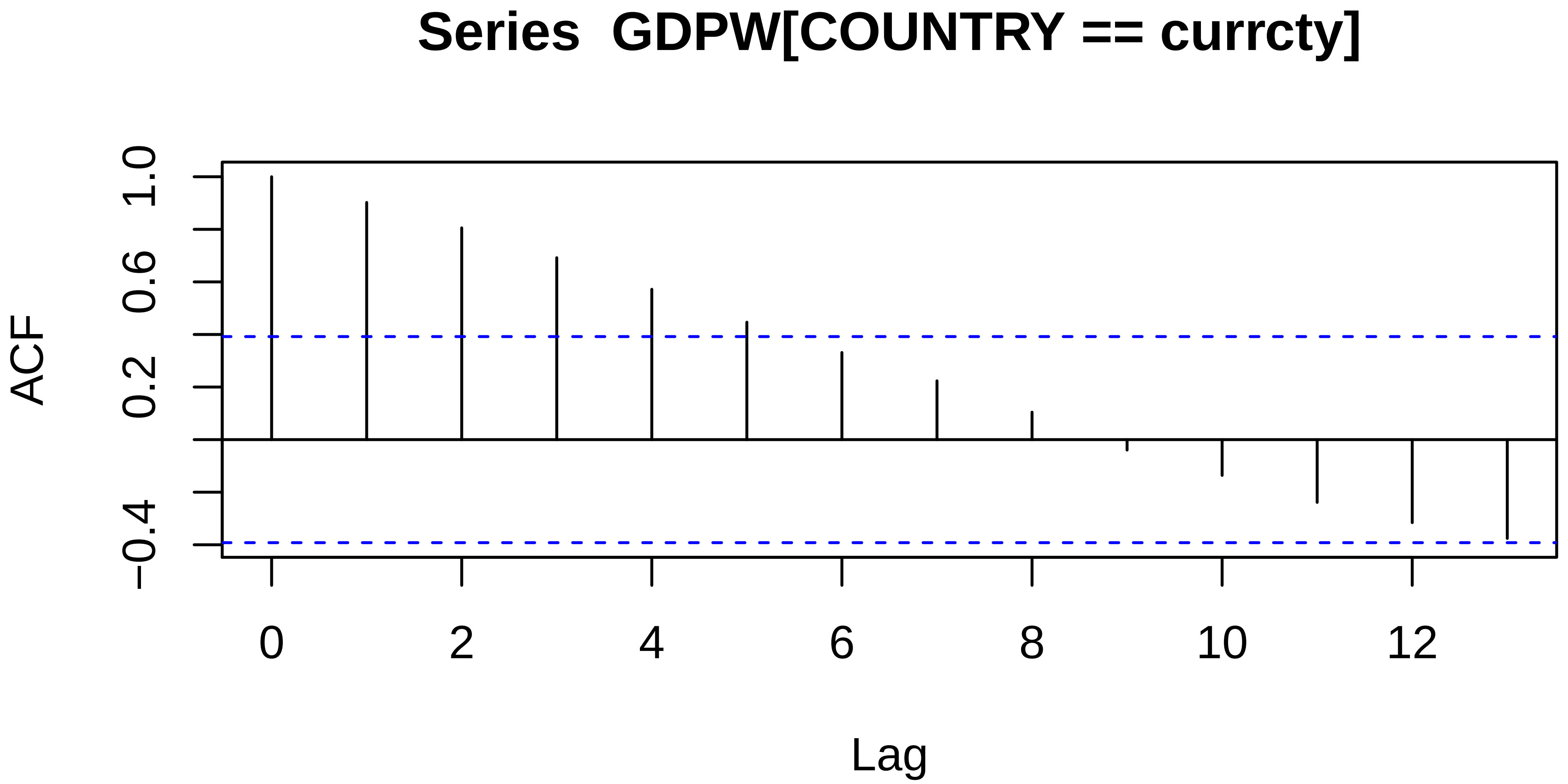
GDP time series for country 3



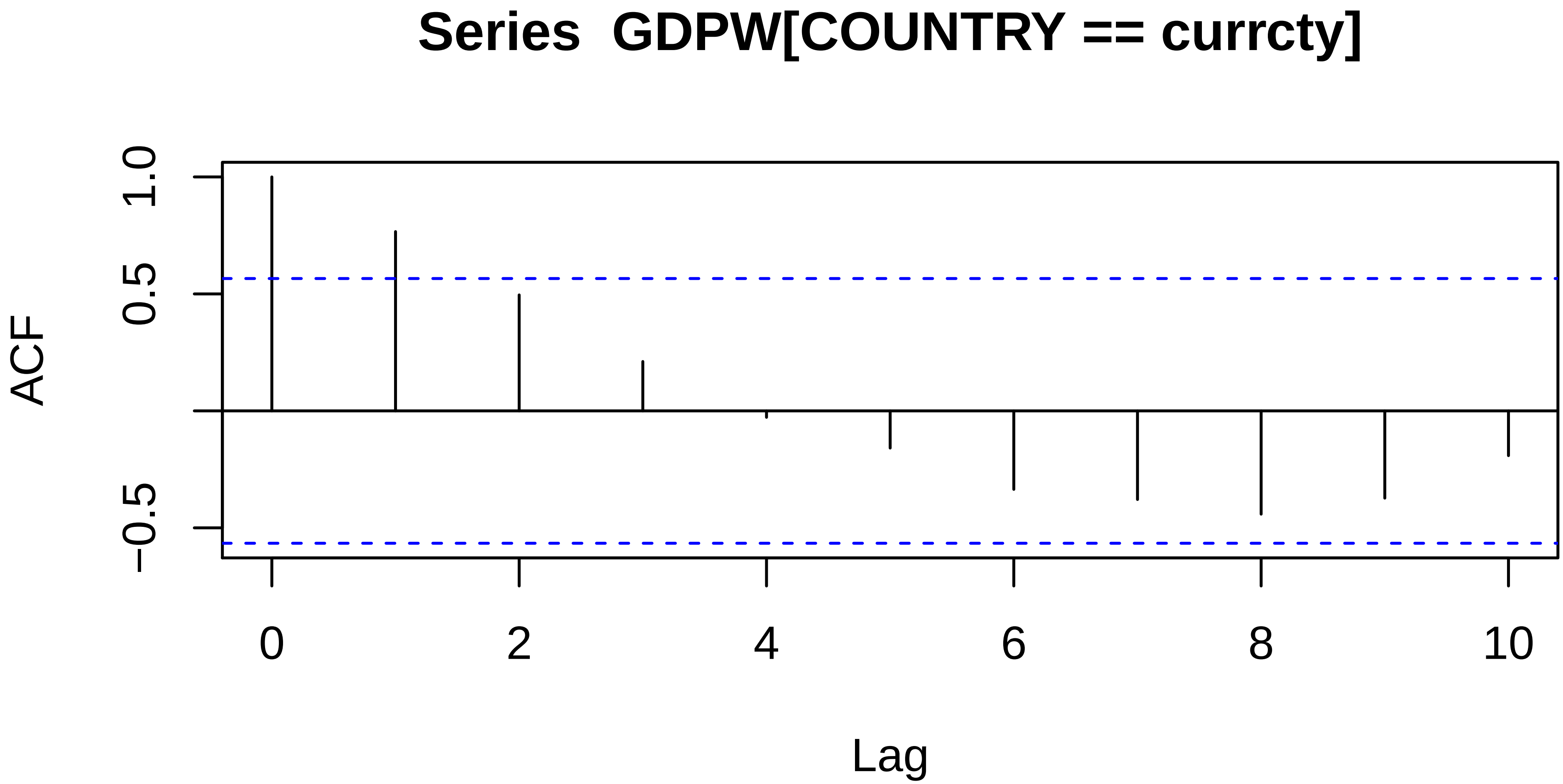
GDP time series for country 4



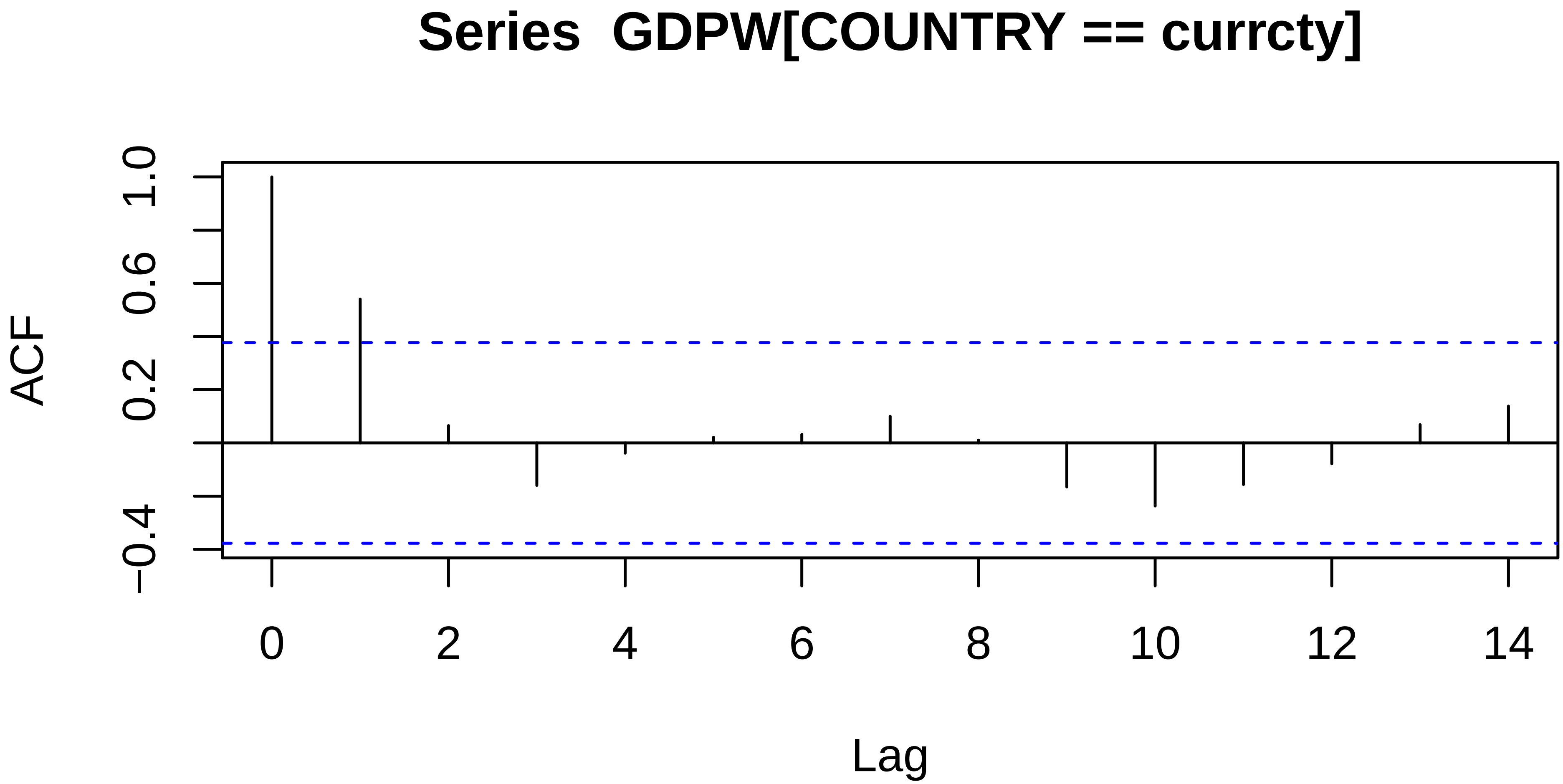
GDP time series for country 113



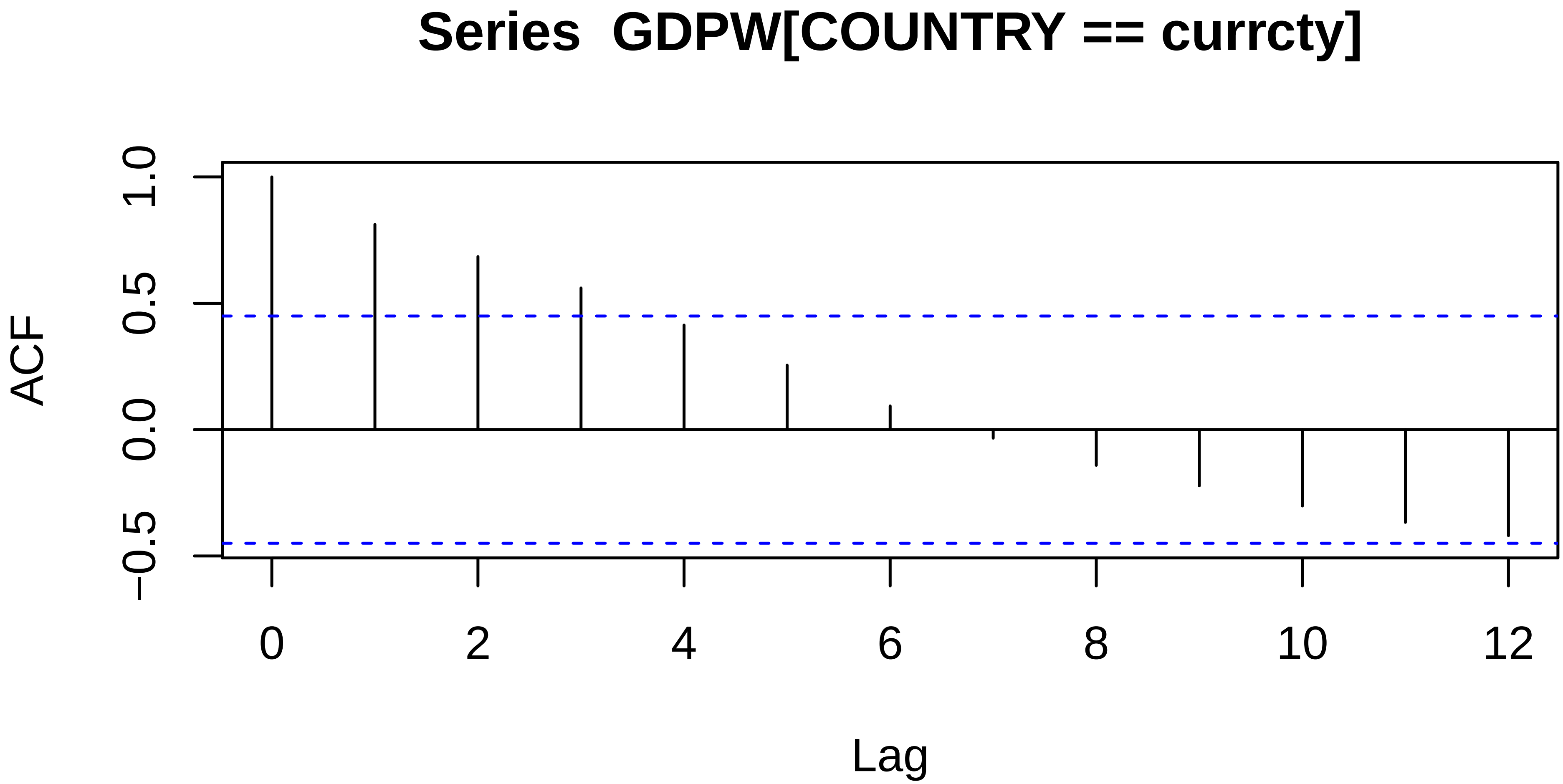
GDP ACF for country 1



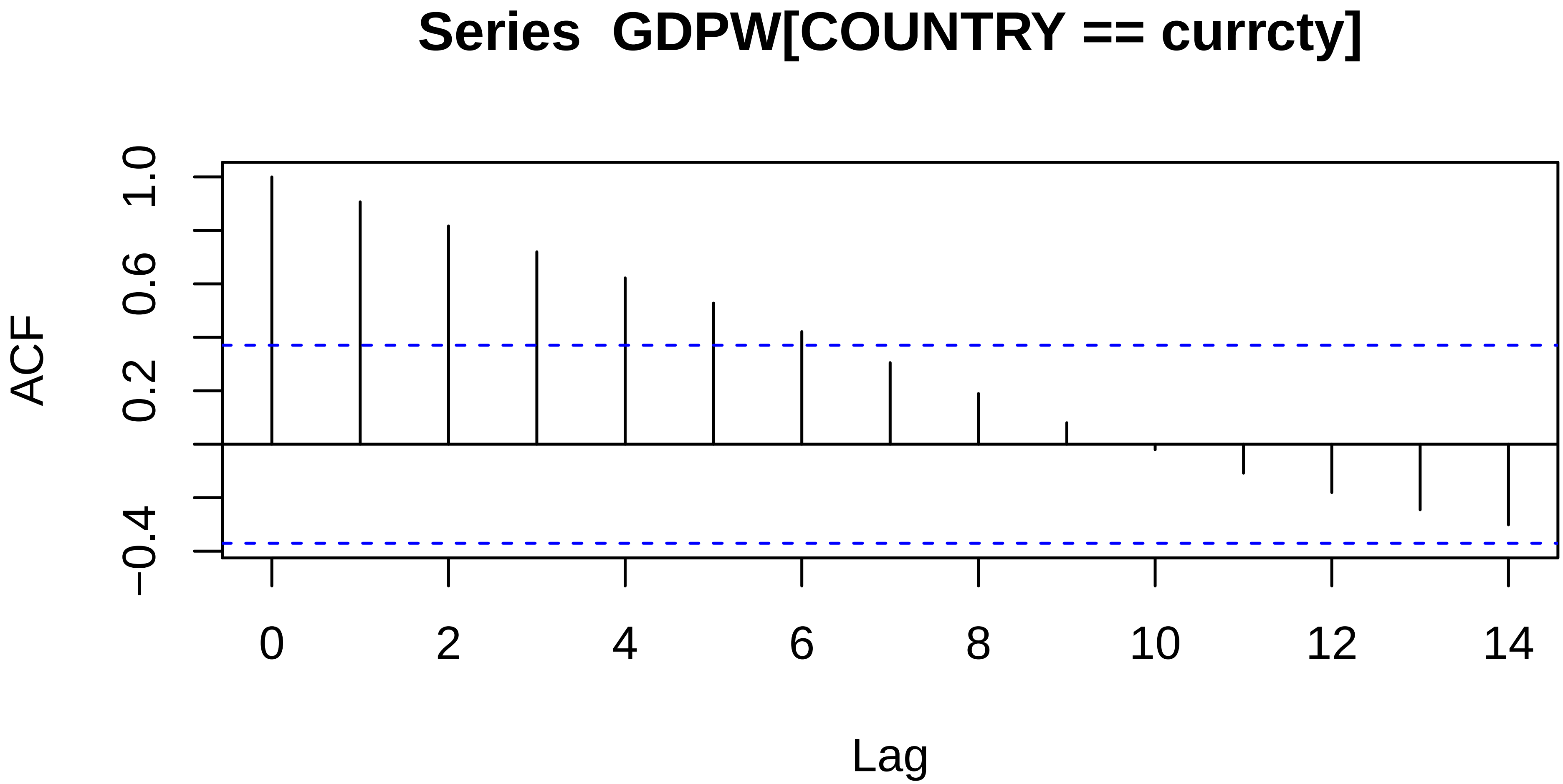
GDP ACF for country 2



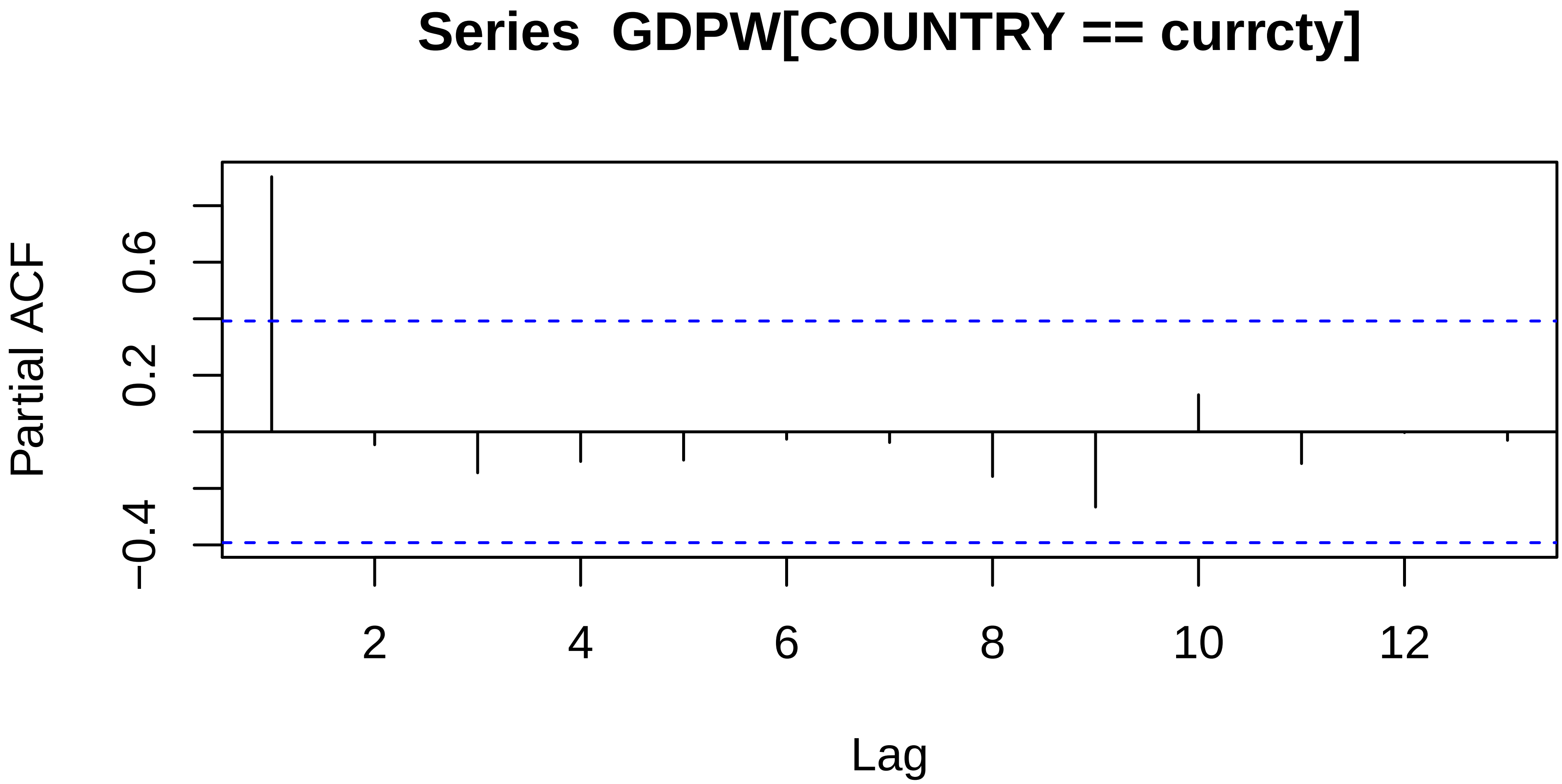
GDP ACF for country 3



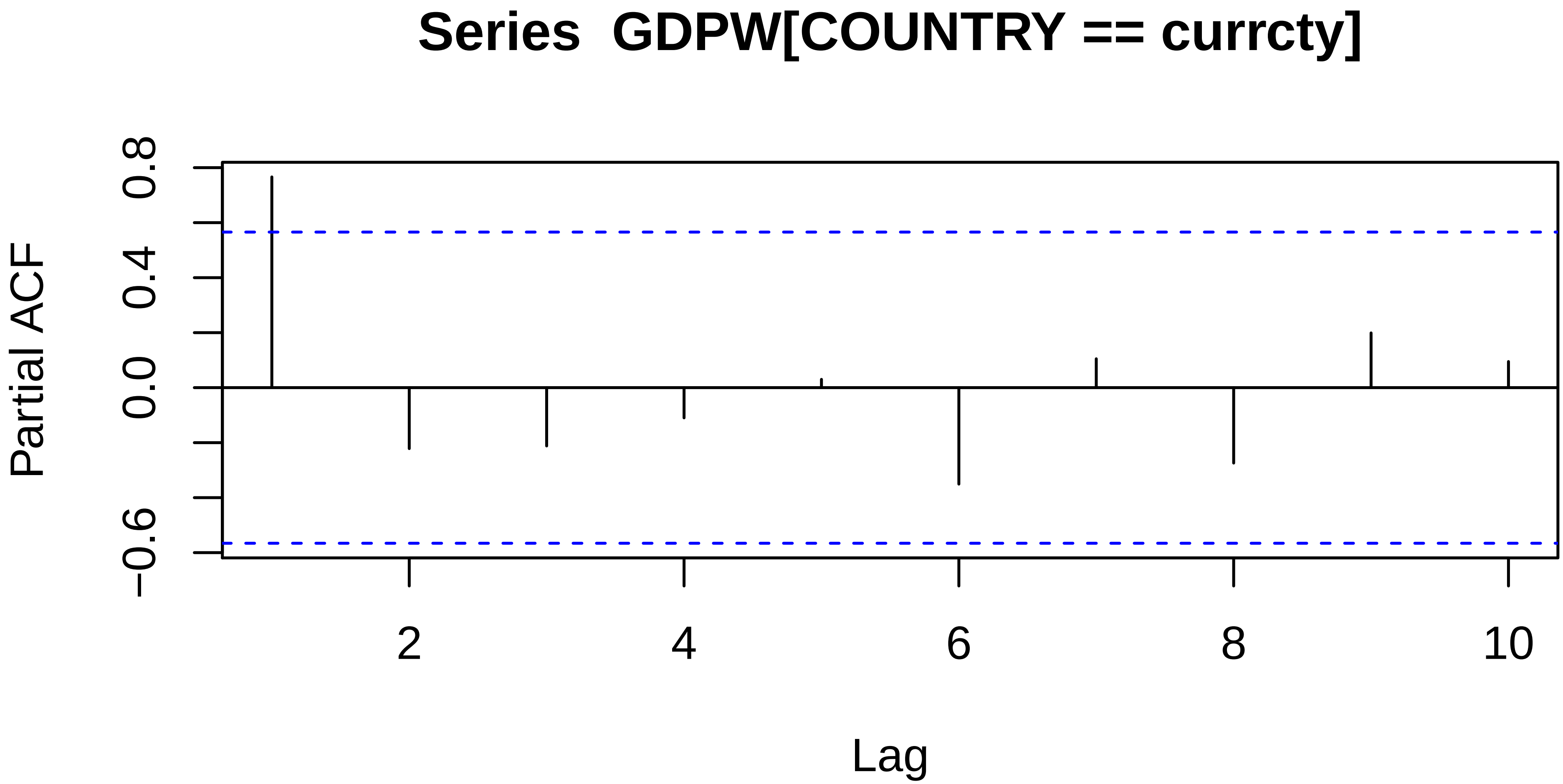
GDP ACF for country 4



