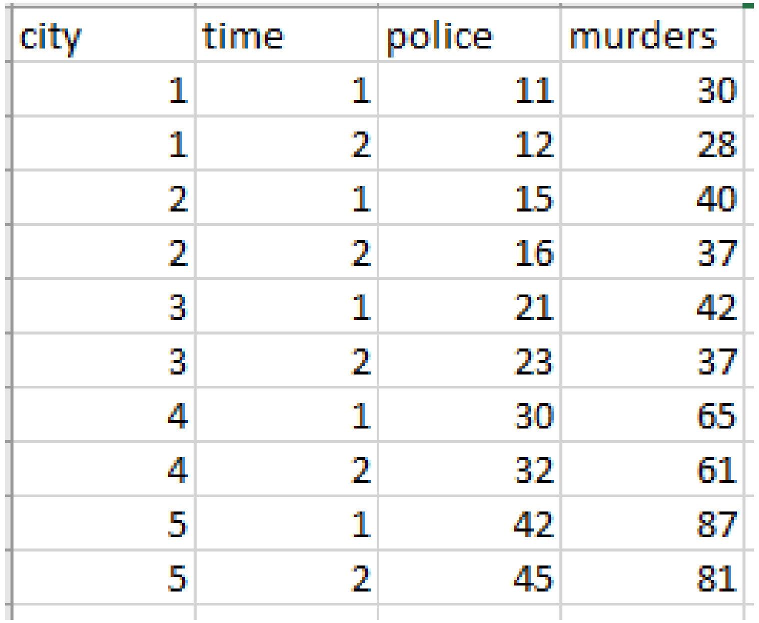
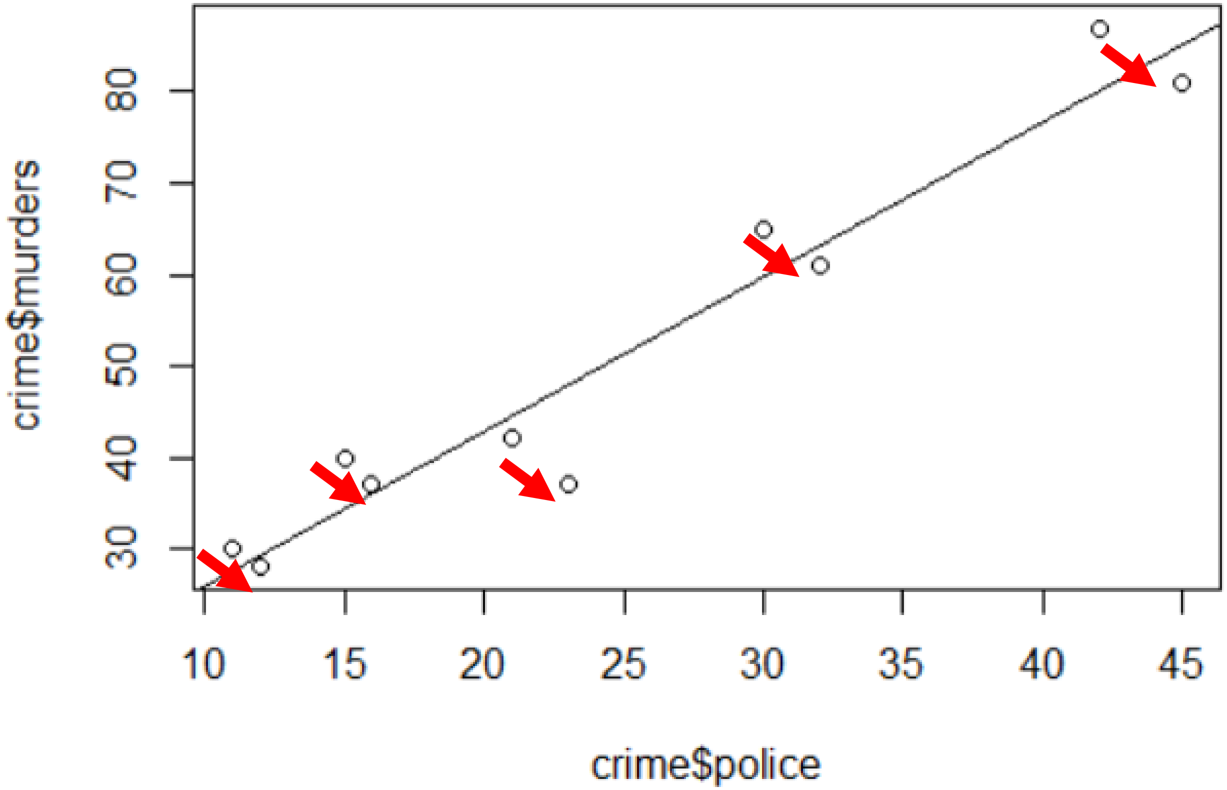
The intuition behind panel regression

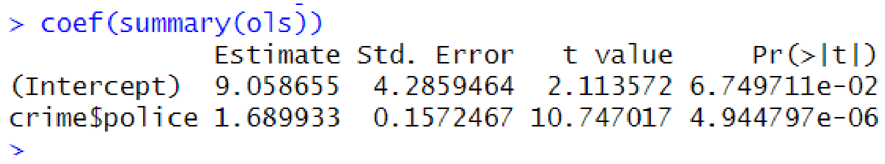
Prof. Alex COAD



city	time	police	murders
1	1	11	30
1	2	12	28
2	1	15	40
2	2	16	37
3	1	21	42
3	2	23	37
4	1	30	65
4	2	32	61
5	1	42	87
5	2	45	81

