

# Time series

EC 421, Set 7

---

Edward Rubin

06 February 2020

# Prologue

# Schedule

## Last Time

Asymptotics, probability limits, and consistency

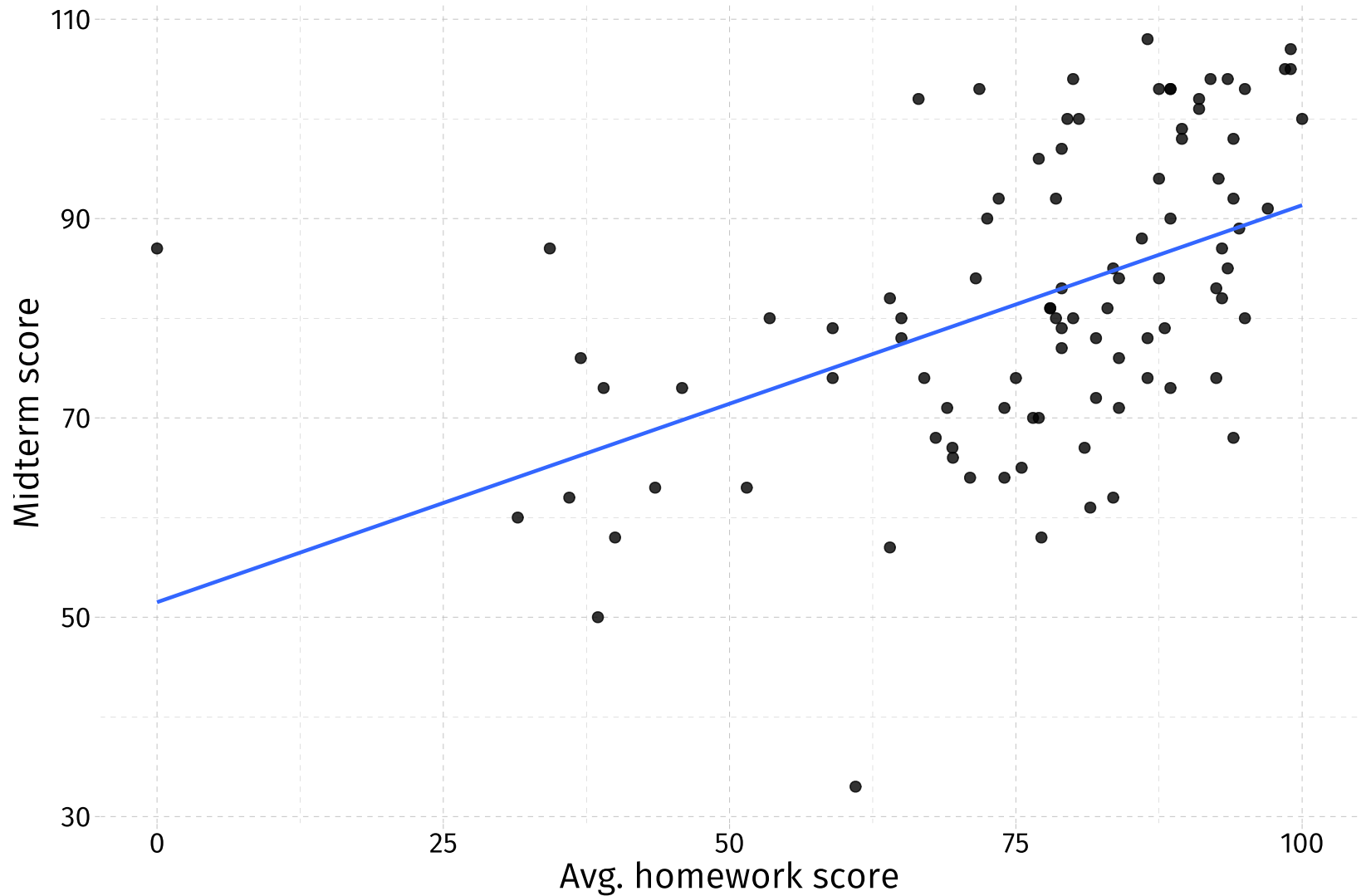
## Today

- Time series

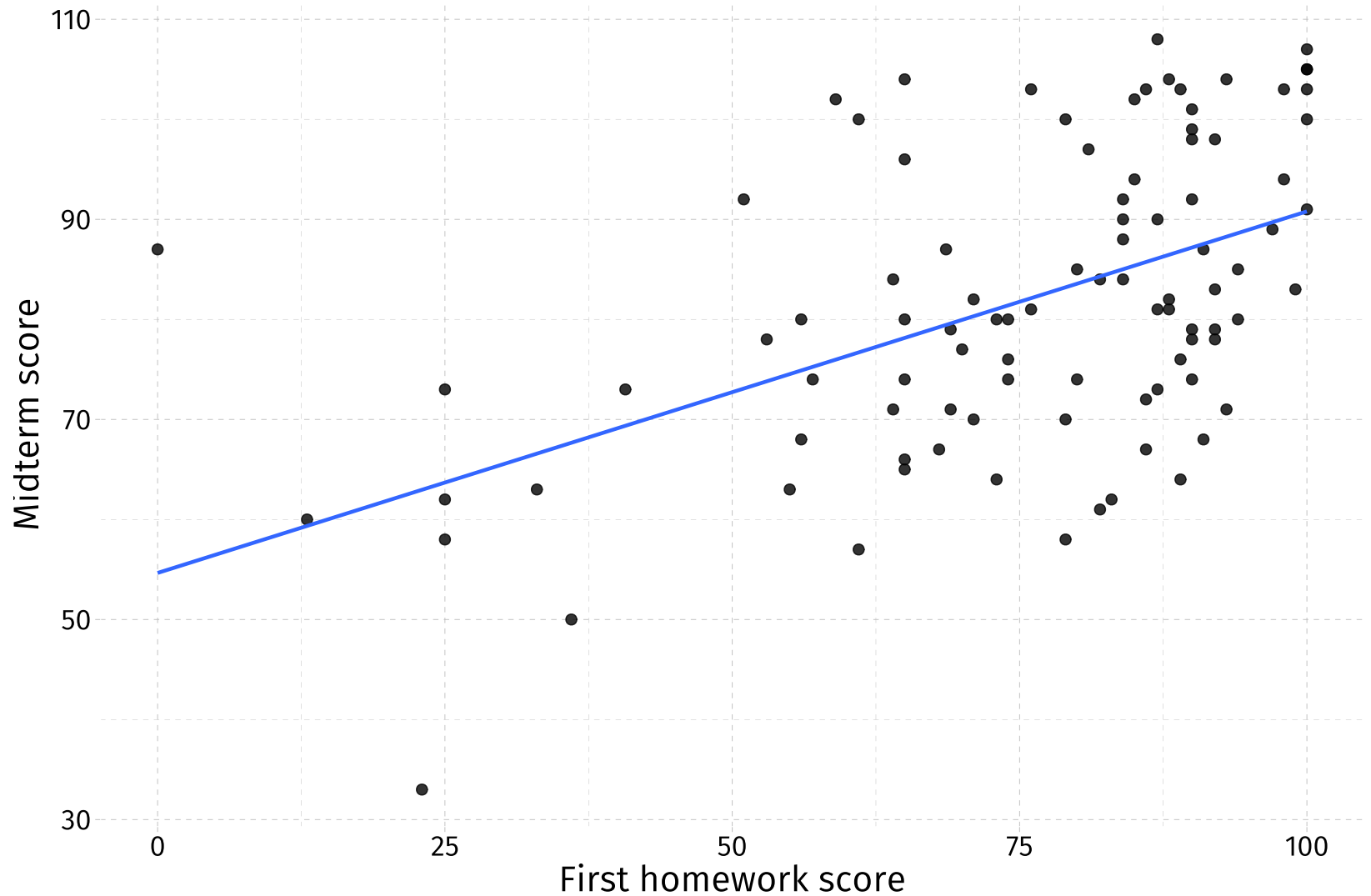
## Upcoming

- Assignment **due Saturday by midnight**
- Midterm **next Thursday** (2/13)