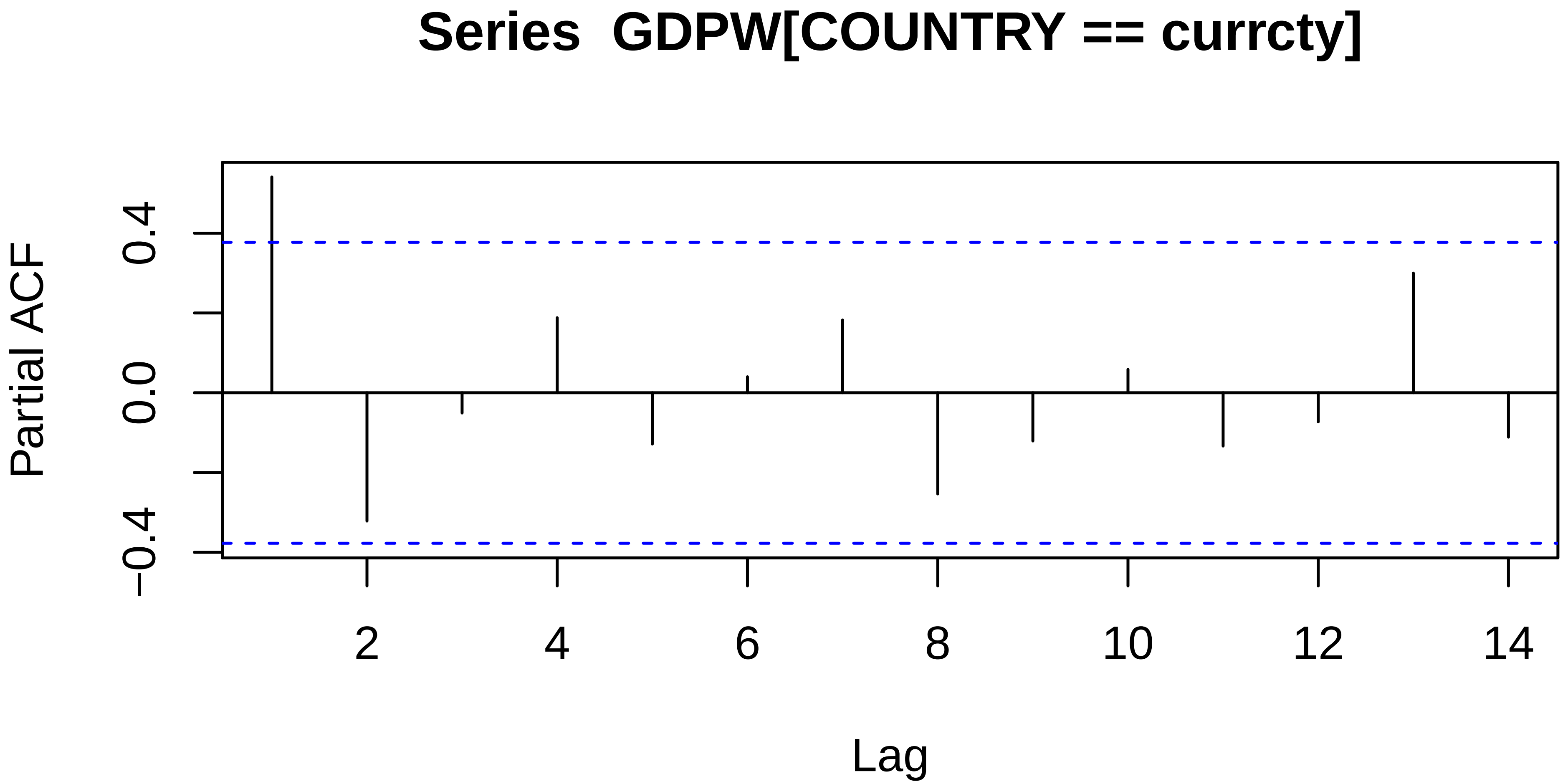
GDP ACF for country 113



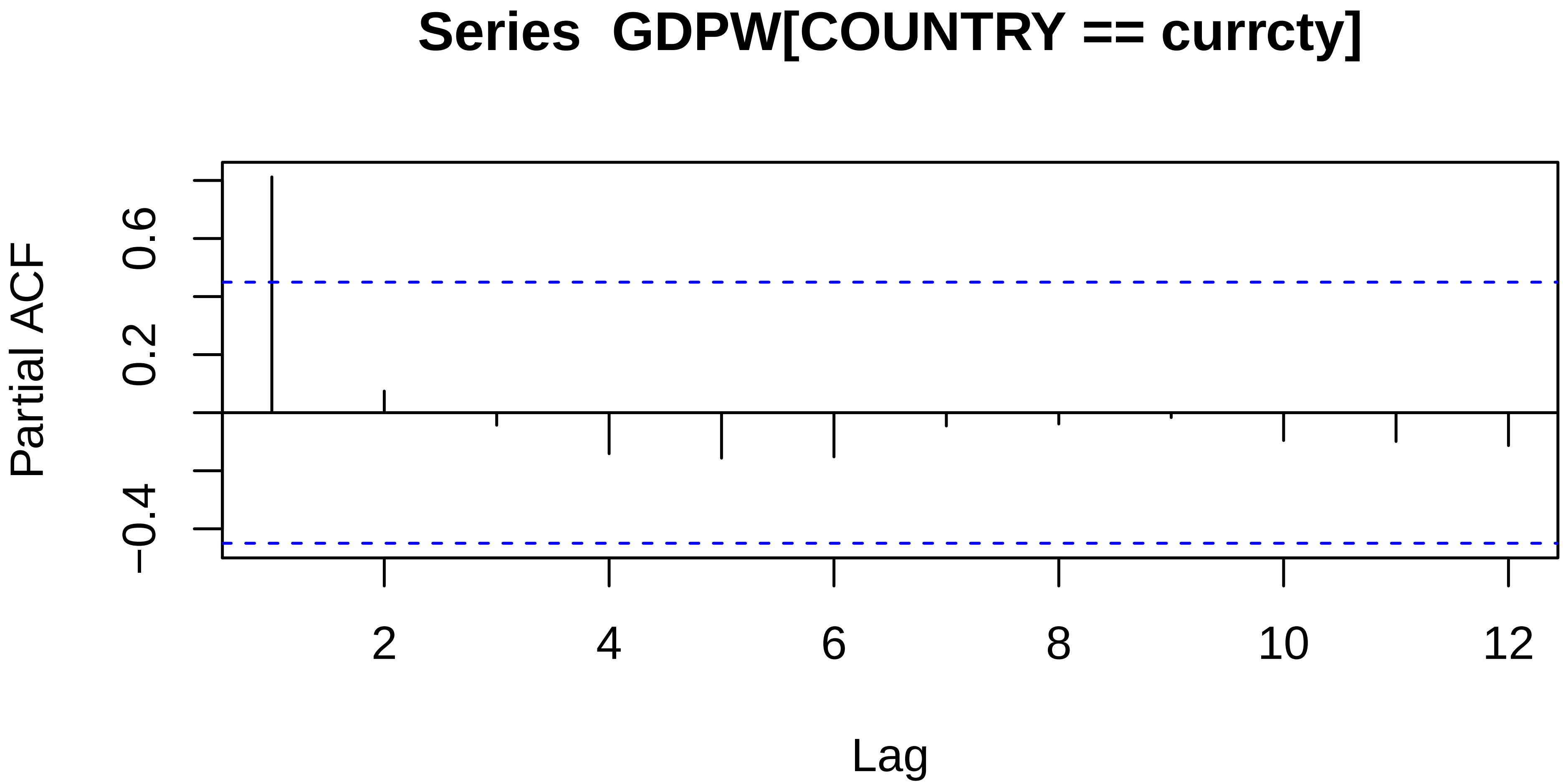
GDP PACF for country 1



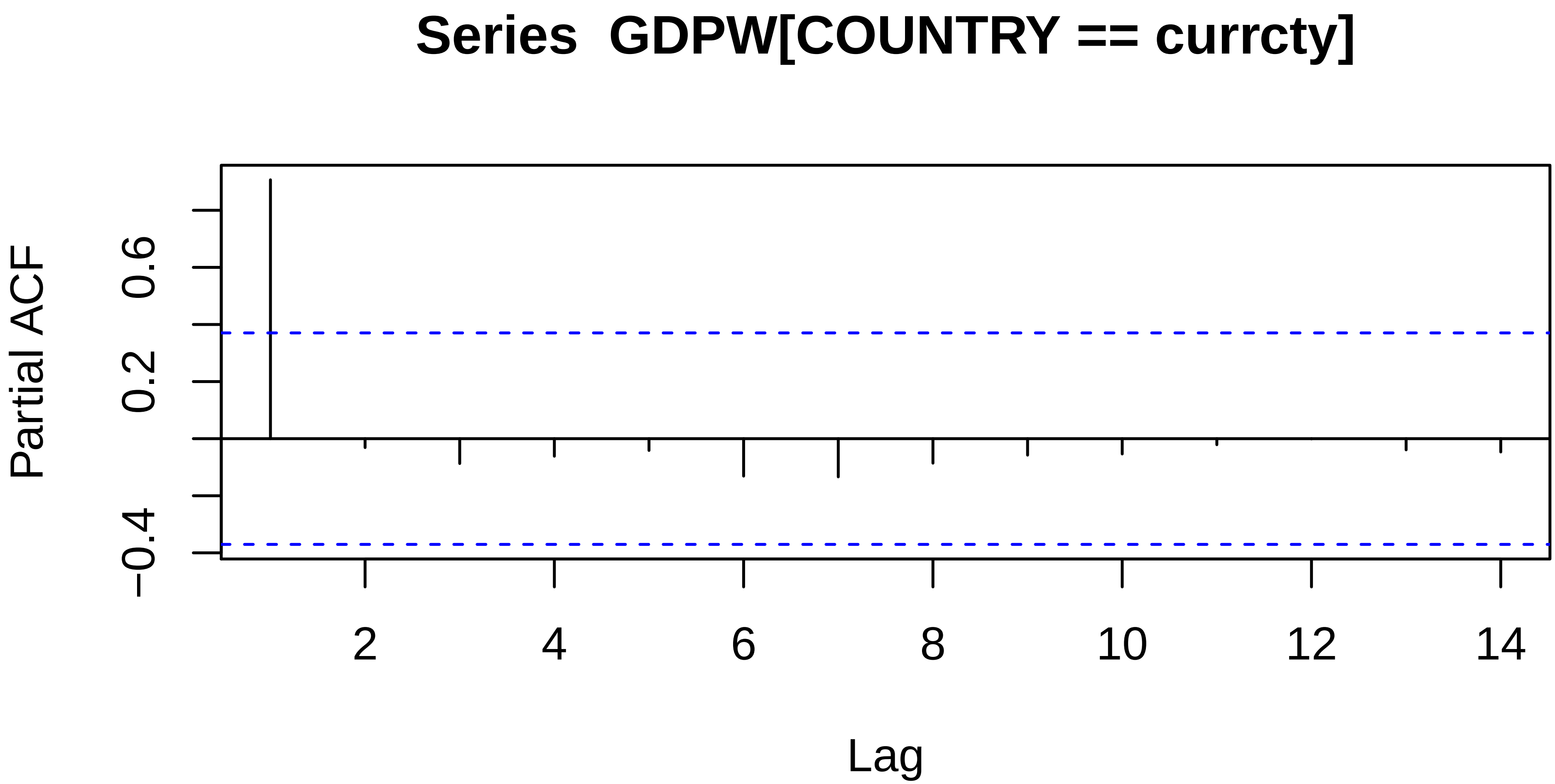
GDP PACF for country 2



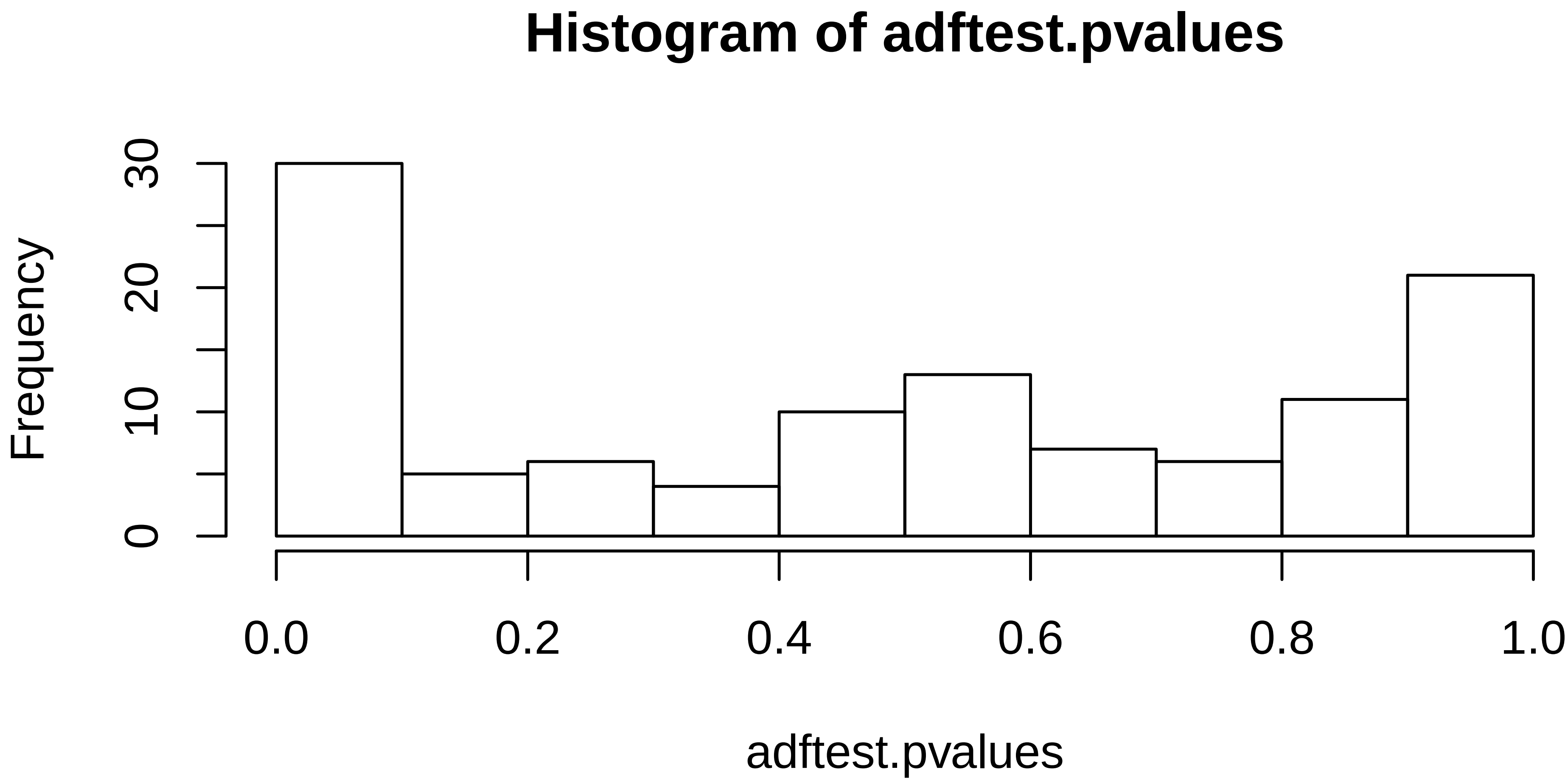
GDP PACF for country 3



GDP PACF for country 4

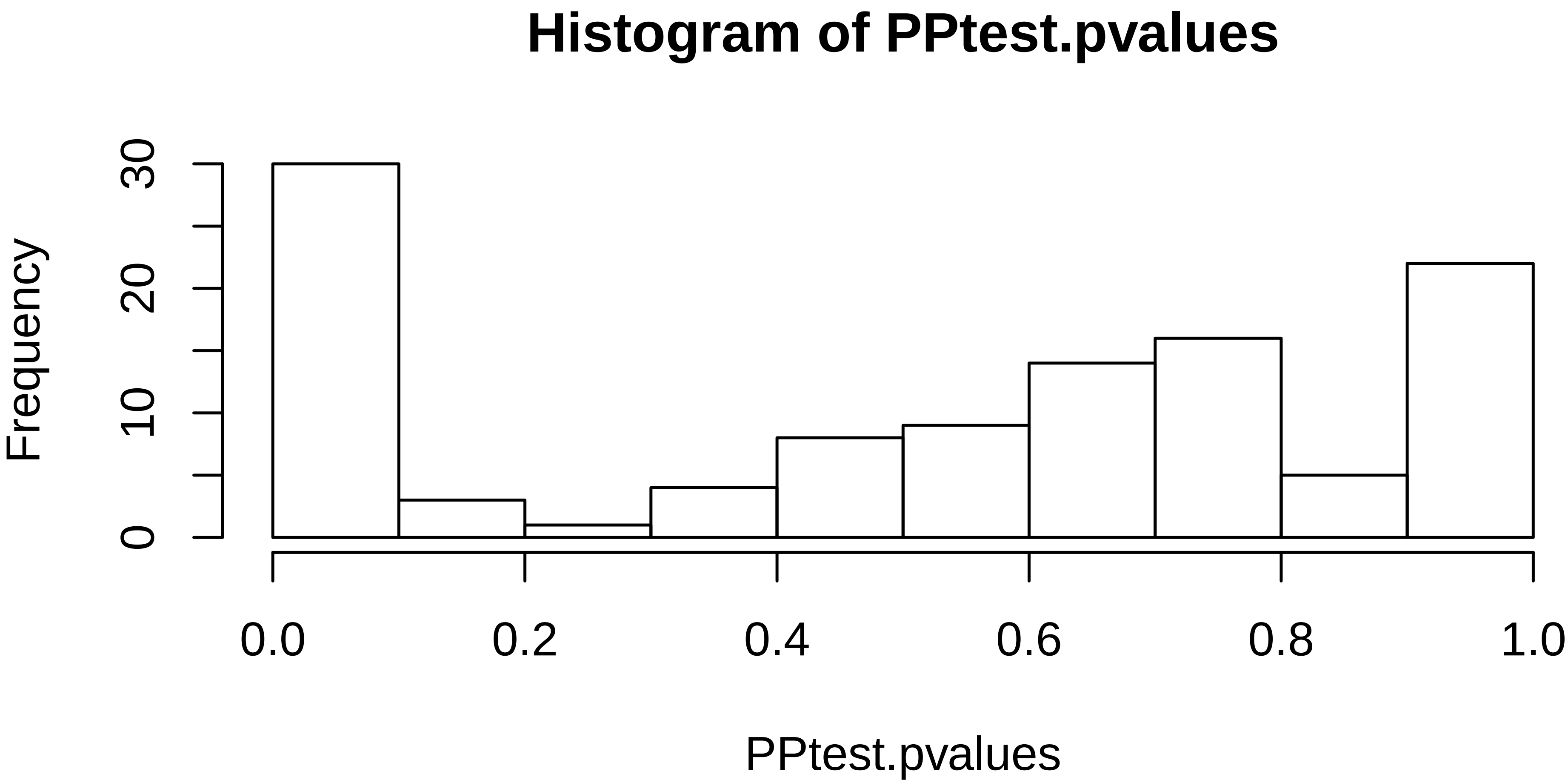


GDP PACF for country 113



Histogram of p-values from ADF tests on GDPW

What would we see if there were no unit roots?