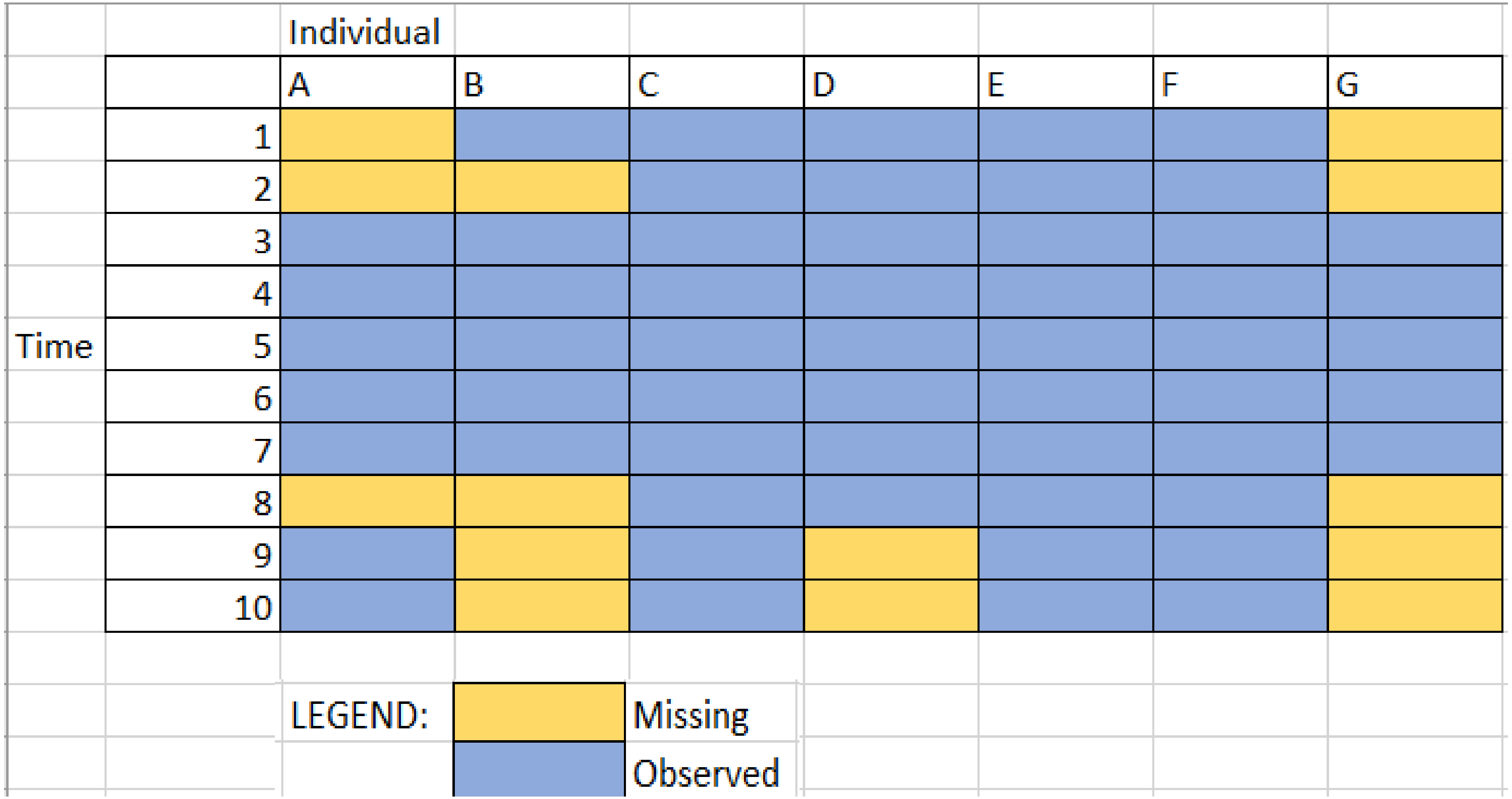
- Naïve OLS: $Murders_i = a + bPolice_i + e_i$
- In each city, adding more police reduces the number of murders
- So, how can the relationship be positive?
- UNOBSERVED HETEROGENEITY: large cities have more police and more murders
  - Omitted Variable Bias (OVB): control for city size?
  - Focus on differences or changes rather than levels?
  - Take into account the different starting points?





```
> coef(summary(ols))
              Estimate Std. Error   t value     Pr(>|t|)
(Intercept)  9.058655  4.2859464  2.113572 6.749711e-02
crime$police 1.689933  0.1572467 10.747017 4.944797e-06
>
```

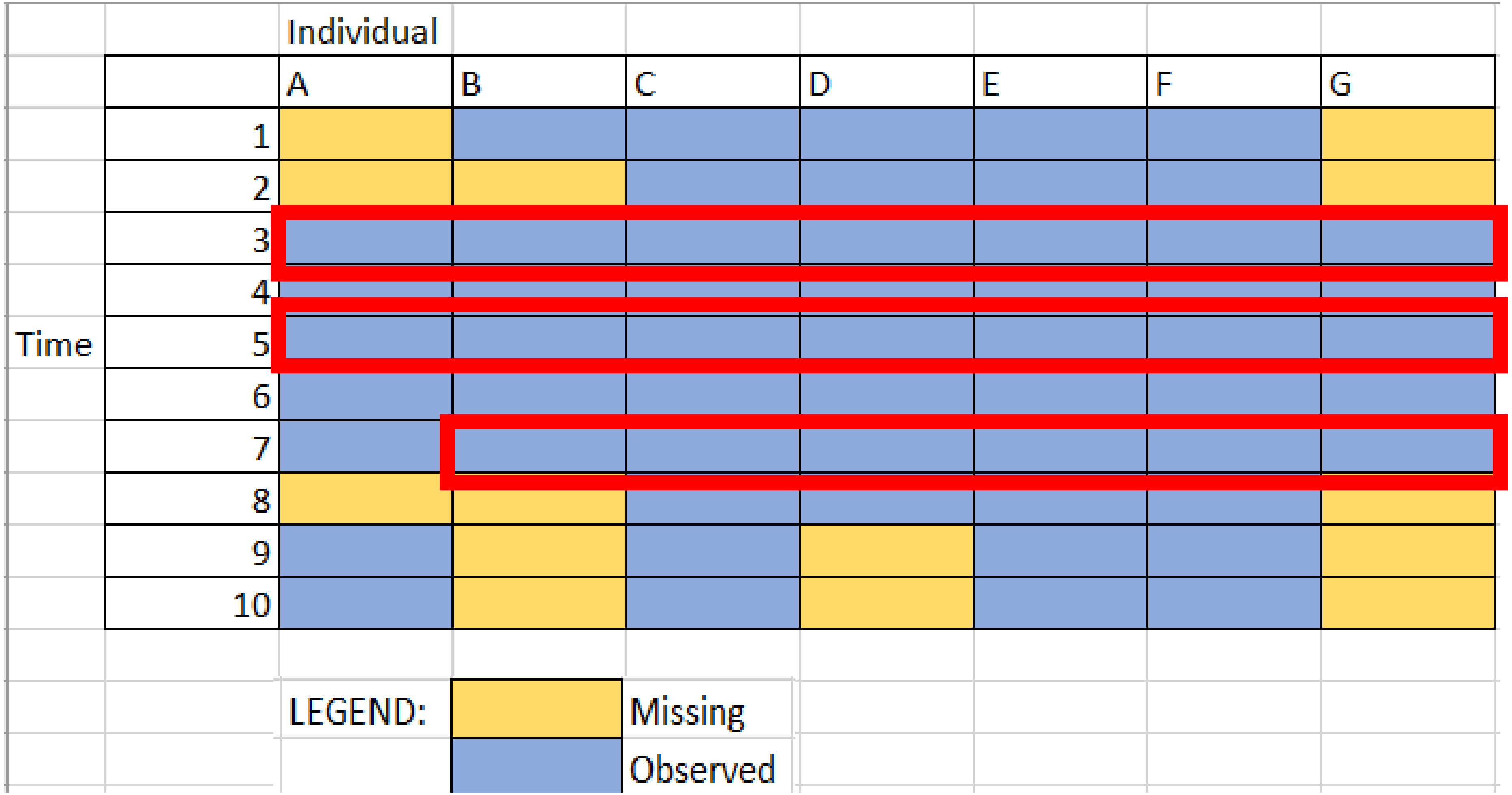
DATA STRUCTURE
Individuals A:G; Time 1:10