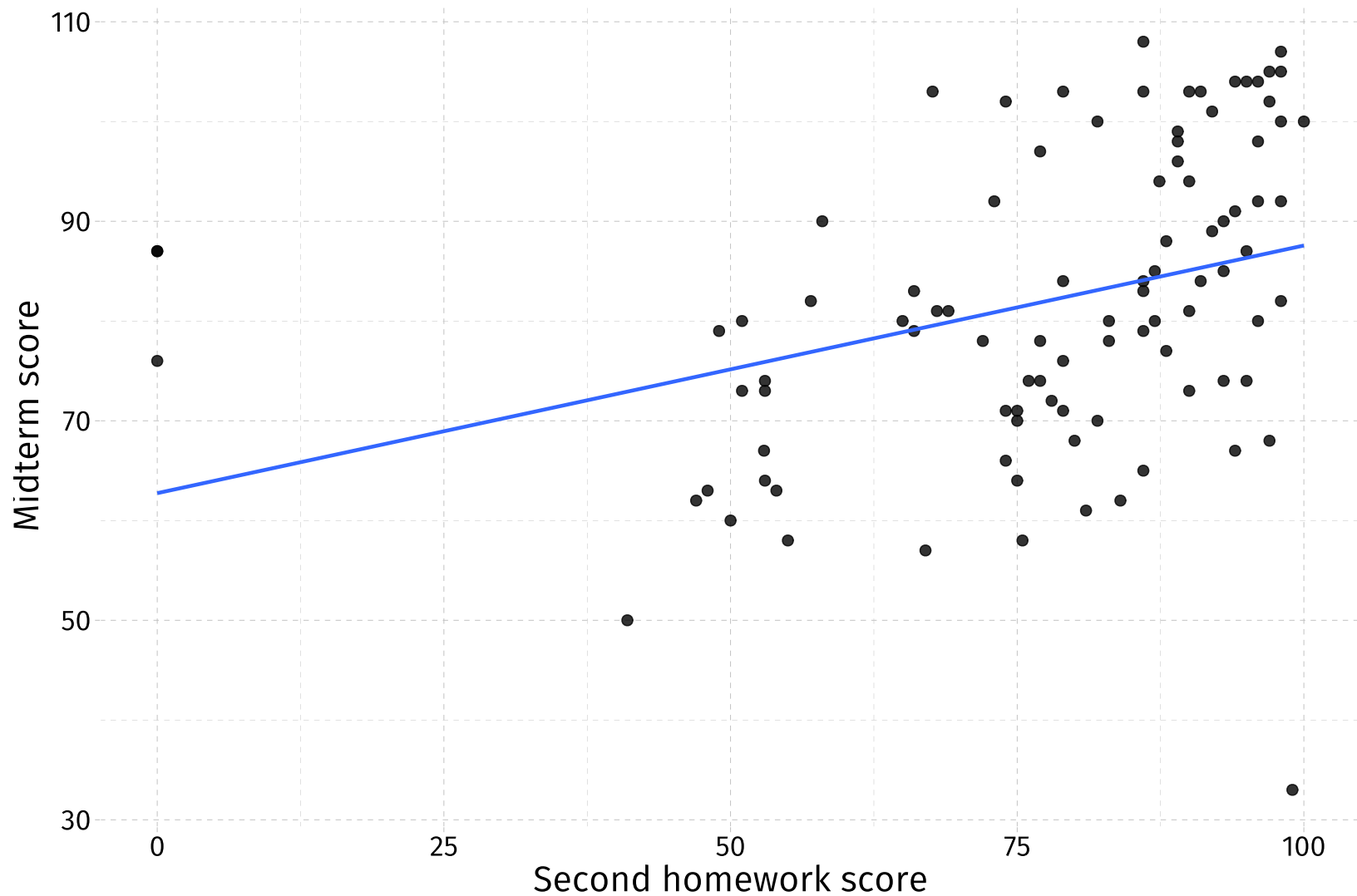
# Midterm,



# Midterm,



# Midterm,



## About our class

1. EC 421 is a **hard class**.
2. EC 421 requires **more math/theory** than most other classes.
3. This **theory is important**—why/when you can trust OLS/regression.
4. With all of this theory, we get **fewer traditional examples**. Proofs and simulations *are* our examples.
5. Midterm will **mix theory, intuition, and application**.

# Example questions

## Theory

In our proof of the consistency of the OLS estimator for  $\beta_1$  (for simple linear regression), we got to the point where we had

$$\text{plim } \hat{\beta}_1 = \beta_1 + \frac{\text{Cov}(x_1, u)}{\text{Var}(x_1)} \quad (1)$$

What does the right-hand side of (1) need to simplify to for the OLS estimator  $\hat{\beta}_1$  to be consistent?



# Example questions

## Intuition

We've shown that omitted variables can cause OLS to be biased and inconsistent.

1. What are the two requirements for an omitted variable to cause bias/inconsistency in OLS?
2. Provide an example of a regression that would suffer from omitted variable bias. Explain why it could be biased.
3. Does leaving out a variable from a regression **always** bias OLS? Explain your answer.