Histogram of p-values from Phillips-Peron tests on GDPW

Choosing AR(p,q) for panel

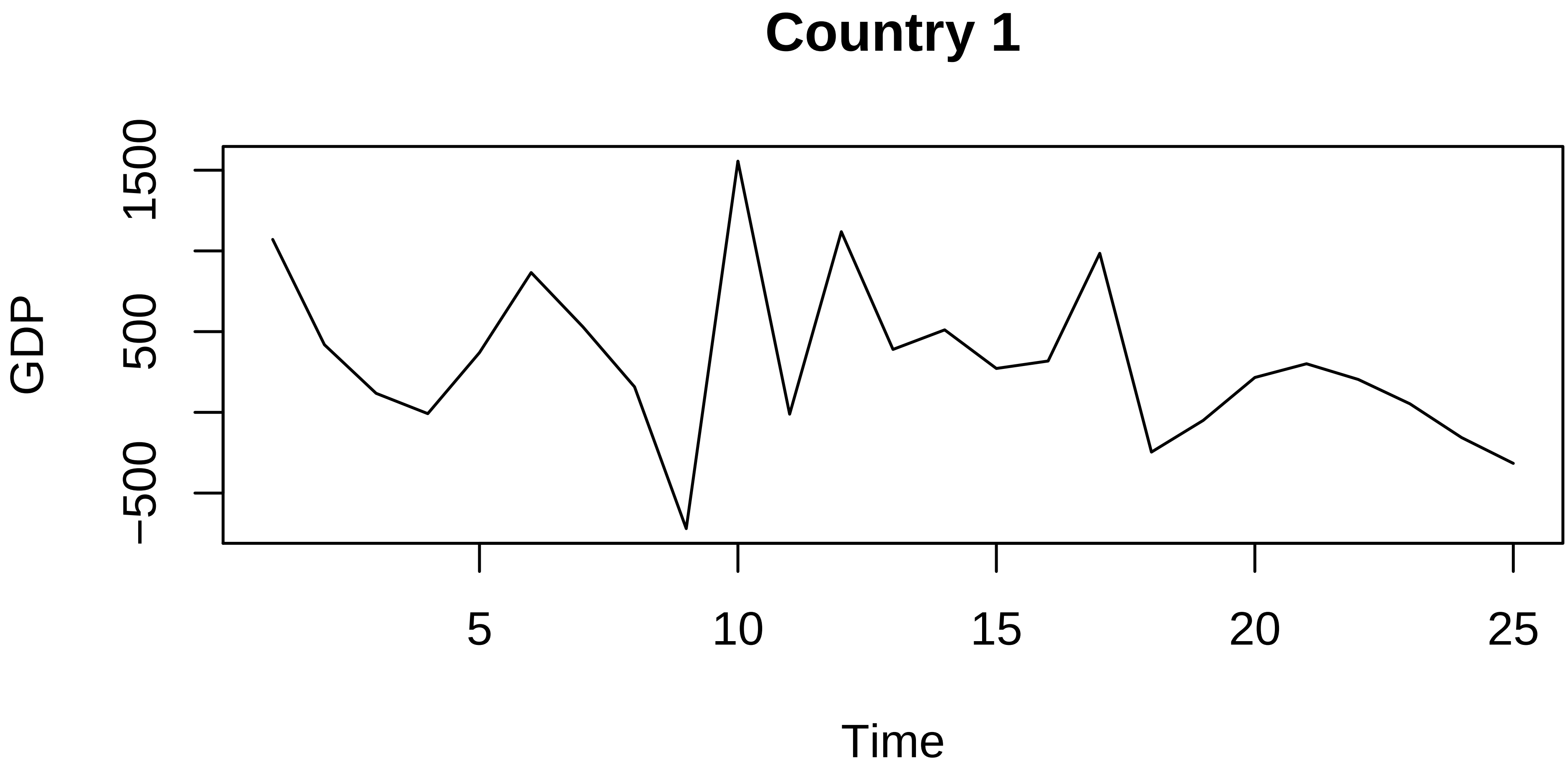
What do we think?

Clearly some heterogeneity

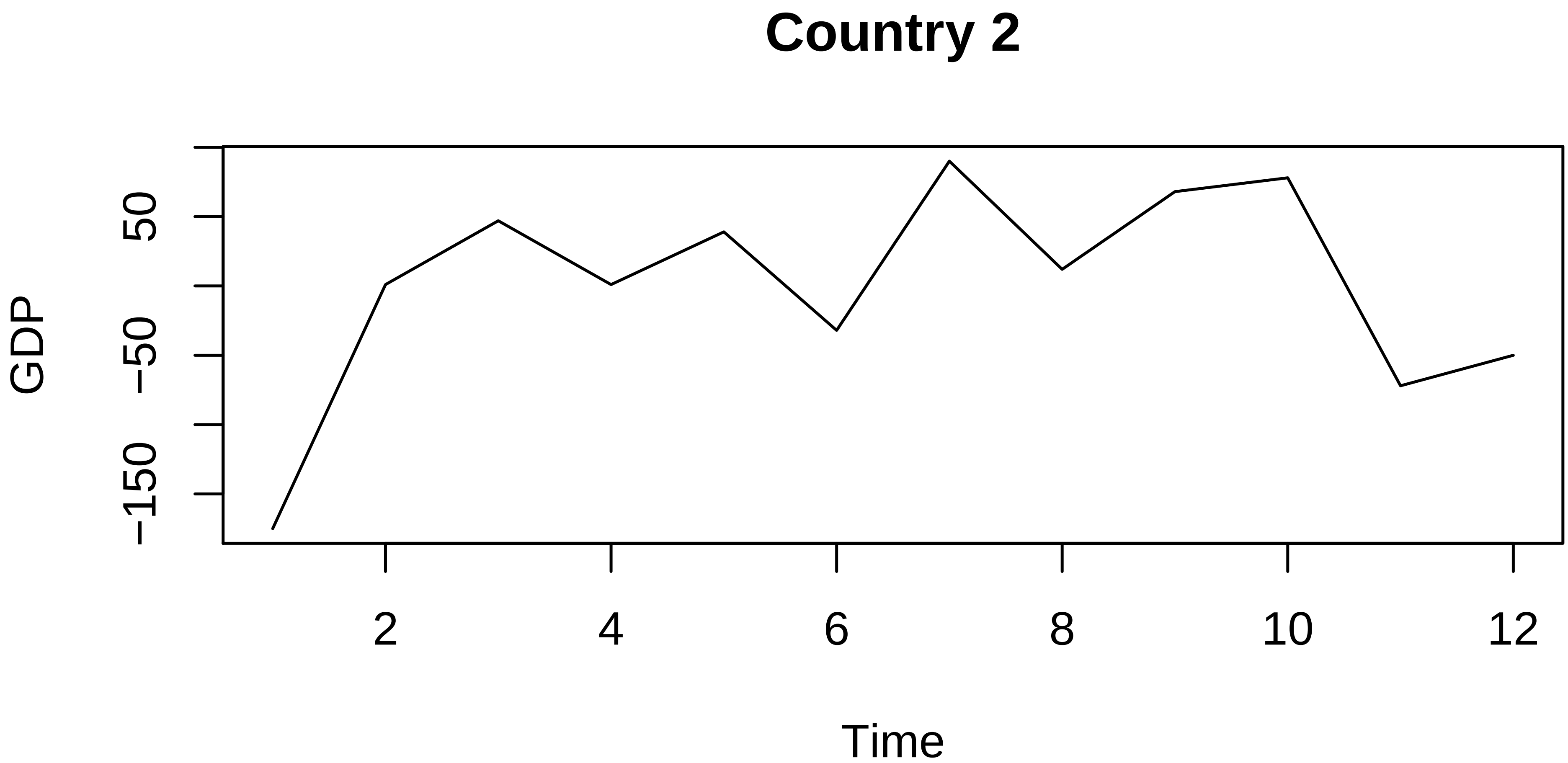
If had to pick one time series specification, choose ARIMA(0,1,0) or ARIMA(1,1,0)

Seems to fit many cases; guards against spurious regression

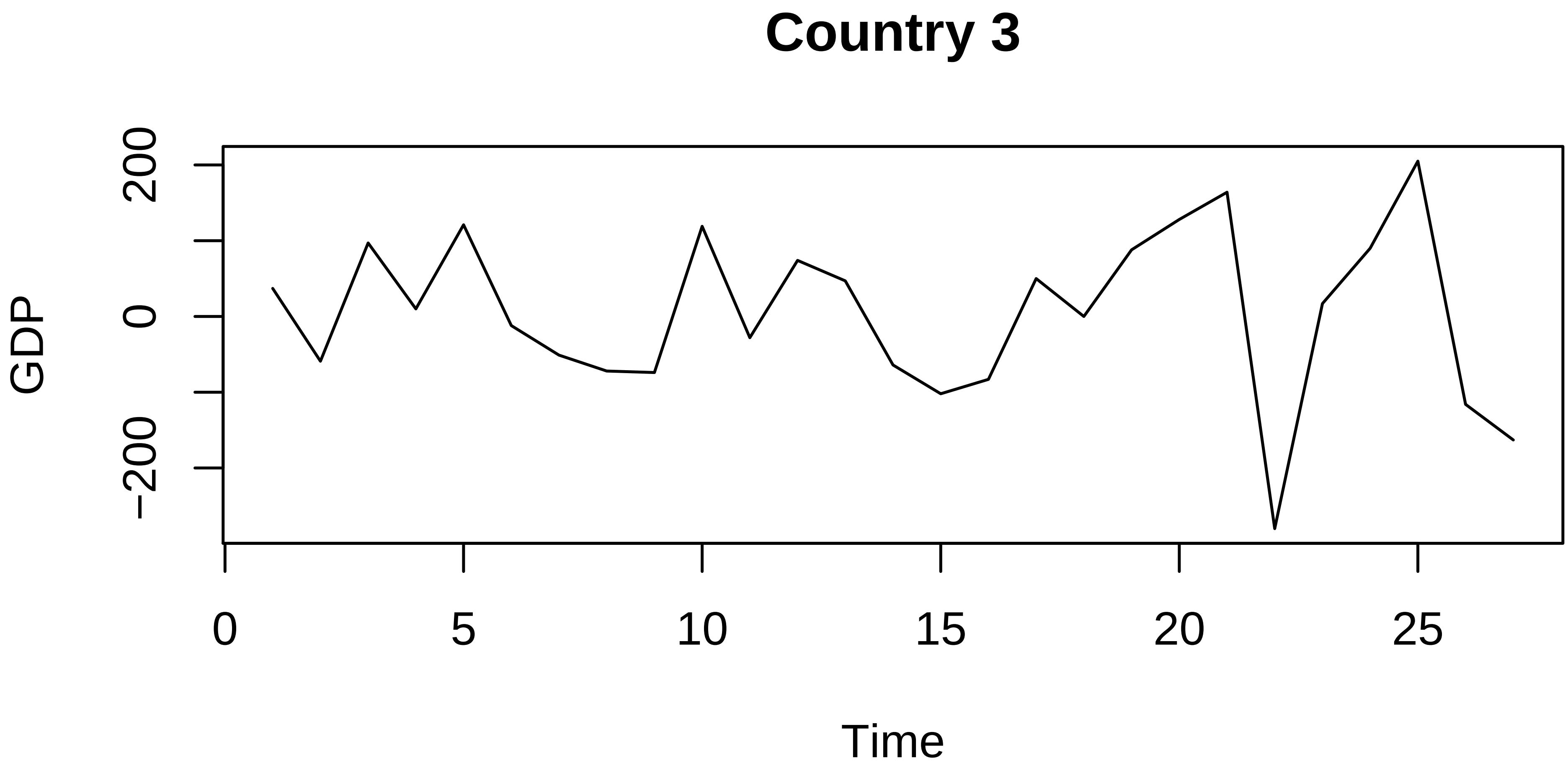
But if we're dubious about imposing a single ARIMA(p,d,q) across units,
we could let them be heterogeneous