		Individual						
		A	B	C	D	E	F	G
	1	🟨	🟦	🟦	🟦	🟦	🟦	🟨
	2	🟨	🟨	🟦	🟦	🟦	🟦	🟨
	3	🟦	🟦	🟦	🟦	🟦	🟦	🟦
	4	🟦	🟦	🟦	🟦	🟦	🟦	🟦
Time	5	🟦	🟦	🟦	🟦	🟦	🟦	🟦
	6	🟦	🟦	🟦	🟦	🟦	🟦	🟦
	7	🟦	🟦	🟦	🟦	🟦	🟦	🟦
	8	🟨	🟨	🟦	🟦	🟦	🟦	🟨
	9	🟦	🟨	🟦	🟨	🟦	🟦	🟨
	10	🟦	🟨	🟦	🟨	🟦	🟦	🟨

LEGEND:	🟨	Missing
	🟦	Observed

Cross-section
1 time period, several entities

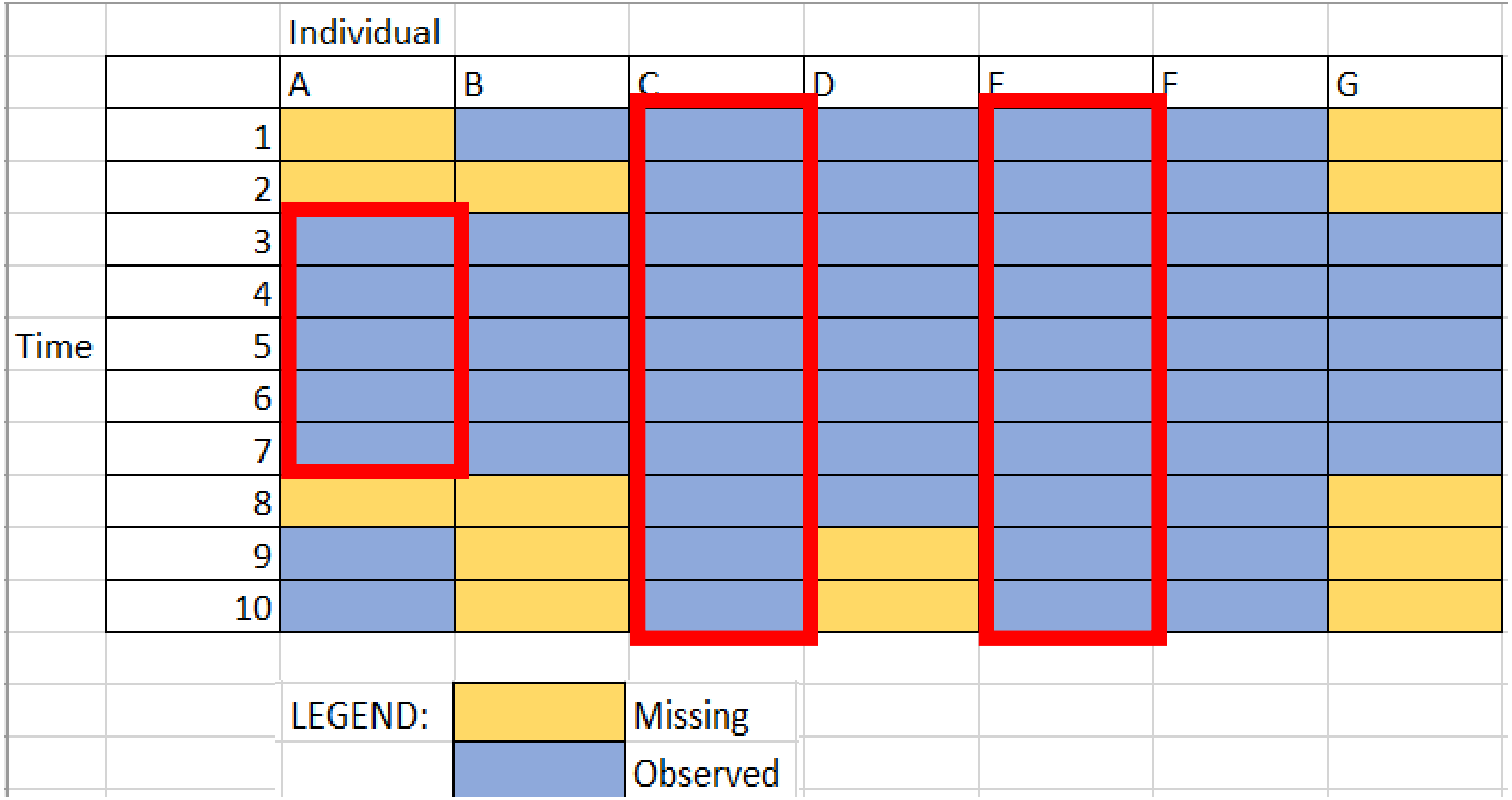


		Individual						
		A	B	C	D	E	F	G
	1	🟨	🟦	🟦	🟦	🟦	🟦	🟨
	2	🟨	🟨	🟦	🟦	🟦	🟦	🟨
	3	🟦	🟦	🟦	🟦	🟦	🟦	🟦
	4	🟦	🟦	🟦	🟦	🟦	🟦	🟦
Time	5	🟦	🟦	🟦	🟦	🟦	🟦	🟦
	6	🟦	🟦	🟦	🟦	🟦	🟦	🟦
	7	🟦	🟦	🟦	🟦	🟦	🟦	🟦
	8	🟨	🟨	🟦	🟦	🟦	🟦	🟨
	9	🟦	🟨	🟦	🟨	🟦	🟦	🟨
	10	🟦	🟨	🟦	🟨	🟦	🟦	🟨