# Example questions

## Application

Your friend is concerned about heteroskedasticity in the regression below.

$$y_i = \hat{\beta}_0 + \hat{\beta}_1 x_{1i} + \hat{\beta}_2 x_{2i} + e_i \quad (2)$$

$$e_i^2 = \hat{\beta}_0 + \hat{\beta}_1 x_{1i} + \hat{\beta}_2 x_{2i} + v_i \quad (3)$$

$$e_i^2 = \hat{\beta}_0 + \hat{\beta}_1 x_{1i} + \hat{\beta}_2 x_{2i} + \hat{\beta}_3 x_{1i}^2 + \hat{\beta}_4 x_{2i}^2 + \hat{\beta}_5 x_{1i} x_{2i} + w_i \quad (4)$$

Because you are such a great friend, you estimated regressions (3) and (4).

The regression in (3) has an  $R^2$  of 0.20, and the regression in (4) has an  $R^2$  of 0.30. You have 100 observations.

1. Calculate the Breusch-Pagan test statistic testing heterosk. in (1).
2. The critical value for the Breusch-Pagan test is 6. Finish the B-P test (state your hypotheses; determine your conclusion).

# Asymptotics and consistency

## *Review*

# Asymptotics and consistency

## Review

1. Compare/contrast the concepts *expected value* and *probability limit*.
2. What does it mean if the estimator  $\hat{\theta}$  is consistent for  $\theta$ ?
3. What is required for an omitted variable to bias OLS estimates of  $\beta_j$ ?
4. Does omitted-variable bias affect the consistency of OLS for  $\beta_j$ ?
5. What can we know about the direction of omitted-variable bias?
6. How does measurement error in an explanatory variable affect the OLS estimate for that variable's effect on the outcome variable?
7. How does measurement error in an outcome variable affect OLS?

# Time-series data

# Time-series data

## Introduction

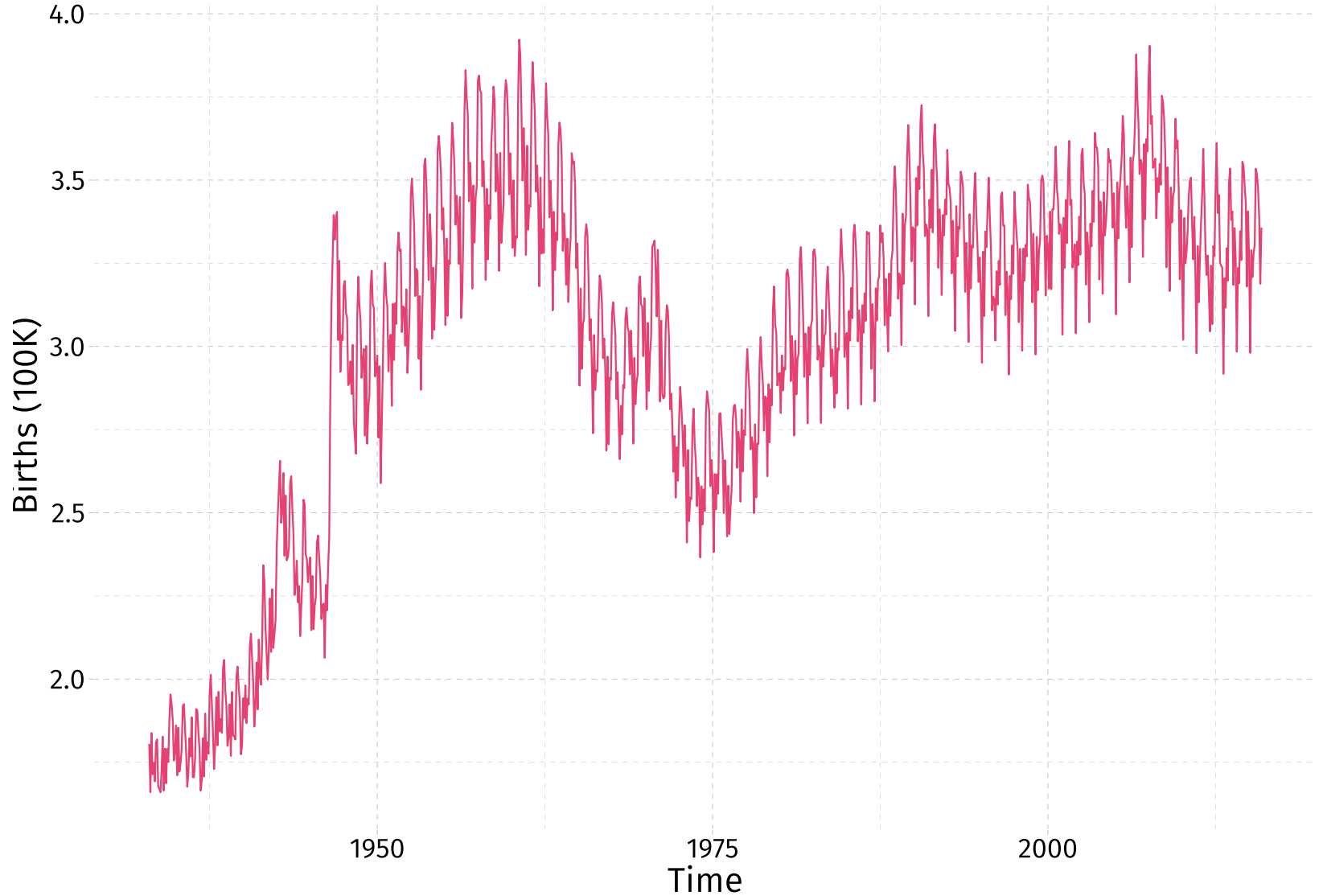
Up to this point, we focused on **cross-sectional data**.