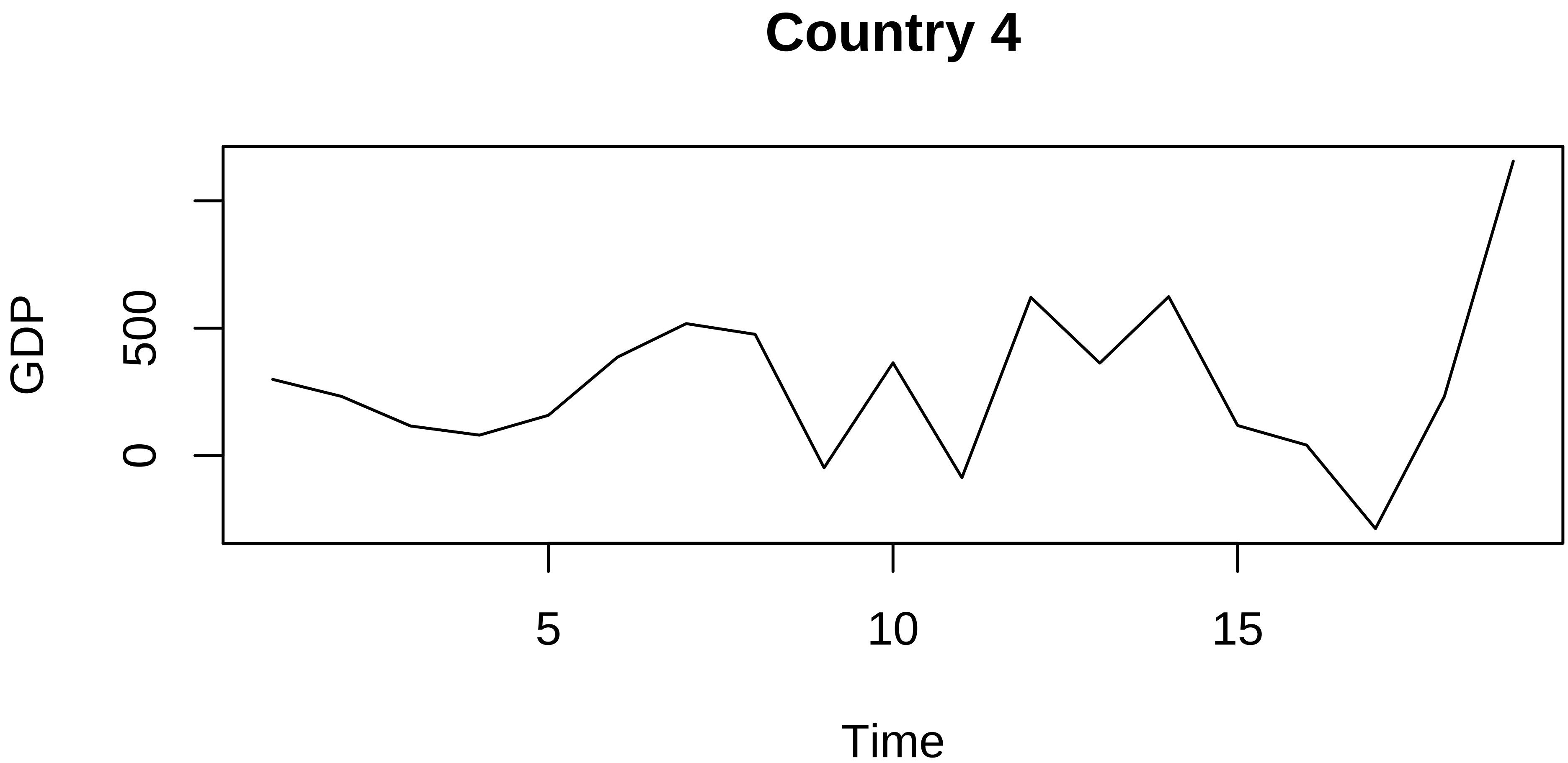
GDPdiff time series for country 1



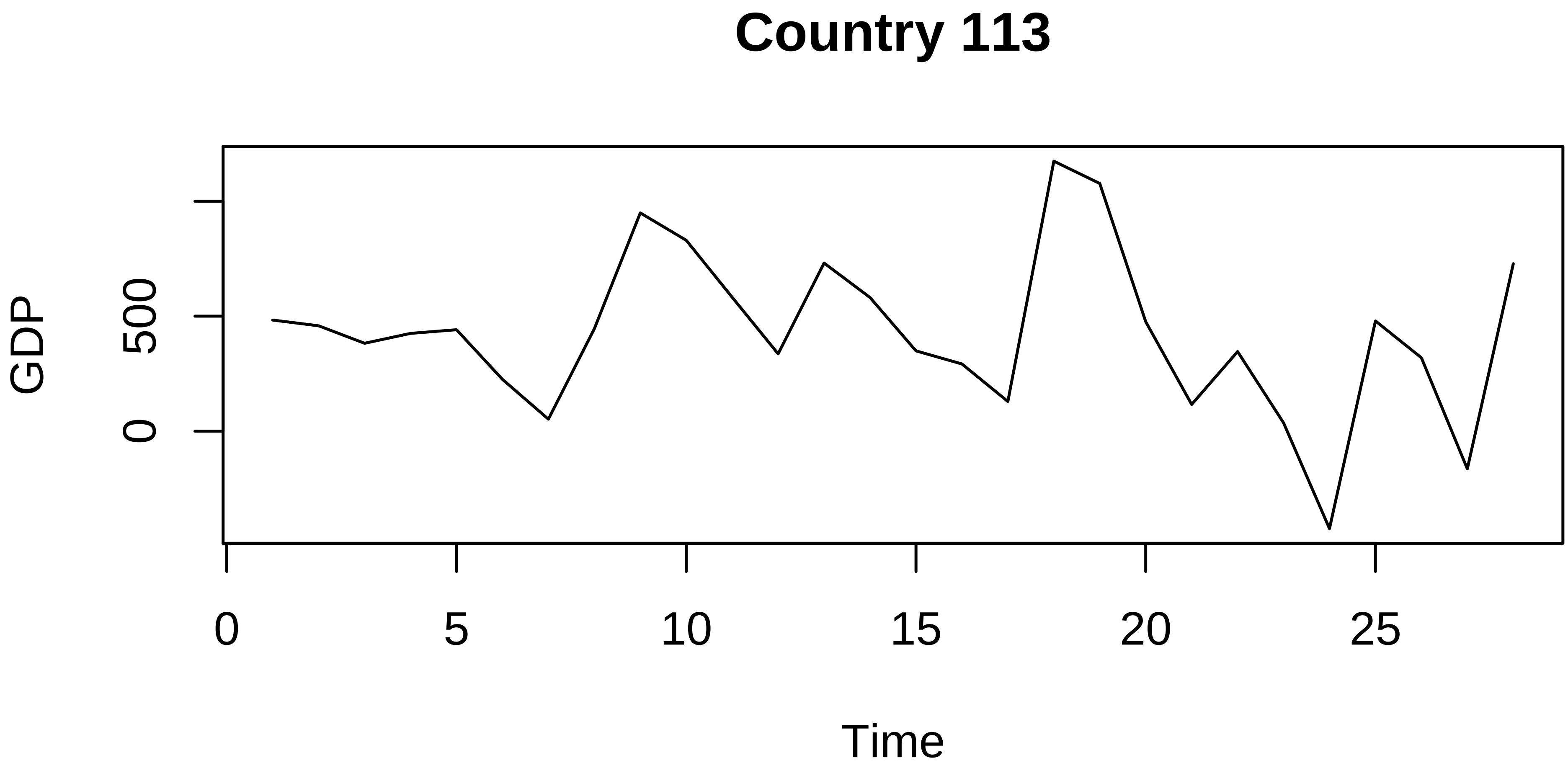
GDPdiff time series for country 2



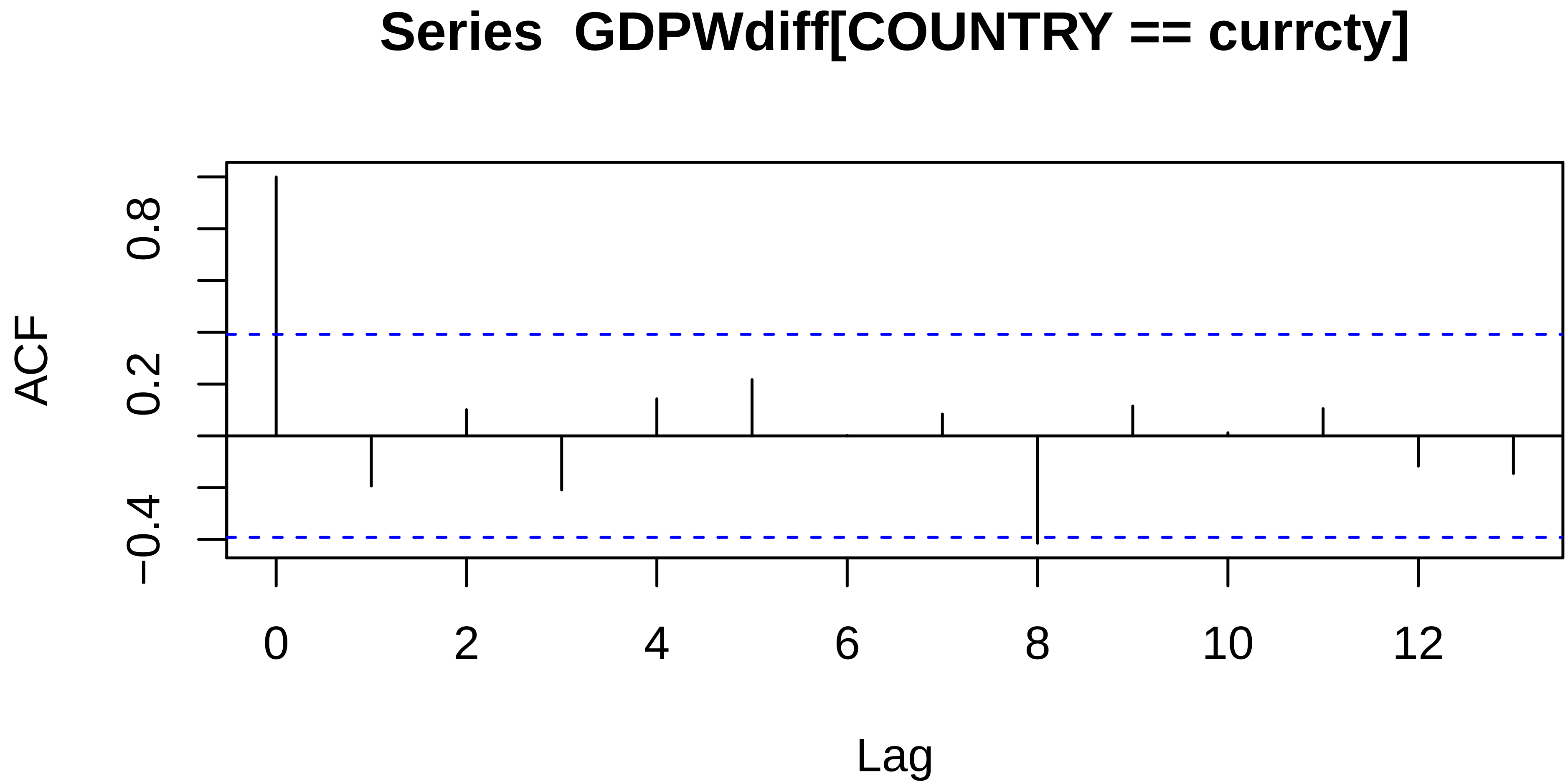
GDPdiff time series for country 3



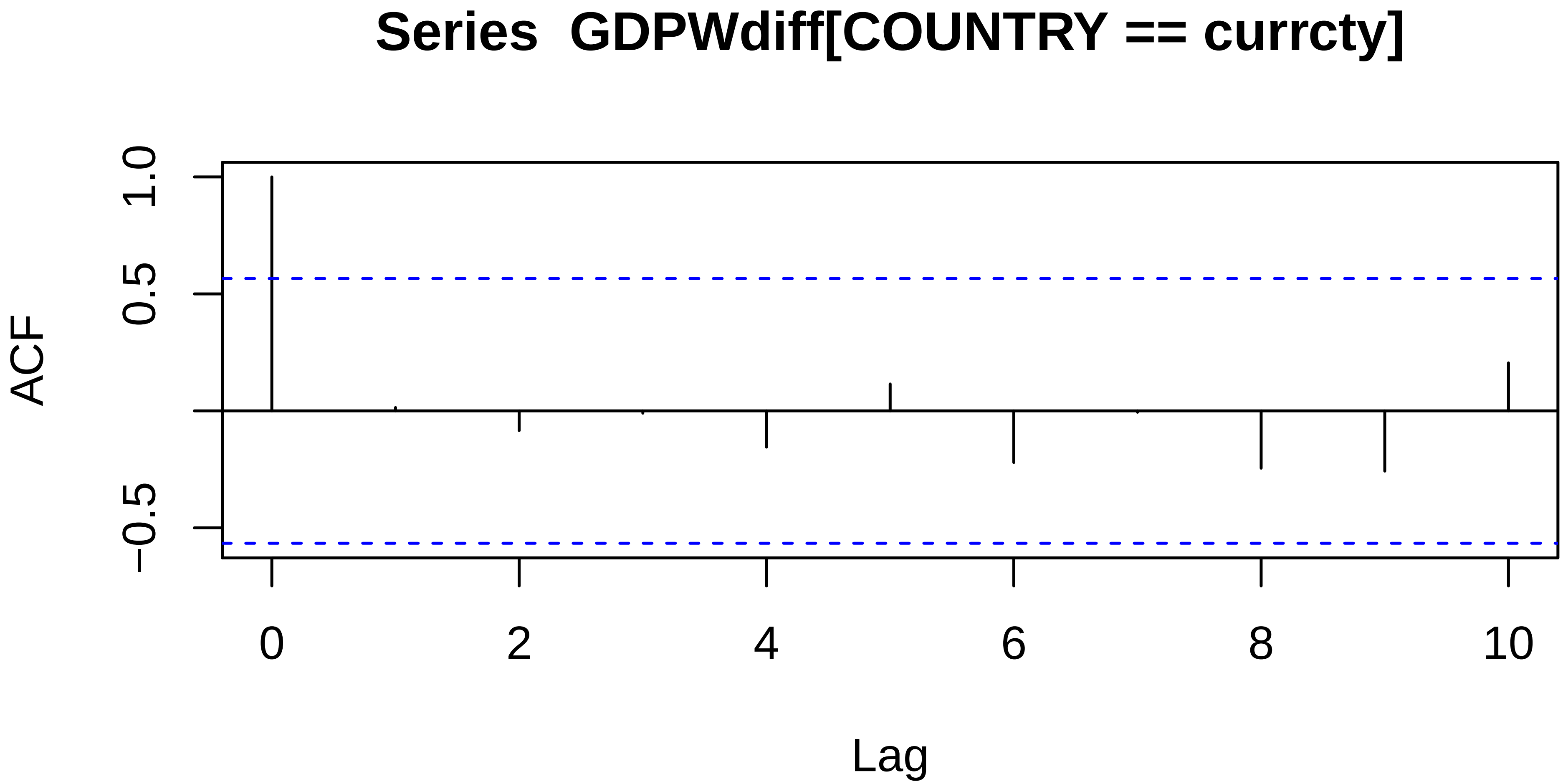
GDPdiff time series for country 4



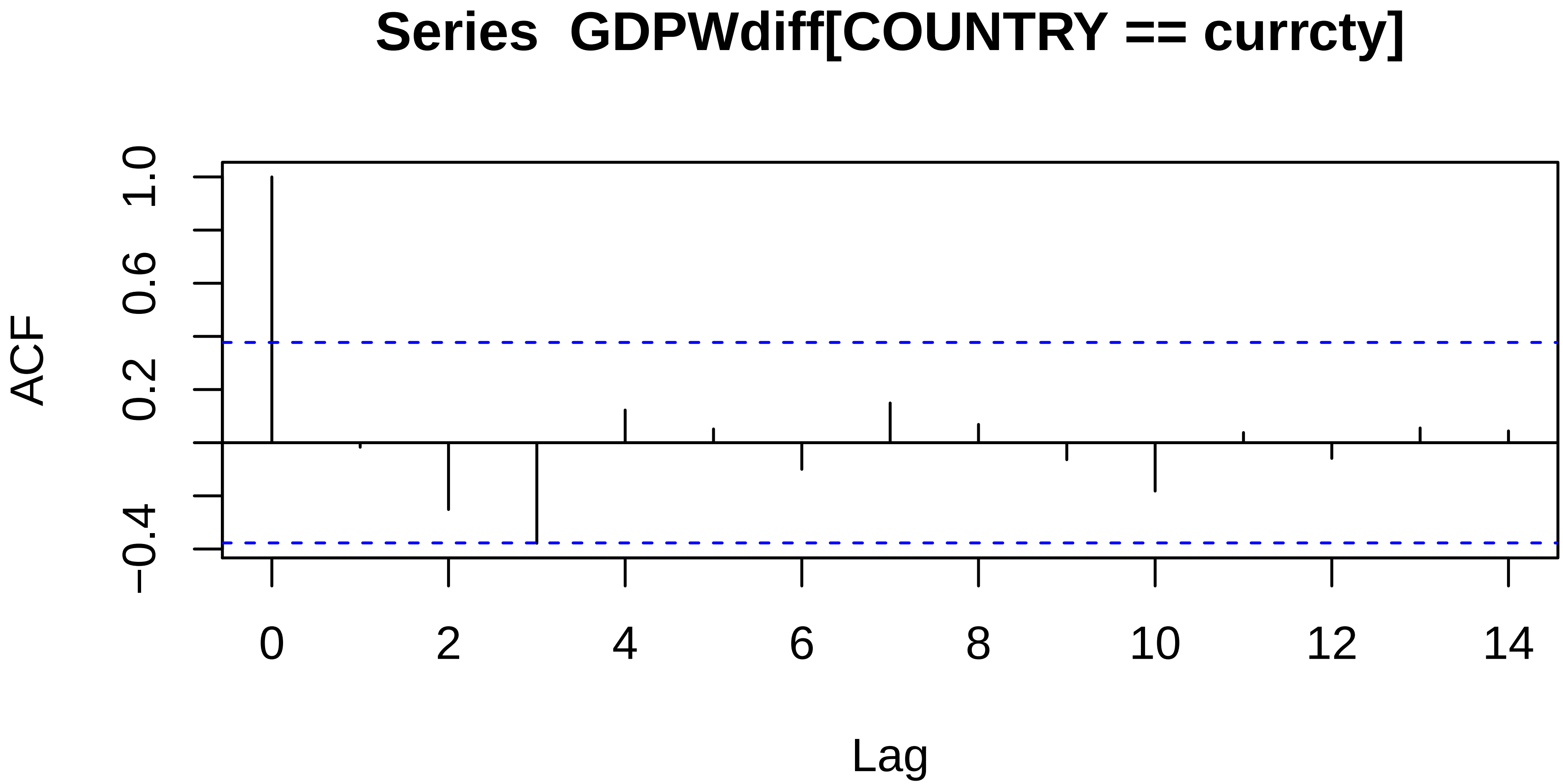
GDP diff time series for country 113



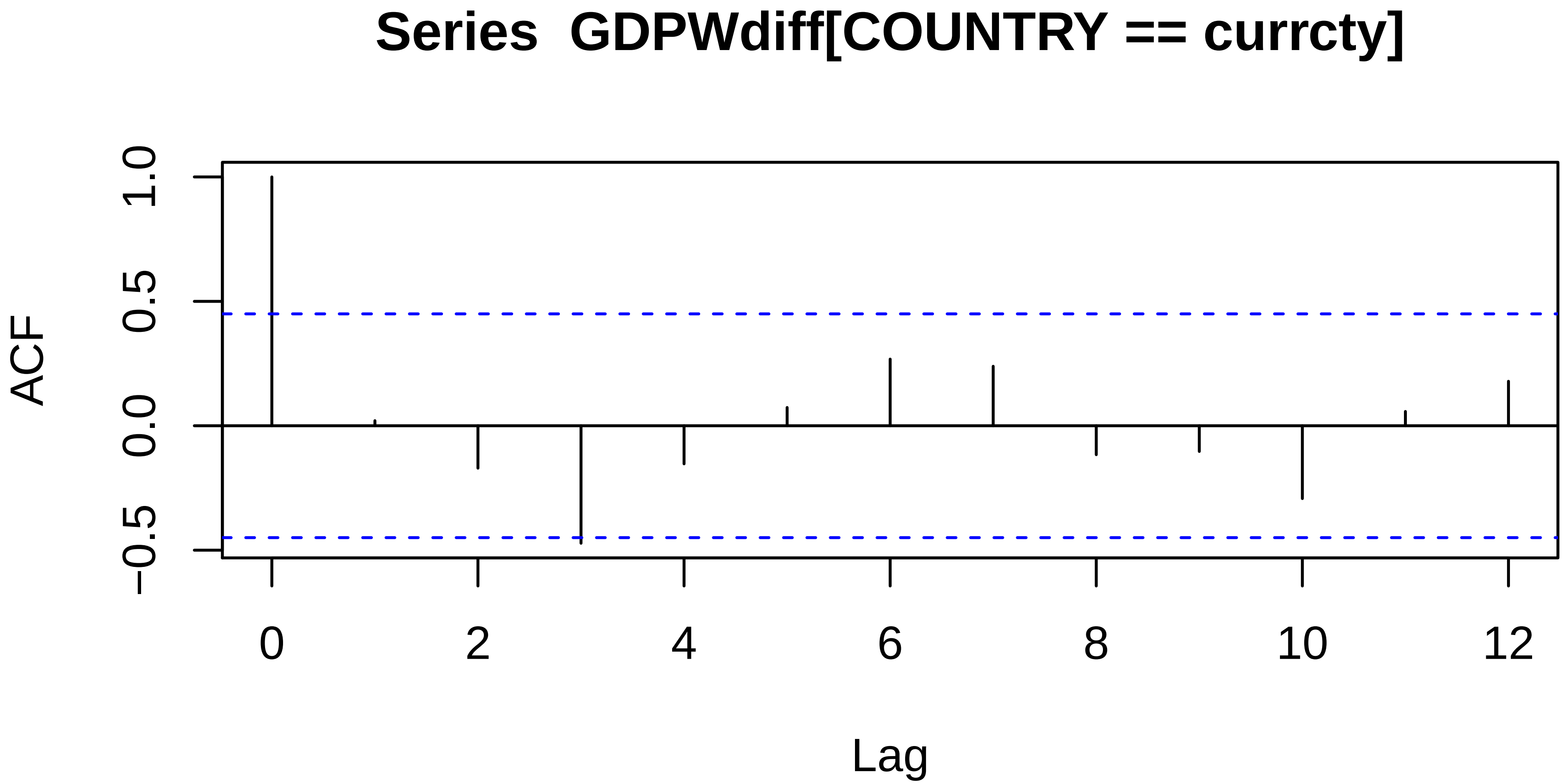
GDPdiff ACF for country 1



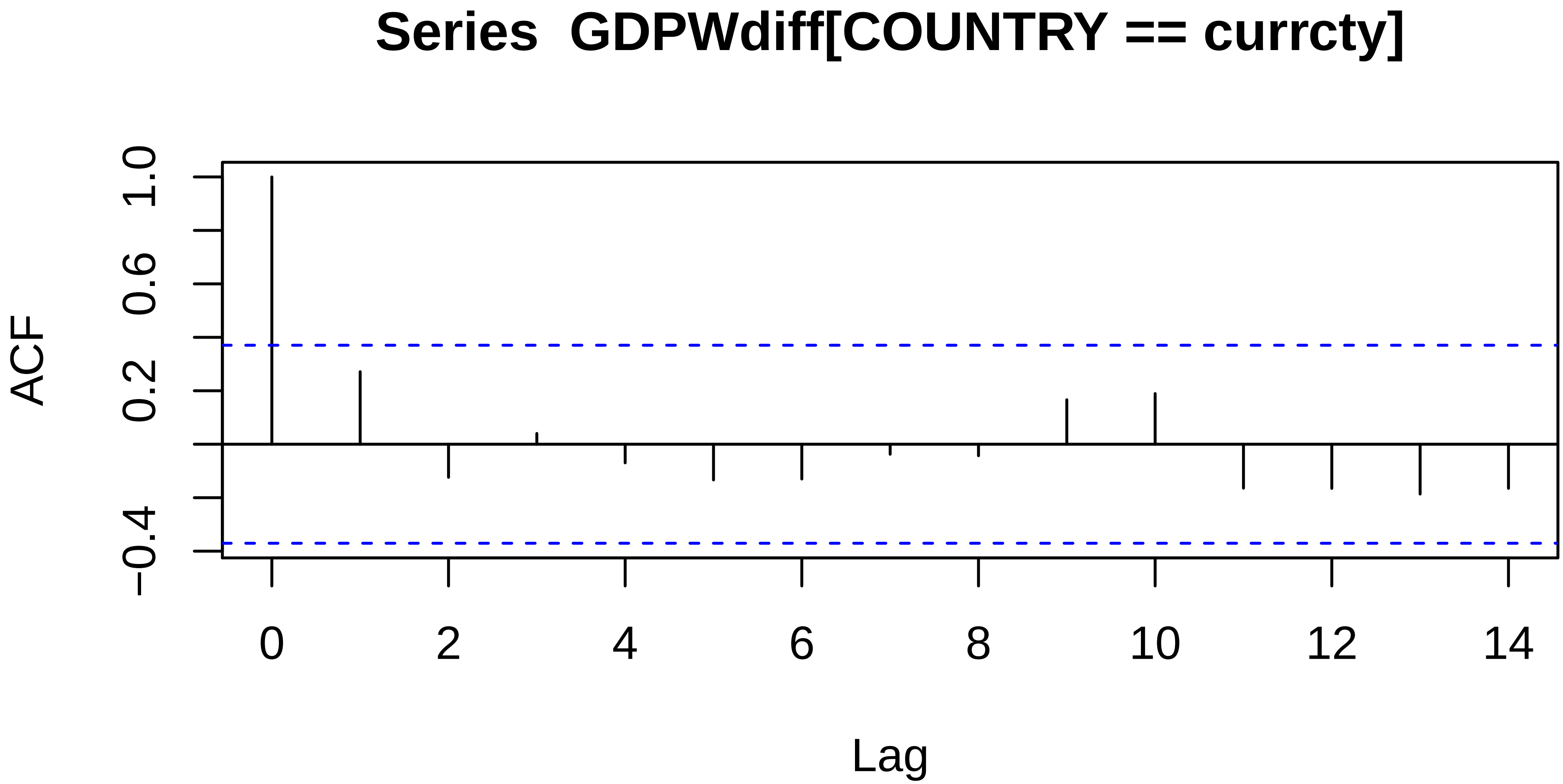
GDPdiff ACF for country 2



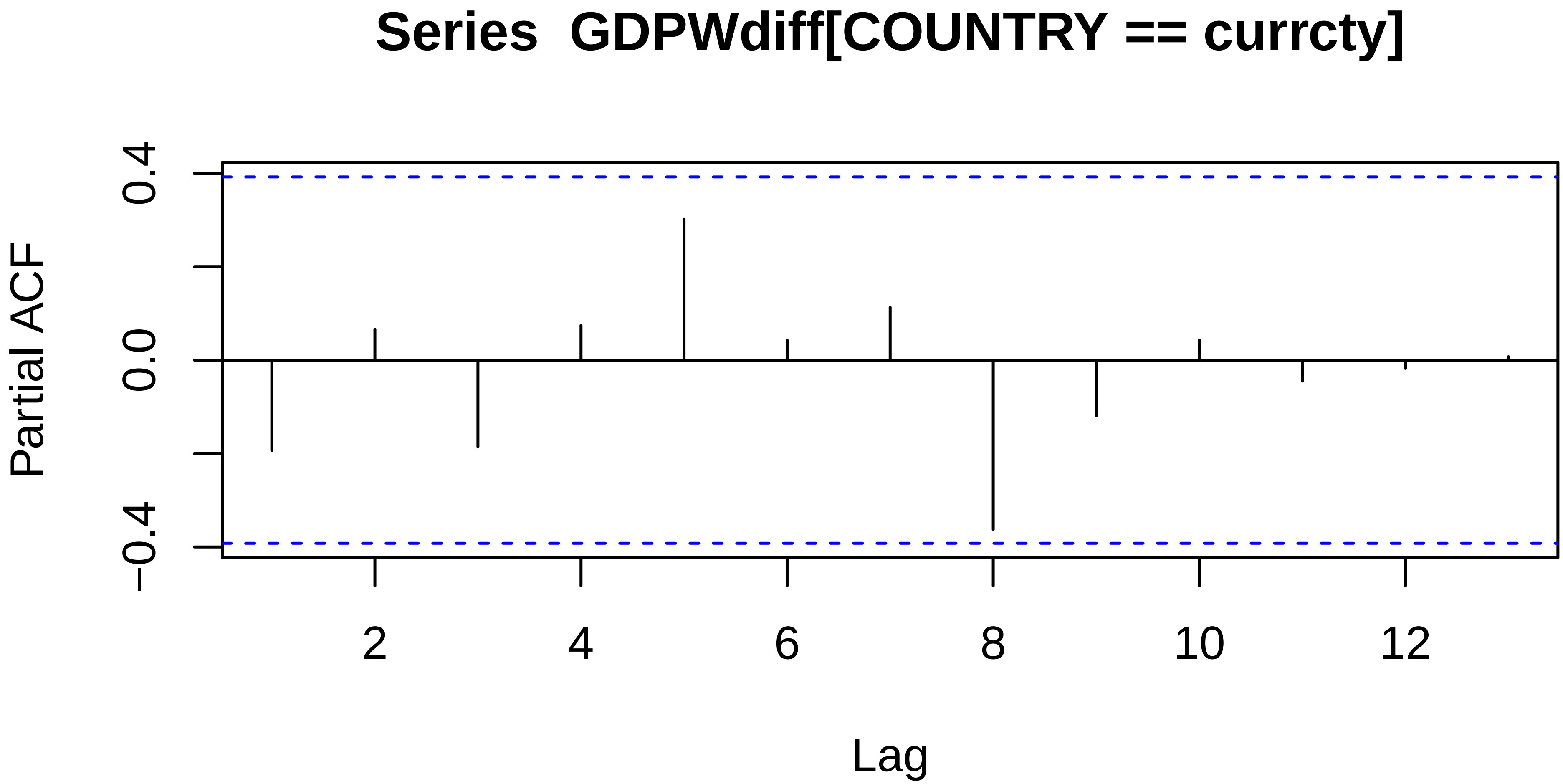
GDPdiff ACF for country 3



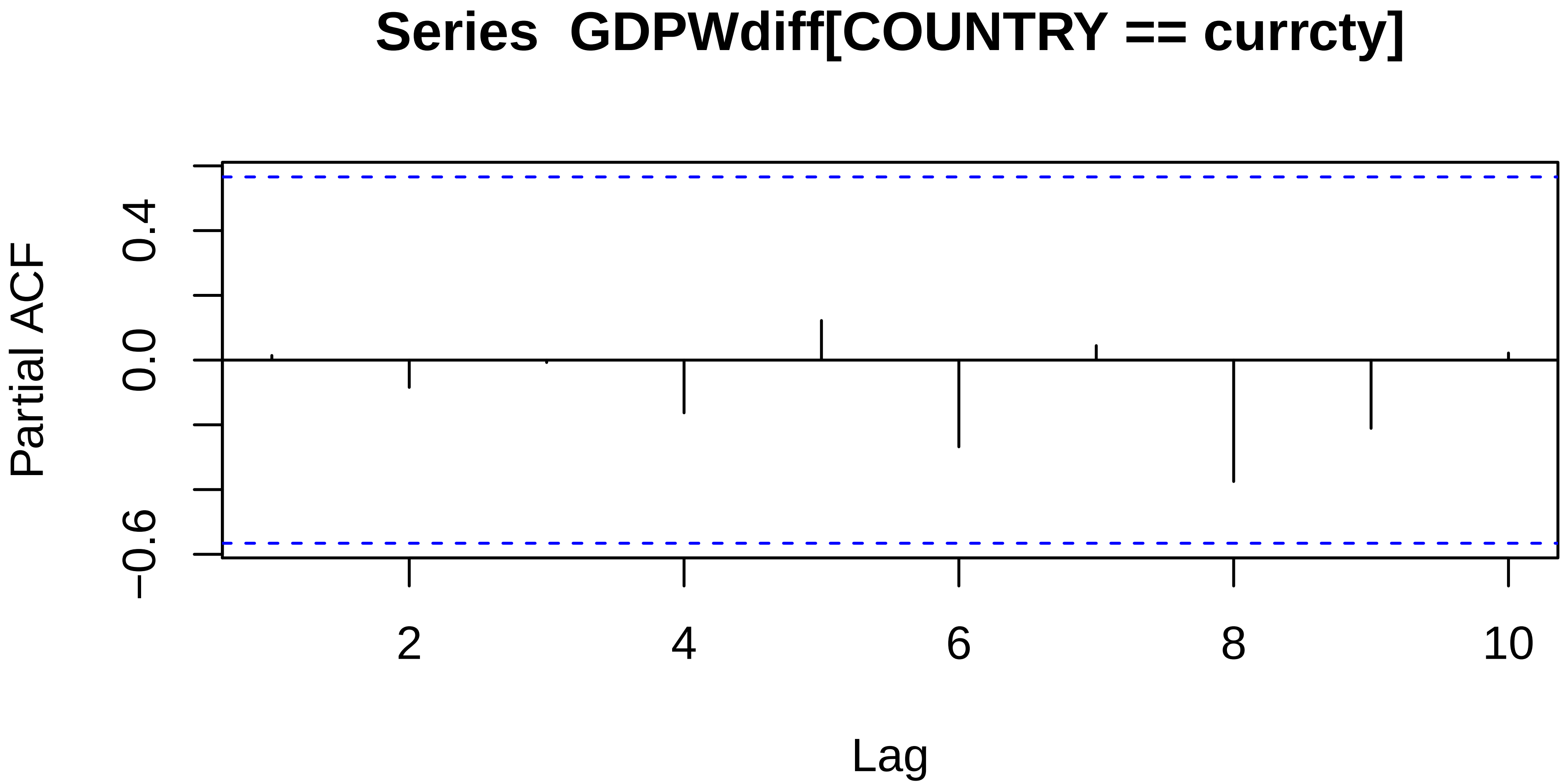
GDPdiff ACF for country 4



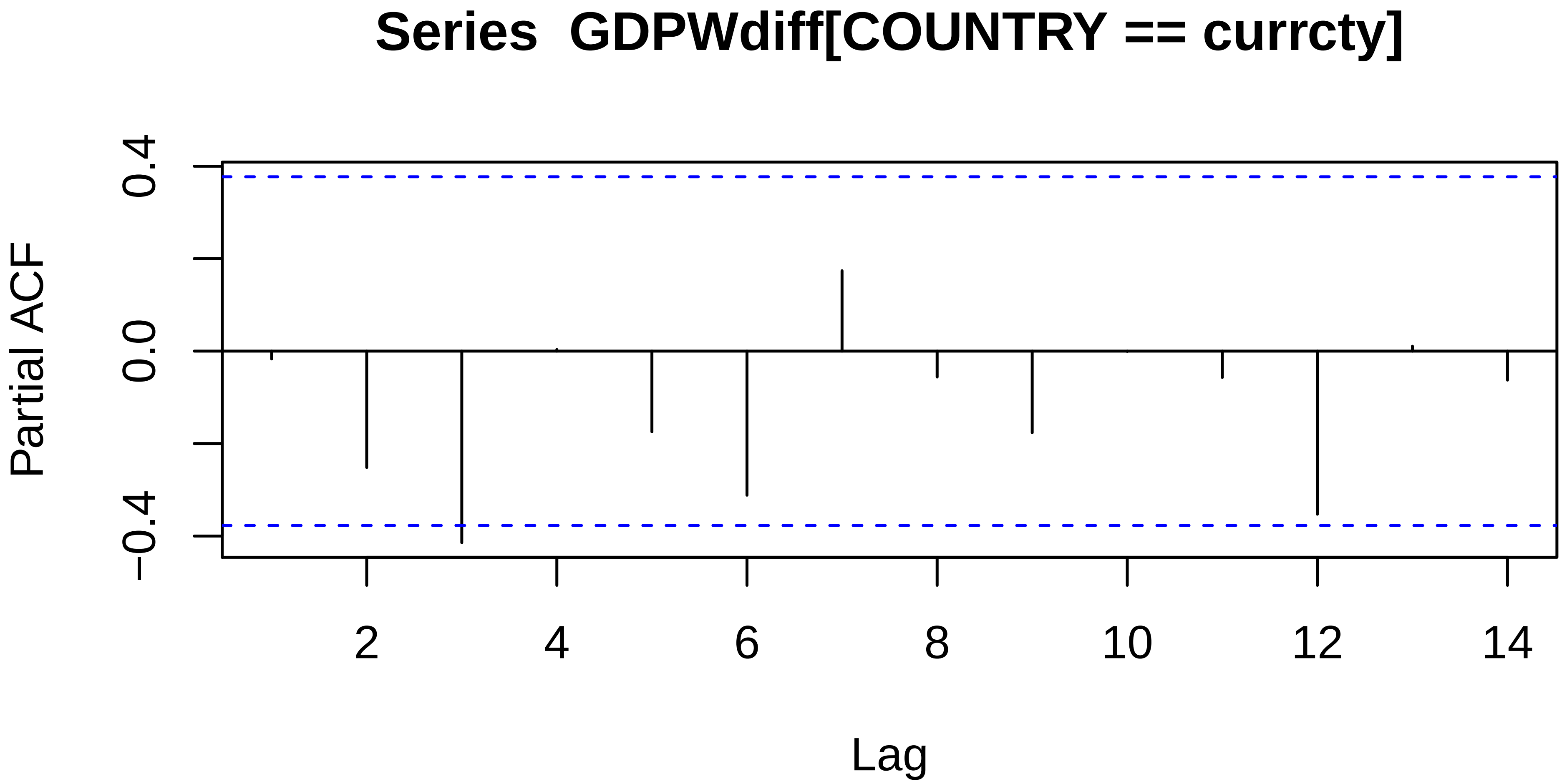
GDPdiff ACF for country 113



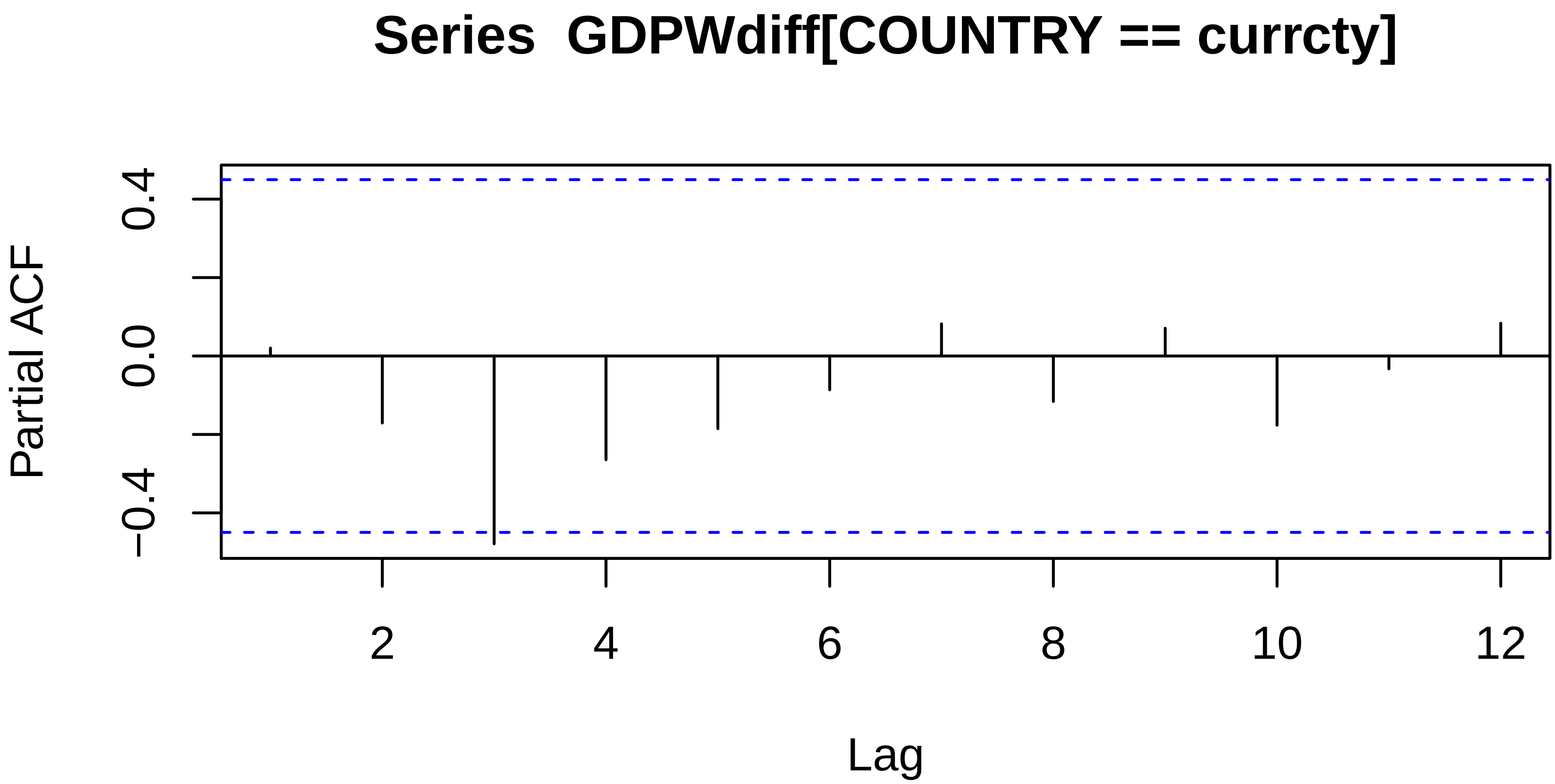
GDPdiff PACF for country 1



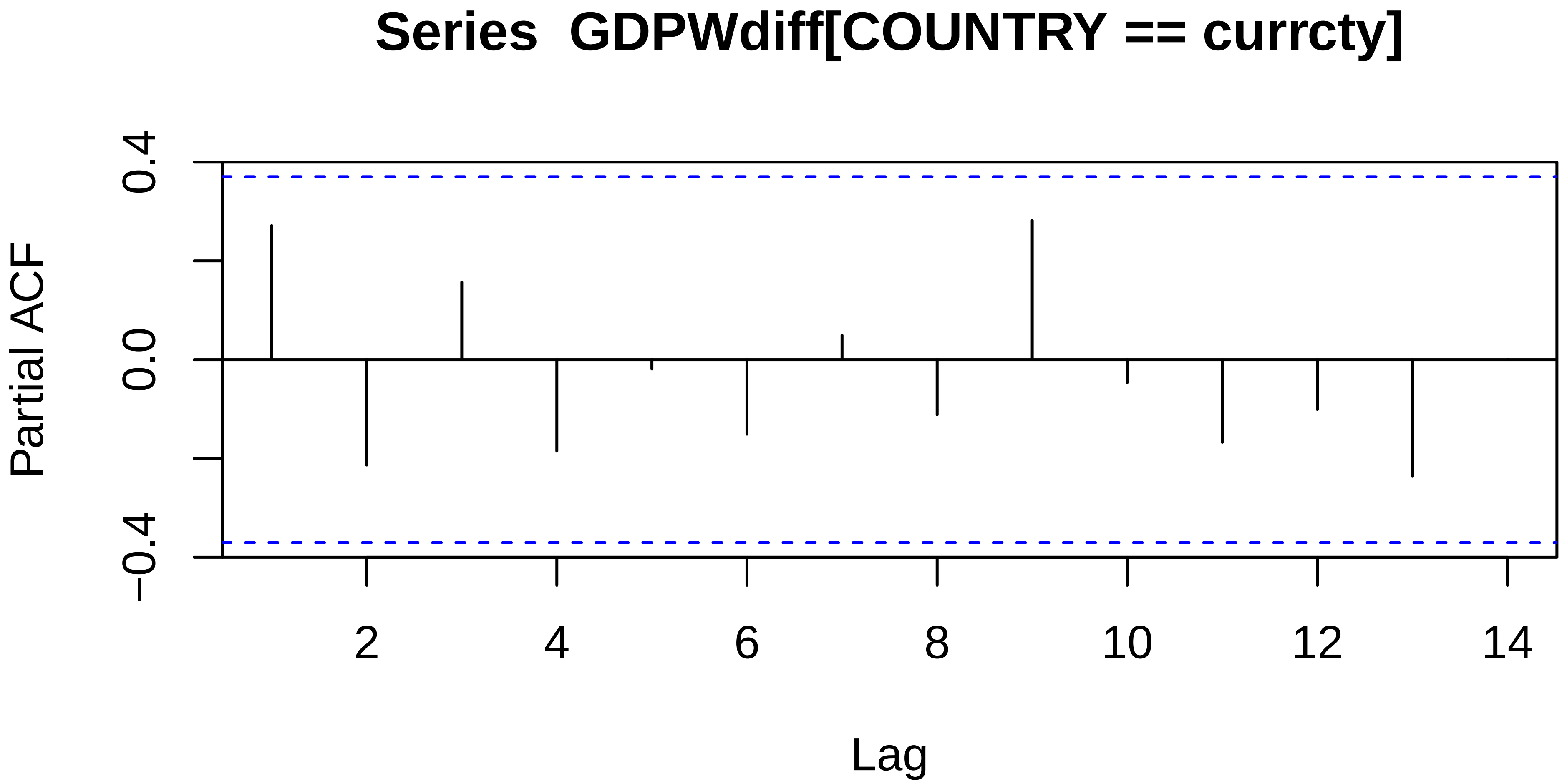
GDPdiff PACF for country 2



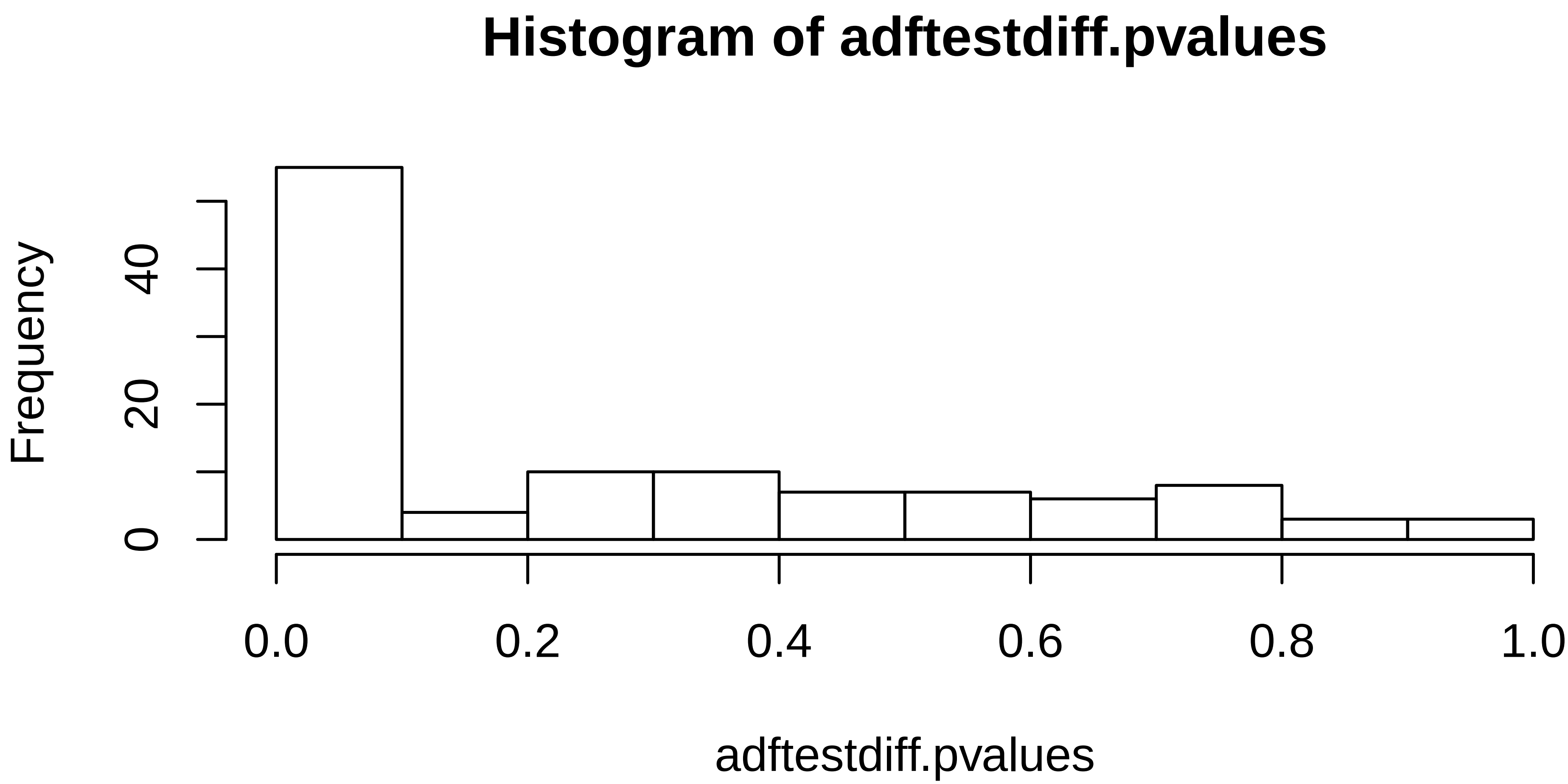
GDPdiff PACF for country 3



GDPdiff PACF for country 4

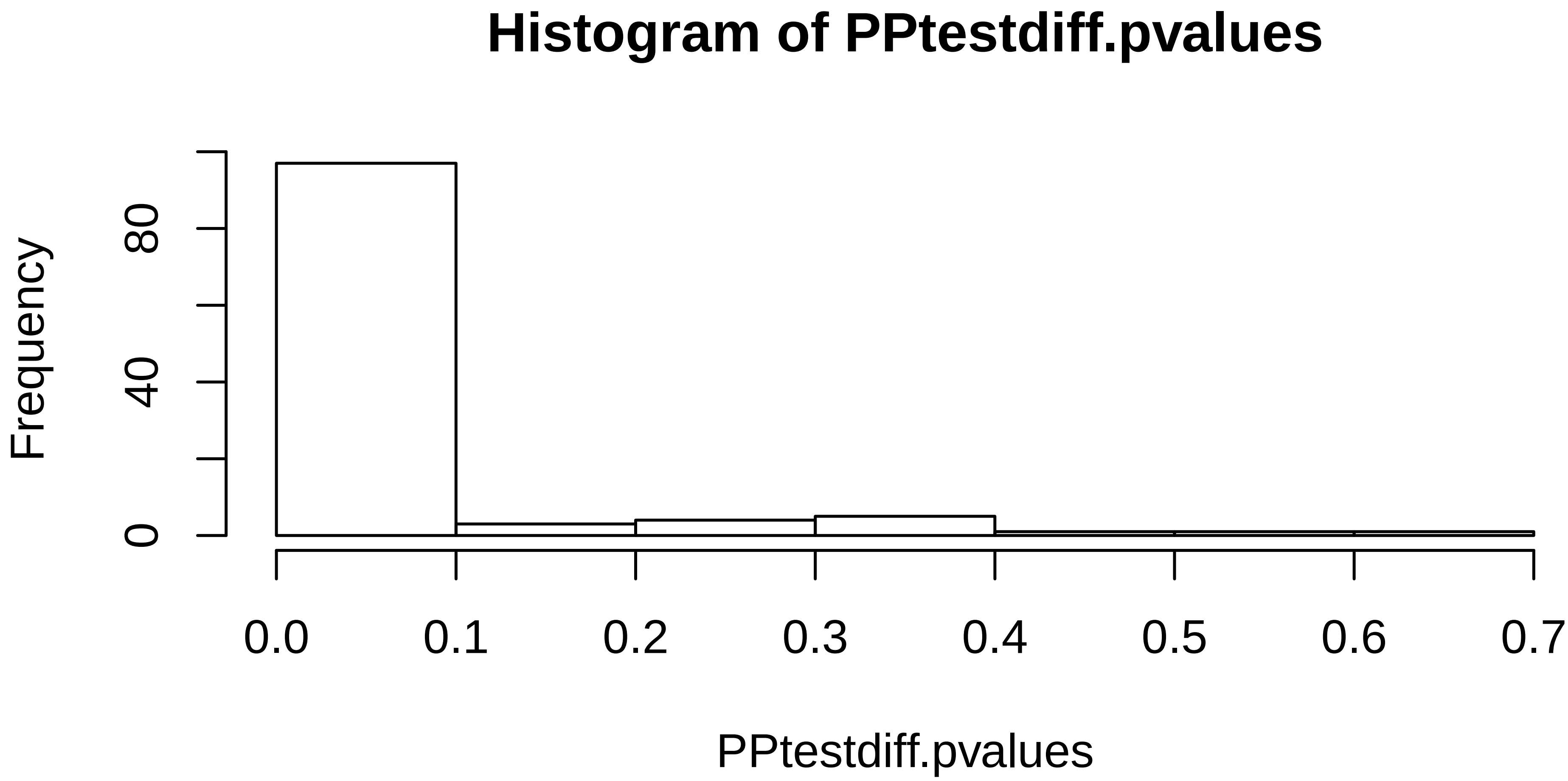


GDPdiff PACF for country 113



Histogram of p-values from ADF tests on GDPWdiff

What is this pattern consistent with?



Histogram of p-values from Phillips-Peron tests on GDPWdiff

Example continued in R demonstration

We will continue this example using the code provided

Remember: our goal is to simulate what happens to this time series after a change in our covariates

And to see if these estimates change when we include random or fixed effects

In particular, we want to see if fixed effects can help us with omitted time invariant variables, which are legion in this example

Panel Wrap-Up

- Bias in Panel with small T
- Heteroskedasticity in Time
- Heteroskedasticity in Panel

Bias in Panels with small T

Consider this panel model:

$$y_{it} = \phi y_{i,t-1} + \mathrm{x}_{it}\boldsymbol{\beta} + \alpha_i + \varepsilon_{it}$$

Remove the fixed effects α_i by differencing (the within estimator)

$$y_{it} - \bar{y}_i = \phi(y_{i,t-1} - \bar{y}_i) + (\mathrm{x}_{it} - \bar{\mathrm{x}}_i)\boldsymbol{\beta} + (\varepsilon_{it} - \bar{\varepsilon}_i)$$

What happens if we estimate this with LS?

Notice that $\bar{y}_i$ must be correlated with the new error term $\varepsilon_{it} - \bar{\varepsilon}_i$

Bias in Panels with small T