LEGEND: 🟨 Missing

🟦 Observed

Time series: several time periods, one entity

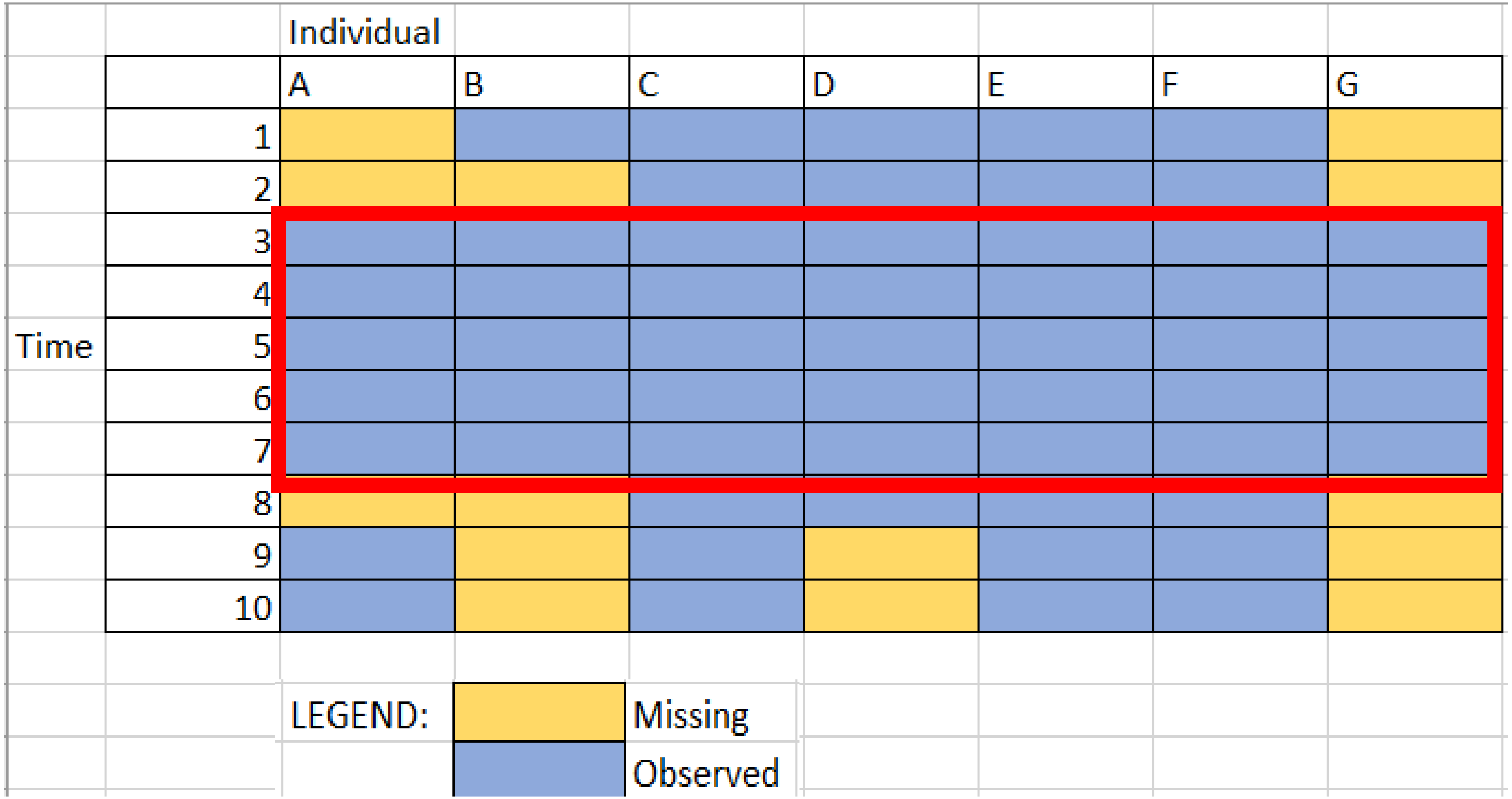


		Individual						
		A	B	C	D	E	F	G
	1	🟨	🟦	🟦	🟦	🟦	🟦	🟨
	2	🟨	🟨	🟦	🟦	🟦	🟦	🟨
	3	🟦	🟦	🟦	🟦	🟦	🟦	🟦
	4	🟦	🟦	🟦	🟦	🟦	🟦	🟦
Time	5	🟦	🟦	🟦	🟦	🟦	🟦	🟦
	6	🟦	🟦	🟦	🟦	🟦	🟦	🟦
	7	🟦	🟦	🟦	🟦	🟦	🟦	🟦
	8	🟨	🟨	🟦	🟦	🟦	🟦	🟨
	9	🟦	🟨	🟦	🟨	🟦	🟦	🟨
	10	🟦	🟨	🟦	🟨	🟦	🟦	🟨