- Sampled *across* a population (e.g., people, counties, countries).
- Sampled at *one moment* in time (e.g., Jan. 1, 2015).
- We had  $n$  individuals, each indexed  $i$  in  $\{1, \dots, n\}$ .

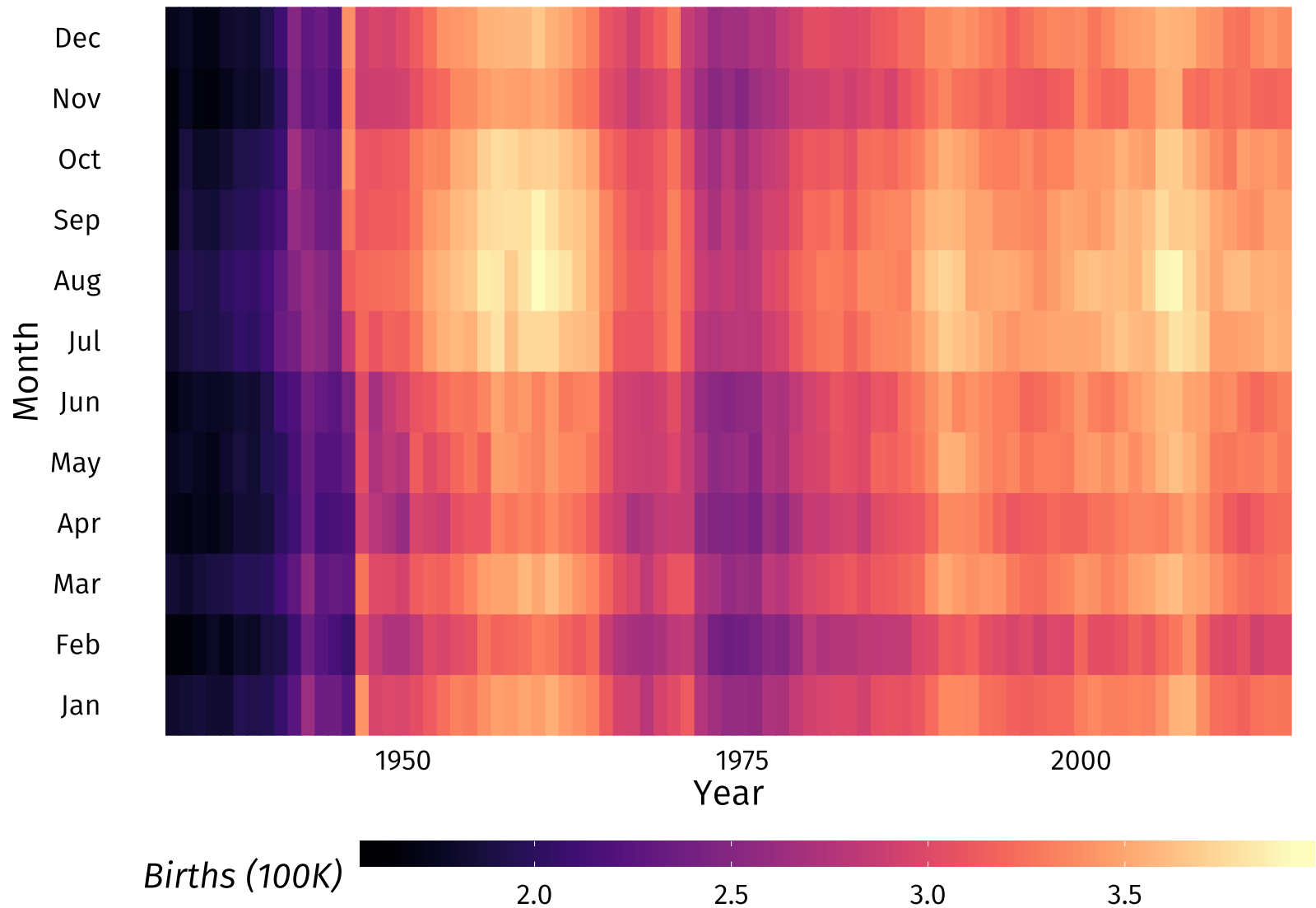
Today, we focus on a different type of data: **time-series data**.