$$y_{it} - \bar{y}_i = \phi(y_{i,t-1} - \bar{y}_i) + (\mathrm{x}_{it} - \bar{\mathrm{x}}_i)\boldsymbol{\beta} + (\varepsilon_{it} - \bar{\varepsilon}_i)$$

When a regressor and the error are correlated, estimates are biased and inconsistent

In this case estimates are inconsistent in N: $\quad \hat{\phi} - \phi \nrightarrow 0 \quad$ as $\quad N \rightarrow \infty$

Bias in Panels with small T

$$y_{it} - \bar{y}_i = \phi(y_{i,t-1} - \bar{y}_i) + (\mathrm{x}_{it} - \bar{\mathrm{x}}_i)\boldsymbol{\beta} + (\varepsilon_{it} - \bar{\varepsilon}_i)$$

When a regressor and the error are correlated, estimates are biased and inconsistent

In this case estimates are inconsistent in N: $\hat{\phi} - \phi \nrightarrow 0$ as $N \rightarrow \infty$

However, estimates are still consistent in T: $\hat{\phi} - \phi \rightarrow 0$ as $T \rightarrow \infty$

Bias in Panels with small T

$$y_{it} - \bar{y}_i = \phi(y_{i,t-1} - \bar{y}_i) + (\mathrm{x}_{it} - \bar{\mathrm{x}}_i)\boldsymbol{\beta} + (\varepsilon_{it} - \bar{\varepsilon}_i)$$

When a regressor and the error are correlated, estimates are biased and inconsistent

In this case estimates are inconsistent in N: $\quad \hat{\phi} - \phi \nrightarrow 0 \quad$ as $\quad N \rightarrow \infty$

However, estimates are still consistent in T: $\quad \hat{\phi} - \phi \rightarrow 0 \quad$ as $\quad T \rightarrow \infty$

The degree of bias is of order $1/T$, so for small T, this is a big problem

Generally, $\quad \hat{\phi} - \phi \rightarrow (\phi - 1)/(T - 1) \quad$ as $\quad N \rightarrow \infty$

Bias in Panels with small T

$$y_{it} - \bar{y}_i = \phi(y_{i,t-1} - \bar{y}_i) + (\mathrm{x}_{it} - \bar{\mathrm{x}}_i)\boldsymbol{\beta} + (\varepsilon_{it} - \bar{\varepsilon}_i)$$

When a regressor and the error are correlated, estimates are biased and inconsistent

In this case estimates are inconsistent in N: $\quad \hat{\phi} - \phi \nrightarrow 0 \quad$ as $\quad N \rightarrow \infty$

However, estimates are still consistent in T: $\quad \hat{\phi} - \phi \rightarrow 0 \quad$ as $\quad T \rightarrow \infty$

The degree of bias is of order $1/T$, so for small T, this is a big problem

Generally, $\quad \hat{\phi} - \phi \rightarrow (\phi - 1)/(T - 1) \quad$ as $\quad N \rightarrow \infty$

ϕ	T	N	Bias $\hat{\phi} - \phi$	% Bias $\lvert(\hat{\phi} - \phi)/\phi\rvert$