LEGEND:	🟨	Missing
	🟦	Observed

Balanced panel

Several time periods, several entities, no missing observations

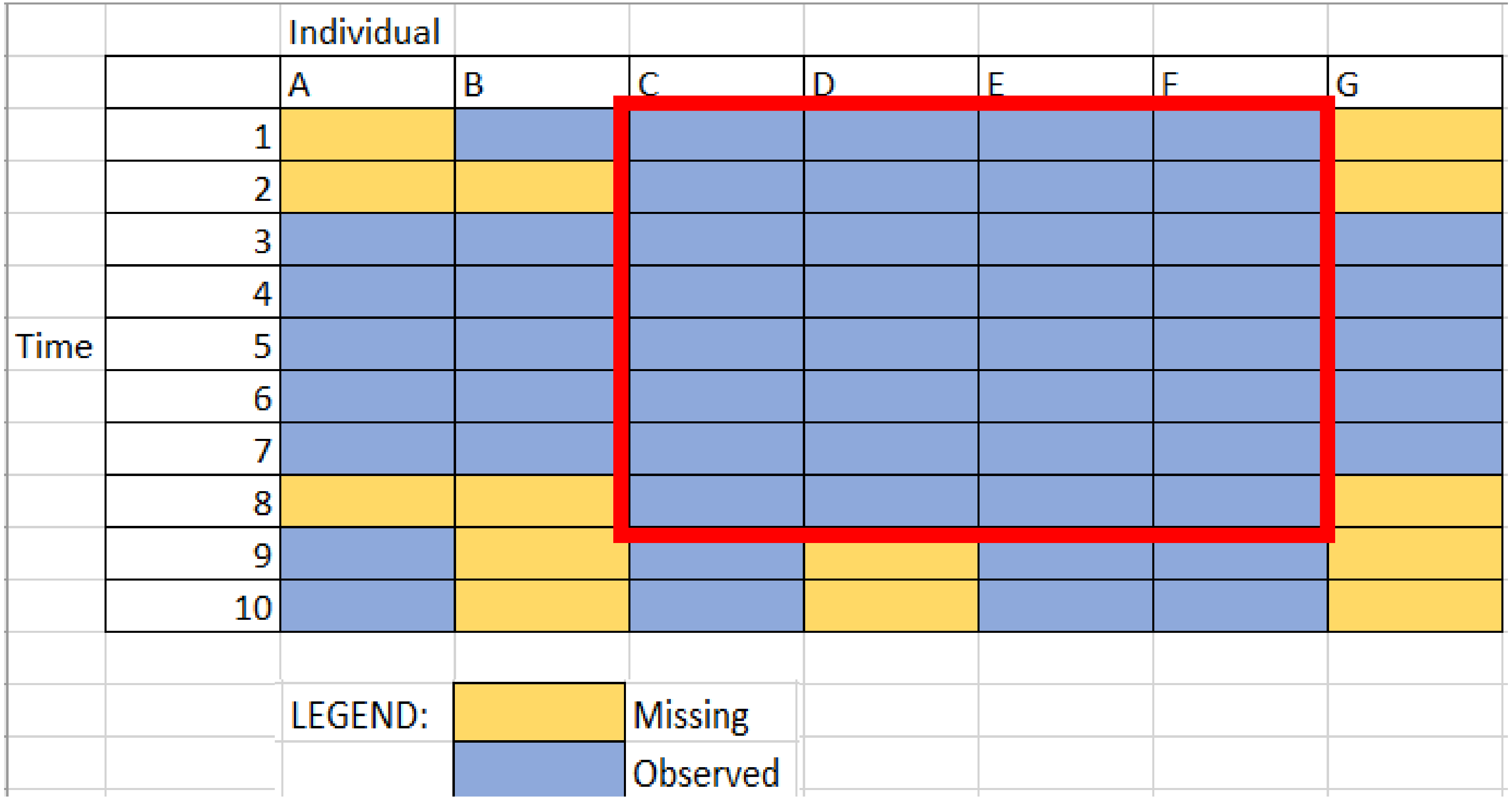


		Individual						
		A	B	C	D	E	F	G
	1	🟨	🟦	🟦	🟦	🟦	🟦	🟨
	2	🟨	🟨	🟦	🟦	🟦	🟦	🟨
	3	🟦	🟦	🟦	🟦	🟦	🟦	🟦
	4	🟦	🟦	🟦	🟦	🟦	🟦	🟦
Time	5	🟦	🟦	🟦	🟦	🟦	🟦	🟦
	6	🟦	🟦	🟦	🟦	🟦	🟦	🟦
	7	🟦	🟦	🟦	🟦	🟦	🟦	🟦
	8	🟨	🟨	🟦	🟦	🟦	🟦	🟨
	9	🟦	🟨	🟦	🟨	🟦	🟦	🟨
	10	🟦	🟨	🟦	🟨	🟦	🟦	🟨