- Sampled within **one unit/individual** (e.g., Oregon).
- Observe **multiple times** for the same unit (e.g., Oregon: 1990–2020).
- We have  **$T$  time periods**, each indexed  $t$  in  $\{1, \dots, T\}$ .

## US monthly births, 1933–2015: Classic time-series graph

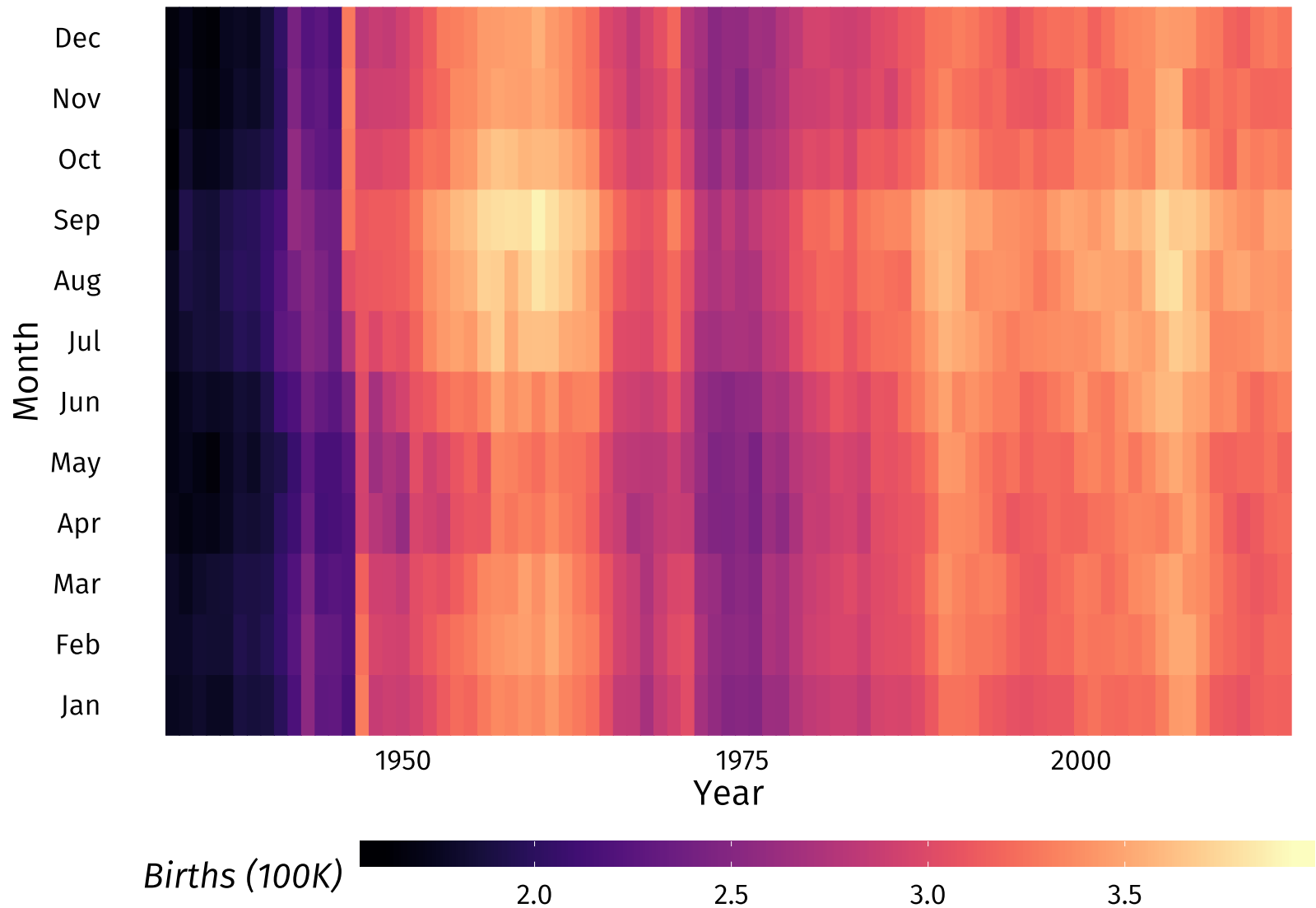


## US monthly births, 1933–2015: Newfangled time-series graph





## US monthly births per 30 days, 1933–2015: Newfangled time-series graph



# Time-series models

# Time-series models

## Introduction

Our model now looks something like

$$\text{Births}_t = \beta_0 + \beta_1 \text{Income}_t + u_t$$

or perhaps

$$\text{Births}_t = \beta_0 + \beta_1 \text{Income}_t + \beta_3 \text{Income}_{t-1} + u_t$$

maybe even

$$\text{Births}_t = \beta_0 + \beta_1 \text{Income}_t + \beta_3 \text{Income}_{t-1} + \beta_4 \text{Births}_{t-1} + u_t$$

where  $t - 1$  denotes the time period prior to  $t$  (*lagged* income or births).

# Time-series models

## Assumptions

1. **New: Weakly persistent outcomes**—essentially,  $x_{t+k}$  in the distant period  $t + k$  is weakly correlated with period  $x_t$  (when  $k$  is "big").
2.  $y_t$  is a **linear function** of its parameters and disturbance.
3. There is **no perfect collinearity** in our data.
4. The  $u_t$  have conditional mean of zero (**exogeneity**),  $E[u_t|X] = 0$ .
5. The  $u_t$  are **homoskedastic** with **zero correlation** between  $u_t$  and  $u_s$ , i.e.,  $\text{Var}(u_t|X) = \text{Var}(u_t) = \sigma^2$  and  $\text{Cor}(u_t, u_s|X) = 0$ .
6. **Normality of disturbances**, i.e.,  $u_t \stackrel{\text{iid}}{\sim} N(0, \sigma^2)$ .