0.5	5	∞	-0.167	0.33
0.5	10	∞	-0.055	0.11
0.5	25	∞	-0.020	0.04
0.5	50	∞	-0.010	0.02
0.5	100	∞	-0.005	0.01
0.5	∞	∞	0	0

So you can judge how worried you should be just from N and T

Bias in Panels with small T

Things that **don't** help:

- More units N
- Purging serial correlation in ε
- Getting the specification of x right

What would help:

One way to deal with correlated covariates & errors is with *instrumental* variables

If x_{it} and ε_{it} are correlated, find some z_{it} which is correlated with x_{it} but not ε_{it}

z_{it} is then an *instrument* for x_{it}

What is an instrument

In words, an instrument z_{it}

- explains part of x_{it},
- but does not *otherwise* explain y_{it},
- so it does not belong in our model,
- but can be used to distinguish the part of x_{it} that influences y_{it} from the part that influences ε_{it}

Background: IV estimation

Consider a bivariate regression

$$y_i = \beta x_i + \varepsilon_i$$

We condition on x_i and take expectations,
assuming no correlation of the error with x_i,

$$\mathrm{E}(y_i|x_i) = \mathrm{E}(\beta x_i|x_i) + \mathrm{E}(\varepsilon_i|x_i)$$

Background: IV estimation

Consider a bivariate regression

$$y_i = \beta x_i + \varepsilon_i$$

We condition on x_i and take expectations,
assuming no correlation of the error with x_i,

$$\begin{aligned}
\mathrm{E}(y_i|x_i) &= \mathrm{E}(\beta x_i|x_i) + \mathrm{E}(\varepsilon_i|x_i) \\
\mathrm{E}(y_i|x_i) &= \beta x_i + 0
\end{aligned}$$

Background: IV estimation

Consider a bivariate regression

$$y_i = \beta x_i + \varepsilon_i$$

We condition on x_i and take expectations,
assuming no correlation of the error with x_i,

$$\begin{aligned}
\mathrm{E}(y_i|x_i) &= \mathrm{E}(\beta x_i|x_i) + \mathrm{E}(\varepsilon_i|x_i) \\
\mathrm{E}(y_i|x_i) &= \beta x_i + 0 \\
\mathrm{E}(y_i|x_i) &= \beta x_i
\end{aligned}$$

Background: IV estimation

What if x_i and ε_i are correlated?

$$\hat{\beta}_{\mathrm{LS}} = \frac{\sum_i x_i y_i}{\sum_i x_i^2}$$

Background: IV estimation

What if x_i and ε_i are correlated?

$$\begin{aligned}\hat{\beta}_{\mathrm{LS}} &= \frac{\sum_i x_i y_i}{\sum_i x_i^2} \\ &= \frac{\sum_i x_i(\beta x_i + \varepsilon_i)}{\sum_i x_i^2}\end{aligned}$$

Background: IV estimation

What if x_i and ε_i are correlated?