LEGEND:	🟨	Missing
	🟦	Observed

Balanced panel

Several time periods, several entities, no missing observations

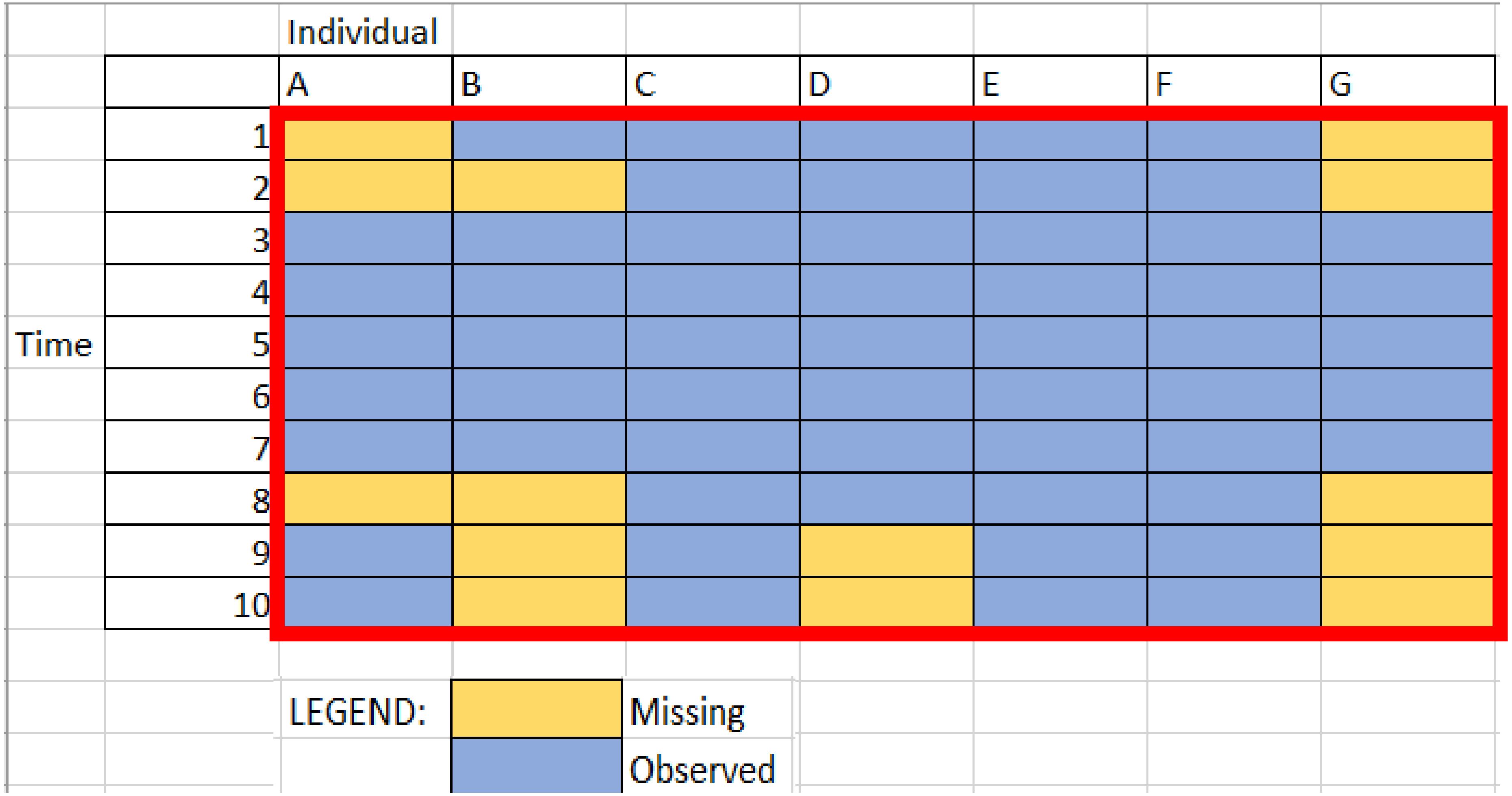


		Individual						
		A	B	C	D	E	F	G
	1	🟨	🟦	🟦	🟦	🟦	🟦	🟨
	2	🟨	🟨	🟦	🟦	🟦	🟦	🟨
	3	🟦	🟦	🟦	🟦	🟦	🟦	🟦
	4	🟦	🟦	🟦	🟦	🟦	🟦	🟦
Time	5	🟦	🟦	🟦	🟦	🟦	🟦	🟦
	6	🟦	🟦	🟦	🟦	🟦	🟦	🟦
	7	🟦	🟦	🟦	🟦	🟦	🟦	🟦
	8	🟨	🟨	🟦	🟦	🟦	🟦	🟨
	9	🟦	🟨	🟦	🟨	🟦	🟦	🟨
	10	🟦	🟨	🟦	🟨	🟦	🟦	🟨

LEGEND:	🟨	Missing
	🟦	Observed

Unbalanced panel

Several time periods, several entities, with missing observations



		Individual						
		A	B	C	D	E	F	G
	1	🟨	🟦	🟦	🟦	🟦	🟦	🟨
	2	🟨	🟨	🟦	🟦	🟦	🟦	🟨
	3	🟦	🟦	🟦	🟦	🟦	🟦	🟦
	4	🟦	🟦	🟦	🟦	🟦	🟦	🟦
Time	5	🟦	🟦	🟦	🟦	🟦	🟦	🟦
	6	🟦	🟦	🟦	🟦	🟦	🟦	🟦
	7	🟦	🟦	🟦	🟦	🟦	🟦	🟦
	8	🟨	🟨	🟦	🟦	🟦	🟦	🟨
	9	🟦	🟨	🟦	🟨	🟦	🟦	🟨
	10	🟦	🟨	🟦	🟨	🟦	🟦	🟨

LEGEND:	🟨	Missing
	🟦	Observed

Panel regression

- $y_{it} = \alpha + \beta x_{it} + \varepsilon_{it}$
- But: there is a time-invariant component in the error term:
- $\varepsilon_{it} = (\eta_i + \nu_{it})$
- Hence: $y_{it} = \alpha + \beta x_{it} + (\eta_i + \nu_{it})$

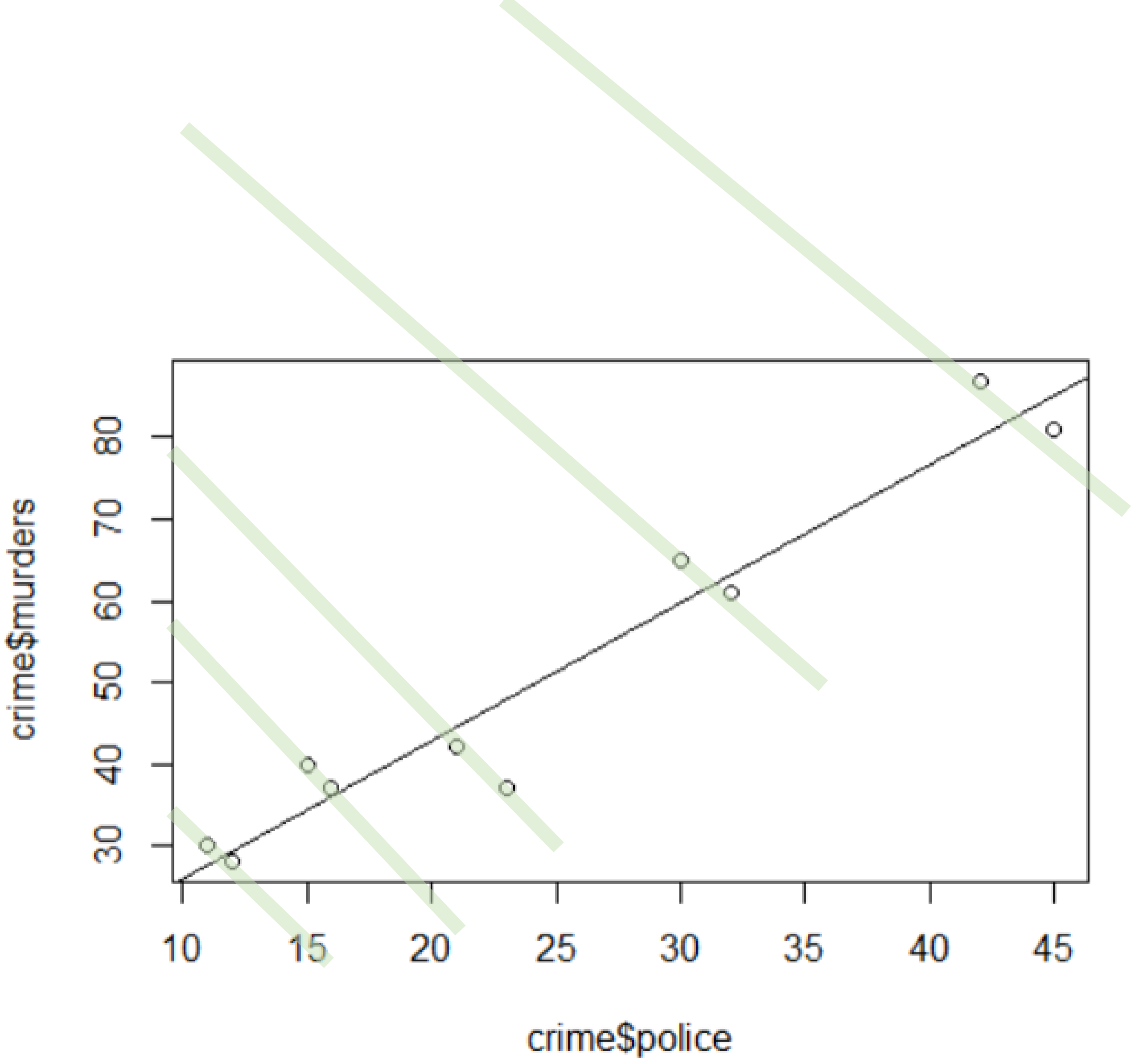
- η_i is usually not directly observed
  - time-invariant
  - individual-specific