# Time-series models

## Model options

Time-series modeling boils down to two classes of models.

1. **Static models:** Do not allow for persistent effect.
2. **Dynamic models:** Allow for persistent effects.
  - Models with **lagged explanatory** variables
  - **Autoregressive, distributed-lag** (ADL) models

# Time-series models

## Model options

### Option 1: Static models

**Static models** assume the outcome depends upon **only the current period**.

$$\text{Births}_t = \beta_0 + \beta_1 \text{Income}_t + u_t$$

Here, we must believe that income **immediately** affects the number of births and does not affect on the numbers of births in the future.

We also need to believe current births do not depend upon previous births.

Can be a very restrictive way to consider time-series data.

# Time-series models

## Model options

### Option 2: **Dynamic models**

**Dynamic models** allow the outcome to depend upon other periods.

# Time-series models

## Model options

**Option 2a: Dynamic models** with lagged explanatory variables

These models allow the outcome to depend upon the explanatory variable(s) in other periods.

$$\text{Births}_t = \beta_0 + \beta_1 \text{Income}_t + \beta_2 \text{Income}_{t-1} + \beta_3 \text{Income}_{t-2} + \beta_4 \text{Income}_{t-3} + u_t$$

Here, income **immediately** affects the number of births *and* affects **future** numbers of births. In other words: Births today depend today's income and lags of income—e.g., last month's income, last year's income, ...

Estimate *total* effects by summing lags' coefficients, e.g.,  $\beta_1 + \beta_2 + \beta_3 + \beta_4$ .

*Note:* We still assume current births don't affect future births.