$$\begin{aligned}\hat{\beta}_{\mathrm{LS}} &= \frac{\sum_i x_i y_i}{\sum_i x_i^2} \\ &= \frac{\sum_i x_i(\beta x_i + \varepsilon_i)}{\sum_i x_i^2} \\ &= \frac{\sum_i x_i^2 \beta + \sum_i x_i \varepsilon_i}{\sum_i x_i^2}\end{aligned}$$

Background: IV estimation

What if x_i and ε_i are correlated?

$$\begin{aligned}
\hat{\beta}_{\mathrm{LS}} &= \frac{\sum_i x_i y_i}{\sum_i x_i^2} \\
&= \frac{\sum_i x_i(\beta x_i + \varepsilon_i)}{\sum_i x_i^2} \\
&= \frac{\sum_i x_i^2 \beta + \sum_i x_i \varepsilon_i}{\sum_i x_i^2} \\
&= \frac{\sum_i x_i^2 \beta}{\sum_i x_i^2} + \frac{\sum_i x_i \varepsilon_i}{\sum_i x_i^2}
\end{aligned}$$

Background: IV estimation

What if x_i and ε_i are correlated?

$$\begin{aligned}
\hat{\beta}_{\mathrm{LS}} &= \frac{\sum_i x_i y_i}{\sum_i x_i^2} \\
&= \frac{\sum_i x_i(\beta x_i + \varepsilon_i)}{\sum_i x_i^2} \\
&= \frac{\sum_i x_i^2 \beta + \sum_i x_i \varepsilon_i}{\sum_i x_i^2} \\
&= \frac{\sum_i x_i^2 \beta}{\sum_i x_i^2} + \frac{\sum_i x_i \varepsilon_i}{\sum_i x_i^2} \\
&= \beta + \frac{\sum_i x_i \varepsilon_i}{\sum_i x_i^2}
\end{aligned}$$

LS is biased

Background: IV estimation

Now try an instrument z_i for x_i

Condition on z_i and take expectations:

$$\mathrm{E}(y_i|z_i) = \mathrm{E}(\beta x_i|z_i) + \mathrm{E}(\varepsilon_i|z_i)$$

Background: IV estimation

Now try an instrument z_i for x_i

Condition on z_i and take expectations:

$$\mathrm{E}(y_i|z_i) = \mathrm{E}(\beta x_i|z_i) + \mathrm{E}(\varepsilon_i|z_i)$$

We assumed that $\mathrm{E}(\varepsilon_i|z_i) = 0$, so we can write

$$\mathrm{E}(y_i|z_i) = \mathrm{E}(\beta x_i|z_i)$$

Background: IV estimation

Now try an instrument z_i for x_i

Condition on z_i and take expectations:

$$\mathrm{E}(y_i|z_i) = \mathrm{E}(\beta x_i|z_i) + \mathrm{E}(\varepsilon_i|z_i)$$

We assumed that $\mathrm{E}(\varepsilon_i|z_i) = 0$, so we can write

$$\mathrm{E}(y_i|z_i) = \mathrm{E}(\beta x_i|z_i)$$

$$\mathrm{E}(y_i|z_i) = \beta\mathrm{E}(x_i|z_i)$$

Background: IV estimation

Now try an instrument z_i for x_i

Condition on z_i and take expectations:

$$\mathrm{E}(y_i|z_i) = \mathrm{E}(\beta x_i|z_i) + \mathrm{E}(\varepsilon_i|z_i)$$

We assumed that $\mathrm{E}(\varepsilon_i|z_i) = 0$, so we can write

$$\begin{aligned}
\mathrm{E}(y_i|z_i) &= \mathrm{E}(\beta x_i|z_i) \\
\mathrm{E}(y_i|z_i) &= \beta \mathrm{E}(x_i|z_i) \\
\beta &= \frac{\mathrm{E}(x_i|z_i)}{\mathrm{E}(y_i|z_i)}
\end{aligned}$$

Solving for β will give us the IV estimator, $\hat{\beta}_{\mathrm{IV}}$

Background: IV estimation

This way of finding the IV estimator is known as the method of moments

So called because we are working just in the expectations of our variables, not with their complete probability distributions

An alternative to maximum likelihood estimation

Background: IV estimation

The IV estimator is consistent if z_i is an instrument:

$$\beta = \frac{\mathrm{E}(x_i|z_i)}{\mathrm{E}(y_i|z_i)}$$

Background: IV estimation

The IV estimator is consistent if z_i is an instrument:

$$\beta = \frac{\mathrm{E}(x_i|z_i)}{\mathrm{E}(y_i|z_i)}$$

$$\hat{\beta}_{\mathrm{IV}} = \frac{\sum_i z_i y_i}{\sum_i z_i x_i}$$

Background: IV estimation

The IV estimator is consistent if z_i is an instrument:

$$\beta = \frac{\mathrm{E}(x_i|z_i)}{\mathrm{E}(y_i|z_i)}$$

$$\hat{\beta}_{\mathrm{IV}} = \frac{\sum_i z_i y_i}{\sum_i z_i x_i}$$

$$= \frac{\sum_i z_i(\beta x_i + \varepsilon_i)}{\sum_i z_i x_i}$$

Background: IV estimation

The IV estimator is consistent if z_i is an instrument:

$$\beta = \frac{\mathrm{E}(x_i|z_i)}{\mathrm{E}(y_i|z_i)}$$

$$\begin{aligned}
\hat{\beta}_{\mathrm{IV}} &= \frac{\sum_i z_i y_i}{\sum_i z_i x_i} \\
&= \frac{\sum_i z_i(\beta x_i + \varepsilon_i)}{\sum_i z_i x_i} \\
&= \frac{\sum_i z_i x_i \beta + \sum_i z_i \varepsilon_i}{\sum_i z_i x_i}
\end{aligned}$$

Background: IV estimation

The IV estimator is consistent if z_i is an instrument:

$$\begin{aligned}
\beta &= \frac{\mathrm{E}(x_i|z_i)}{\mathrm{E}(y_i|z_i)} \\
\hat{\beta}_{\mathrm{IV}} &= \frac{\sum_i z_i y_i}{\sum_i z_i x_i} \\
&= \frac{\sum_i z_i(\beta x_i + \varepsilon_i)}{\sum_i z_i x_i} \\
&= \frac{\sum_i z_i x_i \beta + \sum_i z_i \varepsilon_i}{\sum_i z_i x_i} \\
&= \frac{\sum_i z_i x_i \beta}{\sum_i z_i x_i} + \frac{\sum_i z_i \varepsilon_i}{\sum_i z_i x_i}
\end{aligned}$$

Background: IV estimation

The IV estimator is consistent if z_i is an instrument:

$$\begin{aligned}
\beta &= \frac{\mathrm{E}(x_i|z_i)}{\mathrm{E}(y_i|z_i)} \\
\hat{\beta}_{\mathrm{IV}} &= \frac{\sum_i z_i y_i}{\sum_i z_i x_i} \\
&= \frac{\sum_i z_i(\beta x_i + \varepsilon_i)}{\sum_i z_i x_i} \\
&= \frac{\sum_i z_i x_i \beta + \sum_i z_i \varepsilon_i}{\sum_i z_i x_i} \\
&= \frac{\sum_i z_i x_i \beta}{\sum_i z_i x_i} + \frac{\sum_i z_i \varepsilon_i}{\sum_i z_i x_i} \\
&= \beta + \frac{\sum_i z_i \varepsilon_i}{\sum_i z_i x_i} \\
&= \beta
\end{aligned}$$

IV estimation for Panel

Return to our panel model,

$$y_{it} - \bar{y}_i = \phi(y_{i,t-1} - \bar{y}_i) + (\mathrm{x}_{it} - \bar{\mathrm{x}}_i)\boldsymbol{\beta} + (\varepsilon_{it} - \bar{\varepsilon}_i)$$

Now take first differences to obtain

$$\Delta y_{it} = \phi \Delta y_{i,t-1} + \Delta \mathrm{x}_{it}\boldsymbol{\beta} + \Delta\varepsilon_{it}$$

Note that $\Delta\varepsilon_{it}$ is now MA(1)

If we had instruments for $\Delta y_{i,t-1}$,
we could correct the bias in estimation of ϕ and thus potentially β

We do have some *weak* instruments for $\Delta y_{i,t-1}$, notably its lags

That is, $\Delta y_{i,t-2}$ helps predict $\Delta y_{i,t-1}$, but not $\Delta\varepsilon_{i,t-1}$

Only a weak instrument, though

Using these lags in IV estimation: Anderson-Hsiao estimator

Why are lagged changes in y instruments?

Fixed Effects				Differenced Fixed Effects		
Dependent Variable	Most recent error	Error Term	lag of Dependent Variable	Dep Var	Error Term	lag of Dep Var
$y_{i,t-3} - \bar{y}_i$	$\varepsilon_{i,t-3}$	$\varepsilon_{i,t-3} - \bar{\varepsilon}_i$				
$y_{i,t-2} - \bar{y}_i$	$\varepsilon_{i,t-2}$	$\varepsilon_{i,t-2} - \bar{\varepsilon}_i$	$y_{i,t-3} - \bar{y}_i$	$\Delta y_{i,t-2}$	$\varepsilon_{i,t-2} - \varepsilon_{i,t-3}$	
$y_{i,t-1} - \bar{y}_i$	$\varepsilon_{i,t-1}$	$\varepsilon_{i,t-1} - \bar{\varepsilon}_i$	$y_{i,t-2} - \bar{y}_i$	$\Delta y_{i,t-1}$	$\varepsilon_{i,t-1} - \varepsilon_{i,t-2}$	$\Delta y_{i,t-2}$
$y_{it} - \bar{y}_i$	ε_{it}	$\varepsilon_{it} - \bar{\varepsilon}_i$	$y_{i,t-1} - \bar{y}_i$	Δy_{it}	$\varepsilon_{it} - \varepsilon_{i,t-1}$	$\Delta y_{i,t-1}$

To see which lags of the differences and levels of $y_{it} - \bar{y}_i$

are instruments for the first lag of the difference, $\Delta y_{i,t-1}$,

we need to see which lags are correlated $\Delta y_{i,t-1}$ but not Δy_{it}

Why are lagged changes in y instruments?

Fixed Effects				Differenced Fixed Effects		
Dependent Variable	Most recent error	Error Term	lag of Dependent Variable	Dep Var	Error Term	lag of Dep Var
$y_{i,t-3} - \bar{y}_i$	$\varepsilon_{i,t-3}$	$\varepsilon_{i,t-3} - \bar{\varepsilon}_i$				
$y_{i,t-2} - \bar{y}_i$	$\varepsilon_{i,t-2}$	$\varepsilon_{i,t-2} - \bar{\varepsilon}_i$	$y_{i,t-3} - \bar{y}_i$	$\Delta y_{i,t-2}$	$\varepsilon_{i,t-2} - \varepsilon_{i,t-3}$	
$y_{i,t-1} - \bar{y}_i$	$\varepsilon_{i,t-1}$	$\varepsilon_{i,t-1} - \bar{\varepsilon}_i$	$y_{i,t-2} - \bar{y}_i$	$\Delta y_{i,t-1}$	$\varepsilon_{i,t-1} - \varepsilon_{i,t-2}$	$\Delta y_{i,t-2}$
$y_{it} - \bar{y}_i$	ε_{it}	$\varepsilon_{it} - \bar{\varepsilon}_i$	$y_{i,t-1} - \bar{y}_i$	Δy_{it}	$\varepsilon_{it} - \varepsilon_{i,t-1}$	$\Delta y_{i,t-1}$

Which error terms are incorporated in Δy_{it}?

Why are lagged changes in y instruments?

Fixed Effects				Differenced Fixed Effects		
Dependent Variable	Most recent error	Error Term	lag of Dependent Variable	Dep Var	Error Term	lag of Dep Var
$y_{i,t-3} - \bar{y}_i$	$\varepsilon_{i,t-3}$	$\varepsilon_{i,t-3} - \bar{\varepsilon}_i$				
$y_{i,t-2} - \bar{y}_i$	$\varepsilon_{i,t-2}$	$\varepsilon_{i,t-2} - \bar{\varepsilon}_i$	$y_{i,t-3} - \bar{y}_i$	$\Delta y_{i,t-2}$	$\varepsilon_{i,t-2} - \varepsilon_{i,t-3}$	
$y_{i,t-1} - \bar{y}_i$	$\varepsilon_{i,t-1}$	$\varepsilon_{i,t-1} - \bar{\varepsilon}_i$	$y_{i,t-2} - \bar{y}_i$	$\Delta y_{i,t-1}$	$\varepsilon_{i,t-1} - \varepsilon_{i,t-2}$	$\Delta y_{i,t-2}$
$y_{it} - \bar{y}_i$	ε_{it}	$\varepsilon_{it} - \bar{\varepsilon}_i$	$y_{i,t-1} - \bar{y}_i$	Δy_{it}	$\varepsilon_{it} - \varepsilon_{i,t-1}$	$\Delta y_{i,t-1}$

Which error terms are incorporated in Δy_{it}?

The terms in yellow are differenced to make Δy_{it},

Why are lagged changes in y instruments?

Fixed Effects				Differenced Fixed Effects		
Dependent Variable	Most recent error	Error Term	lag of Dependent Variable	Dep Var	Error Term	lag of Dep Var
$y_{i,t-3} - \bar{y}_i$	$\varepsilon_{i,t-3}$	$\varepsilon_{i,t-3} - \bar{\varepsilon}_i$				
$y_{i,t-2} - \bar{y}_i$	$\varepsilon_{i,t-2}$	$\varepsilon_{i,t-2} - \bar{\varepsilon}_i$	$y_{i,t-3} - \bar{y}_i$	$\Delta y_{i,t-2}$	$\varepsilon_{i,t-2} - \varepsilon_{i,t-3}$	
$y_{i,t-1} - \bar{y}_i$	$\varepsilon_{i,t-1}$	$\varepsilon_{i,t-1} - \bar{\varepsilon}_i$	$y_{i,t-2} - \bar{y}_i$	$\Delta y_{i,t-1}$	$\varepsilon_{i,t-1} - \varepsilon_{i,t-2}$	$\Delta y_{i,t-2}$
$y_{it} - \bar{y}_i$	ε_{it}	$\varepsilon_{it} - \bar{\varepsilon}_i$	$y_{i,t-1} - \bar{y}_i$	Δy_{it}	$\varepsilon_{it} - \varepsilon_{i,t-1}$	$\Delta y_{i,t-1}$

Which error terms are incorporated in Δy_{it}?

The terms in yellow are differenced to make Δy_{it},

and so Δy_{it} includes *only* the difference of errors also in yellow

Why are lagged changes in y instruments?

Fixed Effects				Differenced Fixed Effects		
Dependent Variable	Most recent error	Error Term	lag of Dependent Variable	Dep Var	Error Term	lag of Dep Var
$y_{i,t-3} - \bar{y}_i$	$\varepsilon_{i,t-3}$	$\varepsilon_{i,t-3} - \bar{\varepsilon}_i$				
$y_{i,t-2} - \bar{y}_i$	$\varepsilon_{i,t-2}$	$\varepsilon_{i,t-2} - \bar{\varepsilon}_i$	$y_{i,t-3} - \bar{y}_i$	$\Delta y_{i,t-2}$	$\varepsilon_{i,t-2} - \varepsilon_{i,t-3}$	
$y_{i,t-1} - \bar{y}_i$	$\varepsilon_{i,t-1}$	$\varepsilon_{i,t-1} - \bar{\varepsilon}_i$	$y_{i,t-2} - \bar{y}_i$	$\Delta y_{i,t-1}$	$\varepsilon_{i,t-1} - \varepsilon_{i,t-2}$	$\Delta y_{i,t-2}$
$y_{it} - \bar{y}_i$	ε_{it}	$\varepsilon_{it} - \bar{\varepsilon}_i$	$y_{i,t-1} - \bar{y}_i$	Δy_{it}	$\varepsilon_{it} - \varepsilon_{i,t-1}$	$\Delta y_{i,t-1}$

Is the term in red, $\Delta y_{i,t-2}$, an instrument?

Why are lagged changes in y instruments?

Fixed Effects				Differenced Fixed Effects		
Dependent Variable	Most recent error	Error Term	lag of Dependent Variable	Dep Var	Error Term	lag of Dep Var
$y_{i,t-3} - \bar{y}_i$	$\varepsilon_{i,t-3}$	$\varepsilon_{i,t-3} - \bar{\varepsilon}_i$				
$y_{i,t-2} - \bar{y}_i$	$\varepsilon_{i,t-2}$	$\varepsilon_{i,t-2} - \bar{\varepsilon}_i$	$y_{i,t-3} - \bar{y}_i$	$\Delta y_{i,t-2}$	$\varepsilon_{i,t-2} - \varepsilon_{i,t-3}$	
$y_{i,t-1} - \bar{y}_i$	$\varepsilon_{i,t-1}$	$\varepsilon_{i,t-1} - \bar{\varepsilon}_i$	$y_{i,t-2} - \bar{y}_i$	$\Delta y_{i,t-1}$	$\varepsilon_{i,t-1} - \varepsilon_{i,t-2}$	$\Delta y_{i,t-2}$
$y_{it} - \bar{y}_i$	ε_{it}	$\varepsilon_{it} - \bar{\varepsilon}_i$	$y_{i,t-1} - \bar{y}_i$	Δy_{it}	$\varepsilon_{it} - \varepsilon_{i,t-1}$	$\Delta y_{i,t-1}$

Is the term in red, $\Delta y_{i,t-2}$, an instrument?

The terms in red are differenced to make $\Delta y_{i,t-2}$,

Why are lagged changes in y instruments?

Fixed Effects				Differenced Fixed Effects		
Dependent Variable	Most recent error	Error Term	lag of Dependent Variable	Dep Var	Error Term	lag of Dep Var
$y_{i,t-3} - \bar{y}_i$	$\varepsilon_{i,t-3}$	$\varepsilon_{i,t-3} - \bar{\varepsilon}_i$				
$y_{i,t-2} - \bar{y}_i$	$\varepsilon_{i,t-2}$	$\varepsilon_{i,t-2} - \bar{\varepsilon}_i$	$y_{i,t-3} - \bar{y}_i$	$\Delta y_{i,t-2}$	$\varepsilon_{i,t-2} - \varepsilon_{i,t-3}$	
$y_{i,t-1} - \bar{y}_i$	$\varepsilon_{i,t-1}$	$\varepsilon_{i,t-1} - \bar{\varepsilon}_i$	$y_{i,t-2} - \bar{y}_i$	$\Delta y_{i,t-1}$	$\varepsilon_{i,t-1} - \varepsilon_{i,t-2}$	$\Delta y_{i,t-2}$
$y_{it} - \bar{y}_i$	ε_{it}	$\varepsilon_{it} - \bar{\varepsilon}_i$	$y_{i,t-1} - \bar{y}_i$	Δy_{it}	$\varepsilon_{it} - \varepsilon_{i,t-1}$	$\Delta y_{i,t-1}$

Is the term in red, $\Delta y_{i,t-2}$, an instrument?

The terms in red are differenced to make $\Delta y_{i,t-2}$,

and so $\Delta y_{i,t-2}$ includes *only* the difference of errors in red

And there is no correlation between $\varepsilon_{i,t-2} - \varepsilon_{i,t-3}$ and $\varepsilon_{it} - \varepsilon_{i,t-1}$

So $\Delta y_{i,t-2}$ is an instrument for $\Delta y_{i,t-1}$ in a model of Δy_{it}

IV estimation for Panel

Arellano & Bond noticed that within a dataset with N units and T periods, later observations have more lags available as instruments

For $t = 4$ differenced periods, can create: $\Delta y_{i,t-2}$

For $t = 5$ differenced periods, can create: $\Delta y_{i,t-2}$, $\Delta y_{i,t-3}$

For $t = 6$ differenced periods, can create: $\Delta y_{i,t-2}$, $\Delta y_{i,t-3}$, $\Delta y_{i,t-4}$

etc.