Fixed-effects (a.k.a. “within”) regression

- Subtract the individual’s average over time to every variable
  - “time-demeaning”
  - Estimation in deviations-from-means is called the “within estimator”
- $y_{it} - \bar{y}_i = (x_{it} - \bar{x}_i)\beta + (\mathrm{v}_{it} - \bar{\mathrm{v}}_i)$
- We drop the terms for α and η_i:
- $\alpha - \alpha = 0$
- $\eta_i - \eta_i = 0$

- This is similar to estimating: $y_{it} = \alpha_i + x_{it}\beta + \mathrm{v}_{it}$

- FE is equivalent to estimating: $y_{it} = \alpha_i + x_{it}\beta + \mathrm{v}_{it}$
  - I.e. leaving the individual intercepts free to vary
  - Considering them as parameters to be estimated



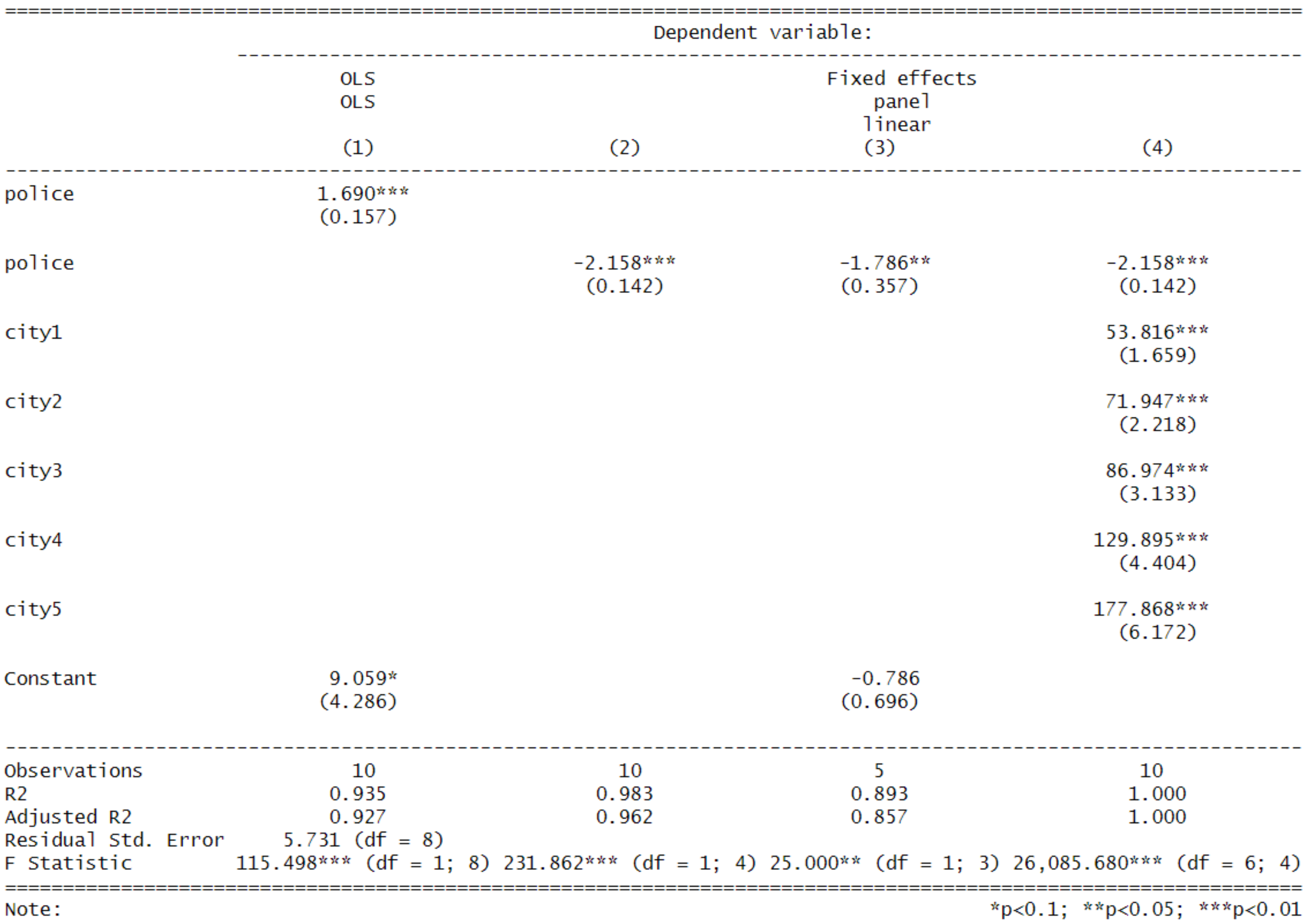
Differencing

- η_i is removed by first-differencing the data
  - Taking lags and subtracting
- $y_{it} = \alpha + \beta x_{it} + (\eta_i + \nu_{it})$
- $y_{i,t-1} = \alpha + \beta x_{i,t-1} + (\eta_i + \nu_{i,t-1})$

- $y_{it} - y_{i,t-1} = \boxed{\alpha - \alpha} + \beta x_{it} - \beta x_{i,t-1} + (\boxed{\eta_i} + \nu_{it}) - (\boxed{\eta_i} + \nu_{i,t-1})$
- $\Delta y_{it} = \beta \Delta x_{it} + \Delta \nu_{it}$

- This is the first difference (FD) estimator
  - Can be consistently estimated by pooled OLS

Crime example revisited



```
=========================================================================================================
                                                Dependent variable:
                     ------------------------------------------------------------------------------------
                              OLS                          Fixed effects
                              OLS                              panel
                                                              linear
                              (1)              (2)              (3)                      (4)
---------------------------------------------------------------------------------------------------------
police                     1.690***
                           (0.157)

police                                      -2.158***         -1.786**                -2.158***
                                             (0.142)          (0.357)                  (0.142)

city1                                                                                  53.816***
                                                                                        (1.659)

city2                                                                                  71.947***
                                                                                        (2.218)

city3                                                                                  86.974***
                                                                                        (3.133)

city4                                                                                 129.895***
                                                                                        (4.404)

city5                                                                                 177.868***
                                                                                        (6.172)

Constant                    9.059*                              -0.786
                           (4.286)                              (0.696)

---------------------------------------------------------------------------------------------------------
Observations                  10               10                 5                        10
R2                          0.935            0.983              0.893                    1.000
Adjusted R2                 0.927            0.962              0.857                    1.000
Residual Std. Error     5.731 (df = 8)
F Statistic         115.498*** (df = 1; 8) 231.862*** (df = 1; 4) 25.000** (df = 1; 3) 26,085.680*** (df = 6; 4)
=========================================================================================================
Note:                                                                        *p<0.1; **p<0.05; ***p<0.01
```

- (1): OLS
  - Wrong, because doesn't take into account unobserved heterogeneity
- (2):FE
  - No constant term
- (3) First differences
  - Similar but not identical to FE
  - Only 5 observations
- (4) LSDV
  - Least Squares Dummy Variable
  - Dummy for each city added
  - Exactly the same coefficient as FE, but more output

FE vs differences

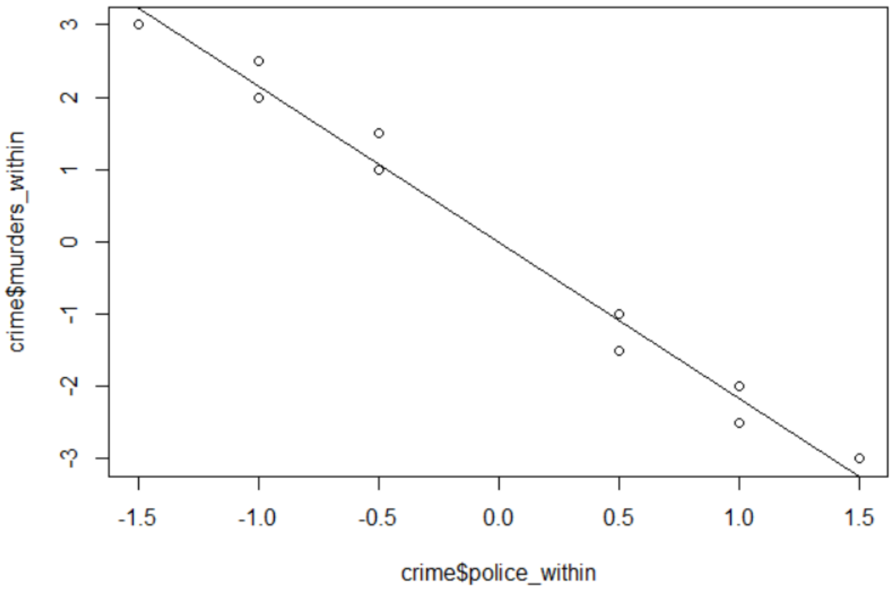
- $y_{it} = \alpha + \beta x_{it} + (\eta_i + \nu_{it})$

- Fixed effects (FE): Estimations of differences-in-means:
- $y_{it} - \bar{y}_i = (x_{it} - \bar{x}_i)\beta + (\nu_{it} - \bar{\nu}_i)$

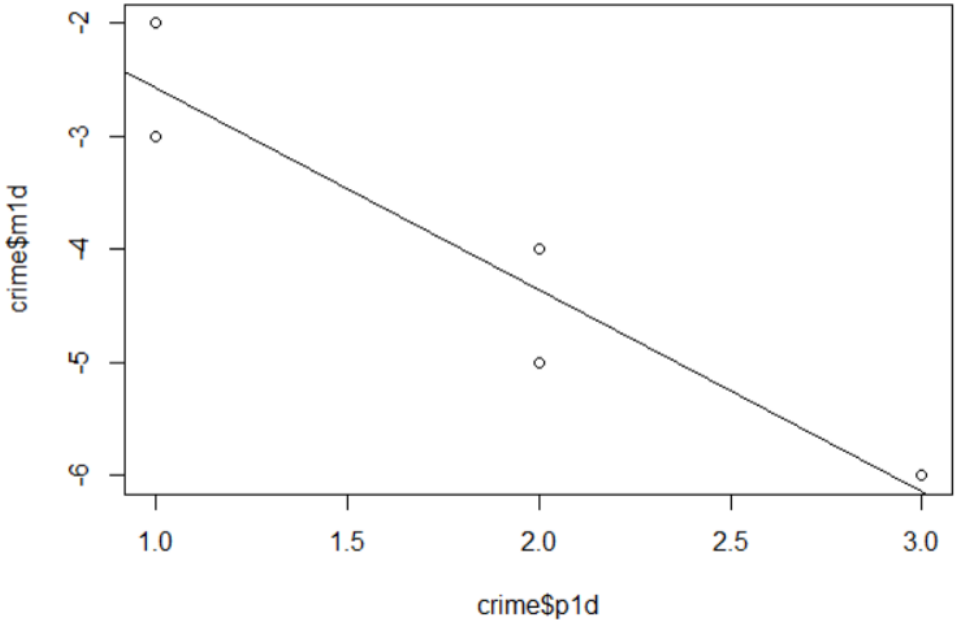
- Taking differences:
- $y_{it} - y_{i,t-1} = \alpha - \alpha + \beta x_{i,t} - \beta x_{i,t-1} + (\eta_i + \nu_{it}) - \left(\eta_i + \nu_{i,t-1}\right)$
- $\Delta y_{it} = \beta \Delta x_{it} + \Delta \nu_{it}$

FE and FD transform the data differently

- Fixed effects



- First differences



References

- Croissant, Y., & Millo, G. (2019). Panel data econometrics with R. John Wiley and Sons, Incorporated.
- Cunningham (2021). Causal inference: the mixtape. Yale University Press. Free to read online: https://scunning.com/mixtape.html
- Huntington-Klein, N. (2021). The effect: An introduction to research design and causality. Chapman and Hall/CRC (New York, USA). DOI https://doi.org/10.1201/9781003226055 Free to read online here: https://theeffectbook.net/ch-FixedEffects.html