Arellano-Bond's difference GMM estimator uses all available instruments in each period

Arellano-Bond's system GMM estimator adds in the lagged levels $y_{i,t-2}$, etc., as additional instruments

Estimation is by Generalized Method of Moments—IV by itself can't handle the varying number of instruments

IV estimation for Panel

Note that for either system or difference GMM, we need to be careful with any AR(p) processes left *after* differencing

That is, if we have an ARIMA(p,1,0), $p > 0$, very recent lags of y or Δy will not be instruments:

If $\Delta y_{i,t}$ is AR(1), will need to start one period earlier on instruments

I.e., if $\Delta y_{i,t}$ is AR(1), then first available instrument is $\Delta y_{i,t-3}$

if $\Delta y_{i,t}$ is AR(2), then first available instrument is $\Delta y_{i,t-4}$

etc.

In `R`, these estimators are available using `pgmm` in the `plm` library

Note that this library requires a special kind of data frame, created with `plm.data`, so that it knows what the units and periods are

Review of heteroskedasticity

Recall that in cross-sectional LS, heteroskedasticity

- is assumed away
- if present, biases our standard errors

We noted two approaches

- Model the heteroskedasticity directly with an appropriate ML model, *or*
- Less optimally, continue to use the wrong method (LS), but try to correct the se's

Heteroskedasticity Robust SEs

We used the following formula to correct LS se's for heteroskedasticty

$$\hat{\mathrm{V}}(\hat{\boldsymbol{\beta}}) = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\hat{\Omega}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}$$

If we use the observed residuals as estimates of the observation-specific variance, we have $\hat{\Omega} = \mathrm{diag}(\hat{\varepsilon}_1^2, \hat{\varepsilon}_2^2, \ldots \hat{\varepsilon}_n^2)$, yielding the (multiply named)

- White standard errors
- robust standard errors
- sandwich standard errors
- heteroskedastcity consistent standard errors

Dynamic heteroskedasticity

As with cross-sectional models, we can model heteroskedasticity directly

One possibility is to let heteroskedasticity evolve dynamically

Dynamic heteroskedasticity

As with cross-sectional models, we can model heteroskedasticity directly

One possibility is to let heteroskedasticity evolve dynamically

We can let heteroskedasticity be (sort-of) "ARMA", under the name "GARCH"
Generalized Autoregressive Conditional Heteroskedasticity:

$$y_t = \mu_t + \varepsilon_t \qquad \varepsilon_t \sim f_{\mathcal{N}}\left(0, \sigma_t^2\right)$$

Dynamic heteroskedasticity

As with cross-sectional models, we can model heteroskedasticity directly

One possibility is to let heteroskedasticity evolve dynamically

We can let heteroskedasticity be (sort-of) "ARMA", under the name "GARCH"
Generalized Autoregressive Conditional Heteroskedasticity:

$$y_t = \mu_t + \varepsilon_t \qquad \varepsilon_t \sim f_{\mathcal{N}}\left(0, \sigma_t^2\right)$$

where

$$\mu_t = \alpha + x_t\beta + \sum_{p=1}^{P} y_{t-p}\phi_p + \sum_{q=1}^{Q} \varepsilon_{t-q}\theta_q$$

Dynamic heteroskedasticity

As with cross-sectional models, we can model heteroskedasticity directly

One possibility is to let heteroskedasticity evolve dynamically

We can let heteroskedasticity be (sort-of) "ARMA", under the name "GARCH"
Generalized Autoregressive Conditional Heteroskedasticity:

$$y_t = \mu_t + \varepsilon_t \qquad \varepsilon_t \sim f_{\mathcal{N}}\left(0, \sigma_t^2\right)$$

where

$$\mu_t = \alpha + x_t\beta + \sum_{p=1}^{P} y_{t-p}\phi_p + \sum_{q=1}^{Q} \varepsilon_{t-q}\theta_q$$

$$\sigma_t^2 = \exp\left(\eta + z_t\gamma\right) + \sum_{c=1}^{C} \sigma_{t-c}^2\lambda_c + \sum_{d=1}^{D} \varepsilon_{t-d}^2\xi_d$$

Dynamic heteroskedasticity

As with cross-sectional models, we can model heteroskedasticity directly

One possibility is to let heteroskedasticity evolve dynamically

We can let heteroskedasticity be (sort-of) "ARMA", under the name "GARCH" Generalized Autoregressive Conditional Heteroskedasticity:

$$y_t = \mu_t + \varepsilon_t \qquad \varepsilon_t \sim f_{\mathcal{N}}\left(0, \sigma_t^2\right)$$

where

$$\mu_t = \alpha + x_t\beta + \sum_{p=1}^{P} y_{t-p}\phi_p + \sum_{q=1}^{Q} \varepsilon_{t-q}\theta_q$$

$$\sigma_t^2 = \exp\left(\eta + z_t\gamma\right) + \sum_{c=1}^{C} \sigma_{t-c}^2\lambda_c + \sum_{d=1}^{D} \varepsilon_{t-d}^2\xi_d$$

In words, y_t is an ARMA(P, Q)-GARCH(C, D) distributed time-series

(Of course, we could leave out x and/or z if we wanted)

Dynamic heteroskedasticity

$$y_t = \mu_t + \varepsilon_t \qquad \varepsilon_t \sim f_{\mathcal{N}}\left(0, \sigma_t^2\right)$$

Dynamic heteroskedasticity

$$y_t = \mu_t + \varepsilon_t \qquad \varepsilon_t \sim f_{\mathcal{N}}\left(0, \sigma_t^2\right)$$

where

$$\mu_t = \alpha + x_t\beta + \sum_{p=1}^{P} y_{t-p}\phi_p + \sum_{q=1}^{Q} \varepsilon_{t-q}\theta_q$$

Dynamic heteroskedasticity

$$y_t = \mu_t + \varepsilon_t \qquad \varepsilon_t \sim f_{\mathcal{N}}\left(0, \sigma_t^2\right)$$

where

$$\mu_t = \alpha + x_t\beta + \sum_{p=1}^{P} y_{t-p}\phi_p + \sum_{q=1}^{Q} \varepsilon_{t-q}\theta_q$$

$$\sigma_t^2 = \exp\left(\eta + z_t\gamma\right) + \sum_{c=1}^{C} \sigma_{t-c}^2\lambda_c + \sum_{d=1}^{D} \varepsilon_{t-d}^2\xi_d$$

Dynamic heteroskedasticity

$$y_t = \mu_t + \varepsilon_t \qquad \varepsilon_t \sim f_{\mathcal{N}}\left(0, \sigma_t^2\right)$$

where

$$\mu_t = \alpha + x_t\beta + \sum_{p=1}^{P} y_{t-p}\phi_p + \sum_{q=1}^{Q} \varepsilon_{t-q}\theta_q$$

$$\sigma_t^2 = \exp\left(\eta + z_t\gamma\right) + \sum_{c=1}^{C} \sigma_{t-c}^2\lambda_c + \sum_{d=1}^{D} \varepsilon_{t-d}^2\xi_d$$

Models like the above are workhorses of financial forecasting

Can estimated by ML as usual

In `R`, `garch()` in the `tseries` package does GARCH

May have to look around a bit for ARMA-GARCH

Appears to be a package called `dynamo` in development that does ARMA-GARCH & forecasts from it here:
`http://cran.r-project.org/web/packages/dynamo/index.html`

Dynamic and panel heteroskedasticity

Three different kinds of heteroskedasticity:

- Cross-sectional
- Dynamic
- Panel

All could be combined

[On Blackboard]

Dynamic and panel heteroskedasticity

Consider a Panel ARMA-GARCH model:

$$y_{it} = \mu_{it} + \varepsilon_{it} \qquad \varepsilon_{it} \sim f_{\mathcal{N}}\left(0, \sigma_{it}^2\right)$$

Dynamic and panel heteroskedasticity

Consider a Panel ARMA-GARCH model:

$$y_{it} = \mu_{it} + \varepsilon_{it} \qquad \varepsilon_{it} \sim f_{\mathcal{N}}\left(0, \sigma_{it}^2\right)$$

$$\mu_{it} = \alpha_i + x_{it}\beta + \sum_{p=1}^{P} y_{i,t-p}\phi_p + \sum_{q=1}^{Q} \varepsilon_{i,t-q}\theta_q$$

Dynamic and panel heteroskedasticity

Consider a Panel ARMA-GARCH model:

$$y_{it} = \mu_{it} + \varepsilon_{it} \qquad \varepsilon_{it} \sim f_{\mathcal{N}}\left(0, \sigma^2_{it}\right)$$

$$\mu_{it} = \alpha_i + x_{it}\beta + \sum_{p=1}^{P} y_{i,t-p}\phi_p + \sum_{q=1}^{Q} \varepsilon_{i,t-q}\theta_q$$

$$\sigma^2_{it} = \exp\left(\eta_i + z_{it}\gamma\right) + \sum_{c=1}^{C} \sigma^2_{i,t-c}\lambda_c + \sum_{d=1}^{D} \varepsilon^2_{i,t-d}\xi_d$$

Dynamic and panel heteroskedasticity

Consider a Panel ARMA-GARCH model:

$$y_{it} = \mu_{it} + \varepsilon_{it} \qquad \varepsilon_{it} \sim f_{\mathcal{N}}\left(0, \sigma^2_{it}\right)$$

$$\mu_{it} = \alpha_i + x_{it}\beta + \sum_{p=1}^{P} y_{i,t-p}\phi_p + \sum_{q=1}^{Q} \varepsilon_{i,t-q}\theta_q$$

$$\sigma^2_{it} = \exp\left(\eta_i + z_{it}\gamma\right) + \sum_{c=1}^{C} \sigma^2_{i,t-c}\lambda_c + \sum_{d=1}^{D} \varepsilon^2_{i,t-d}\xi_d$$

This model includes two sources of panel heteroskedasticity:

- η_i, a random or fixed effect in the variance function, and
- $\boldsymbol{\lambda}_i$, $\boldsymbol{\xi}_i$, which make the variance time dependent

Dynamic and panel heteroskedasticity

We could simplify this model to an AR(1) with panel heteroskedasticity:

$$y_{it} = \mu_{it} + \varepsilon_{it} \qquad \varepsilon_{it} \sim f_{\mathcal{N}}\left(0, \sigma_i^2\right)$$

Dynamic and panel heteroskedasticity

We could simplify this model to an AR(1) with panel heteroskedasticity:

$$y_{it} = \mu_{it} + \varepsilon_{it} \qquad \varepsilon_{it} \sim f_{\mathcal{N}}\left(0, \sigma_i^2\right)$$

$$\mu_{it} = \alpha_i + x_{it}\beta + y_{i,t-p}\phi$$

$$\sigma_i^2 = \exp\left(\eta_i\right)$$

Only source of heteroskedasticity is now η_i:
panel heteroskedasticity, not dynamic heteroskedasticity

This is the model Beck & Katz advocate as a baseline for comparative politics.

They suggest estimating by LS, and then correcting the se's for the omission of η_i
(what they call "panel-corrected standard errors", PCSEs)

Note that treating α_i as a fixed effect
and, in the presence of unit roots, differencing y_{it}, would also be prudent

Panel-corrected standard errors

To calculate panel-corrected standard errors, we need to estimate the right variance-covariance matrix

We need an estimate of the variance-covariance matrix, $\hat{\Omega}$, which we can plug in to the GLS formula for the var-cov of regressors:

$$\mathrm{Cov}(\beta) = (\mathbf{X}'\mathbf{X})^{-1}(\mathbf{X}'\Omega\mathbf{X})(\mathbf{X}'\mathbf{X})^{-1}$$

In linear regression, $\Omega = \sigma^2 \mathrm{I}_N$, or the variance of the errors times an identity matrix

Panel-corrected standard errors

To calculate panel-corrected standard errors, we need to estimate the right variance-covariance matrix

We need an estimate of the variance-covariance matrix, $\hat{\Omega}$, which we can plug in to the GLS formula for the var-cov of regressors:

$$\mathrm{Cov}(\beta) = (\mathbf{X}'\mathbf{X})^{-1}(\mathbf{X}'\Omega\mathbf{X})(\mathbf{X}'\mathbf{X})^{-1}$$

In linear regression, $\Omega = \sigma^2 \mathrm{I}_N$, or the variance of the errors times an identity matrix

To make PCSEs, suppose Ω is $NT \times NT$ block-diagonal
with an $N \times N$ matrix Σ of contemporaneous covariances on diagonal

Panel-corrected standard errors

To calculate panel-corrected standard errors, we need to estimate the right variance-covariance matrix

We need an estimate of the variance-covariance matrix, $\hat{\Omega}$, which we can plug in to the GLS formula for the var-cov of regressors:

$$\mathrm{Cov}(\beta) = (\mathbf{X}'\mathbf{X})^{-1}(\mathbf{X}'\Omega\mathbf{X})(\mathbf{X}'\mathbf{X})^{-1}$$

In linear regression, $\Omega = \sigma^2 \mathrm{I}_N$, or the variance of the errors times an identity matrix

To make PCSEs, suppose Ω is $NT \times NT$ block-diagonal
with an $N \times N$ matrix Σ of contemporaneous covariances on diagonal

In other words, allow for unit or contemporaneous heteroskedaticity that stays the same over time

Visualizing this large matrix is tricky

Note that "$NT \times NT$ block-diagonal" means we are ordering the observations first by time, then by unit (reverse of our usual practice)

Panel-corrected standard errors

$\Omega_{NT\times NT} =$

$$
\begin{bmatrix}
\sigma^2_{\varepsilon_{1\cdot}} & \cdots & \sigma_{\varepsilon_{1\cdot},\varepsilon_{i\cdot}} & \cdots & \sigma_{\varepsilon_{1\cdot},\varepsilon_{N\cdot}} & \cdots\cdots & 0 & \cdots & 0 & \cdots & 0 & \cdots\cdots & 0 & \cdots & 0 & \cdots & 0 \\
\vdots & \ddots & \vdots & & \vdots & & \vdots & \ddots & \vdots & & \vdots & & \vdots & \ddots & \vdots & & \vdots \\
\sigma_{\varepsilon_{i\cdot},\varepsilon_{1\cdot}} & \cdots & \sigma^2_{\varepsilon_{i\cdot}} & \cdots & \sigma_{\varepsilon_{i\cdot},\varepsilon_{N\cdot}} & \cdots\cdots & 0 & \cdots & 0 & \cdots & 0 & \cdots\cdots & 0 & \cdots & 0 & \cdots & 0 \\
\vdots & & \vdots & \ddots & \vdots & & \vdots & & \vdots & \ddots & \vdots & & \vdots & & \vdots & \ddots & \vdots \\
\sigma_{\varepsilon_{N\cdot},\varepsilon_{1\cdot}} & \cdots & \sigma_{\varepsilon_{N\cdot},\varepsilon_{i\cdot}} & \cdots & \sigma^2_{\varepsilon_{N\cdot}} & \cdots\cdots & 0 & \cdots & 0 & \cdots & 0 & \cdots\cdots & 0 & \cdots & 0 & \cdots & 0 \\
\vdots & & \vdots & & \vdots & \ddots & \vdots & & \vdots & & \vdots & & \vdots & & \vdots & & \vdots \\
\vdots & & \vdots & & \vdots & \ddots & \vdots & & \vdots & & \vdots & & \vdots & & \vdots & & \vdots \\
0 & \cdots & 0 & \cdots & 0 & \cdots\cdots & \sigma^2_{\varepsilon_{1\cdot}} & \cdots & \sigma_{\varepsilon_{1\cdot},\varepsilon_{i\cdot}} & \cdots & \sigma_{\varepsilon_{1\cdot},\varepsilon_{N\cdot}} & \cdots\cdots & 0 & \cdots & 0 & \cdots & 0 \\
\vdots & \ddots & \vdots & & \vdots & & \vdots & \ddots & \vdots & & \vdots & & \vdots & \ddots & \vdots & & \vdots \\
0 & \cdots & 0 & \cdots & 0 & \cdots\cdots & \sigma_{\varepsilon_{i\cdot},\varepsilon_{1\cdot}} & \cdots & \sigma^2_{\varepsilon_{i\cdot}} & \cdots & \sigma_{\varepsilon_{i\cdot},\varepsilon_{N\cdot}} & \cdots\cdots & 0 & \cdots & 0 & \cdots & 0 \\
\vdots & & \vdots & \ddots & \vdots & & \vdots & & \vdots & \ddots & \vdots & & \vdots & & \vdots & \ddots & \vdots \\
0 & \cdots & 0 & \cdots & 0 & \cdots\cdots & \sigma_{\varepsilon_{N\cdot},\varepsilon_{1\cdot}} & \cdots & \sigma_{\varepsilon_{N\cdot},\varepsilon_{i\cdot}} & \cdots & \sigma^2_{\varepsilon_{N\cdot}} & \cdots\cdots & 0 & \cdots & 0 & \cdots & 0 \\
\vdots & & \vdots & & \vdots & & \vdots & & \vdots & & \vdots & \ddots & \vdots & & \vdots & & \vdots \\
\vdots & & \vdots & & \vdots & & \vdots & & \vdots & & \vdots & \ddots & \vdots & & \vdots & & \vdots \\
0 & \cdots & 0 & \cdots & 0 & \cdots\cdots & 0 & \cdots & 0 & \cdots & 0 & \cdots\cdots & \sigma^2_{\varepsilon_{1\cdot}} & \cdots & \sigma_{\varepsilon_{1\cdot},\varepsilon_{i\cdot}} & \cdots & \sigma_{\varepsilon_{1\cdot},\varepsilon_{N\cdot}} \\
\vdots & \ddots & \vdots & & \vdots & & \vdots & \ddots & \vdots & & \vdots & & \vdots & \ddots & \vdots & & \vdots \\
0 & \cdots & 0 & \cdots & 0 & \cdots\cdots & 0 & \cdots & 0 & \cdots & 0 & \cdots\cdots & \sigma_{\varepsilon_{i\cdot},\varepsilon_{1\cdot}} & \cdots & \sigma^2_{\varepsilon_{i\cdot}} & \cdots & \sigma_{\varepsilon_{i\cdot},\varepsilon_{N\cdot}} \\
\vdots & & \vdots & \ddots & \vdots & & \vdots & & \vdots & \ddots & \vdots & & \vdots & & \vdots & \ddots & \vdots \\
0 & \cdots & 0 & \cdots & 0 & \cdots\cdots & 0 & \cdots & 0 & \cdots & 0 & \cdots\cdots & \sigma_{\varepsilon_{N\cdot},\varepsilon_{1\cdot}} & \cdots & \sigma_{\varepsilon_{N\cdot},\varepsilon_{i\cdot}} & \cdots & \sigma^2_{\varepsilon_{N\cdot}}
\end{bmatrix}
$$

Panel-corrected standard errors

Instead, suppose Ω is $NT \times NT$ block-diagonal
with an $N \times N$ matrix Σ of contemporaneous covariances on diagonal

Panel-corrected standard errors

Instead, suppose Ω is $NT \times NT$ block-diagonal
with an $N \times N$ matrix Σ of contemporaneous covariances on diagonal

In other words, allow for unit or contemporaneous heteroskedaticity that stays the same over time

Beck and Katz (1995) show can estimate Σ using LS residuals $e_{i,t}$:

$$\hat{\Sigma}_{i,j} = \sum_{t=1}^{T} \frac{e_{i,t} e_{j,t}}{T}$$

And then use $\hat{\Sigma}$ to calculate the covariance matrix.

Panel-corrected standard errors

Monte Carlo experiments show panel-coorrected standard errors are "correct" unless contemporaneous correlation is very high.

(Note: alternative is to estimate random effects in variance by ML.
Tends to produce similar results.)

Beck and Katz suggest using LS with PCSEs and lagged DVs as a baseline model

Most practioners think fixed effects should also be used

Most important: getting the right lags structure & including FEs where appropriate

PCSEs, or choice of estimation strategy is a much smaller concern

In `R`, package `pcse` will calculate PCSEs for a linear regression

Getting them for ML models would take (considerable) extra work