# Time-series models

## Model options

### Option 2b: Autoregressive distributed-lag (ADL) models

These models allow the outcome to depend upon the explanatory variable(s) and/or the outcome variable in prior periods.

$$\text{Births}_t = \beta_0 + \beta_1 \text{Income}_t + \beta_2 \text{Income}_{t-1} + \beta_3 \text{Births}_{t-1} + u_t$$

Here, current income affects **current** births and **future** births.

In addition, **current births affect future births**—we're allowing lags of the outcome variable.

# Autoregressive distributed-lag models

## Numbers of lags

ADL models are often specified as  $\text{ADL}(p, q)$ , where

- $p$  is the (maximum) number of **lags** for the outcome variable.
- $q$  is the (maximum) number of **lags** for explanatory variables.

Example:  $\text{ADL}(1, 0)$

$$\text{Births}_t = \beta_0 + \beta_1 \text{Income}_t + \beta_2 \text{Births}_{t-1} + u_t$$

Example:  $\text{ADL}(2, 2)$

$$\begin{aligned} \text{Births}_t = & \beta_0 + \beta_1 \text{Income}_t + \beta_2 \text{Income}_{t-1} + \beta_3 \text{Income}_{t-2} \\ & + \beta_4 \text{Births}_{t-1} + \beta_5 \text{Births}_{t-2} + u_t \end{aligned}$$

# Autoregressive distributed-lag models

## Complexity

Due to their lags, ADL models actually estimate even more complex relationships than you might first guess.

Consider ADL(1, 0):  $\text{Births}_t = \beta_0 + \beta_1 \text{Income}_t + \beta_2 \text{Births}_{t-1} + u_t$

Write out the model for period  $t - 1$ :

$$\text{Births}_{t-1} = \beta_0 + \beta_1 \text{Income}_{t-1} + \beta_2 \text{Births}_{t-2} + u_{t-1}$$

which we can substitute in for  $\text{Births}_{t-1}$  in the first equation, *i.e.*,

$$\text{Births}_t = \beta_0 + \beta_1 \text{Income}_t + \underbrace{\beta_2(\beta_0 + \beta_1 \text{Income}_{t-1} + \beta_2 \text{Births}_{t-2} + u_{t-1})}_{\text{Births}_{t-1}} + u_t$$

# Autoregressive distributed-lag models

## Complexity

Continuing...

$$\begin{aligned}\text{Births}_t &= \beta_0 + \beta_1 \text{Income}_t + \\ &\quad \underbrace{\beta_2(\beta_0 + \beta_1 \text{Income}_{t-1} + \beta_2 \text{Births}_{t-2} + u_{t-1})}_{\text{Births}_{t-1}} + u_t \\ &= \beta_0 (1 + \beta_2) + \beta_1 \text{Income}_t + \beta_1 \beta_2 \text{Income}_{t-1} + \\ &\quad \beta_2^2 \text{Births}_{t-2} + u_t + \beta_2 u_{t-1}\end{aligned}$$

We could then substitute in the equation for  $\text{Births}_{t-2}$ ,  $\text{Births}_{t-3}$ , ...

# Autoregressive distributed-lag models

## Complexity

Eventually we arrive at

$$\begin{aligned}\text{Births}_t = & \beta_0 (1 + \beta_2 + \beta_2^2 + \beta_2^3 + \cdots) + \\ & \beta_1 (\text{Income}_t + \beta_2 \text{Income}_{t-1} + \beta_2^2 \text{Income}_{t-2} + \cdots) + \\ & u_t + \beta_2 u_{t-1} + \beta_2^2 u_{t-2} + \cdots\end{aligned}$$

## The point?

By including just **one lag of the dependent variable**—as in a ADL(1, 0)—we implicitly include for *many lags* of the explanatory variables and disturbances.<sup>†</sup>

<sup>†</sup> These lags enter into the equation in a very specific way—not the most flexible specification.

# Autoregressive distributed-lag models

## The partial-adjustment model

There are times that actually want to model an individual's **desired amount**, rather than her **actual amount**, but we are unable to observe the desired level.

*Partial-adjustment models* help us model this situation.

# Autoregressive distributed-lag models

## The partial-adjustment model

### Example

We want to know how the **desired number of cigarettes**,  $\widetilde{\text{Cig}}_t$ , changes with the current period's cigarette tax, *e.g.*,

$$\widetilde{\text{Cig}}_t = \beta_0 + \beta_1 \text{Tax}_t + u_t \quad (\text{A})$$

Imagine **actual cigarette consumption**,  $\text{Cig}_t$ , doesn't change immediately (*e.g.*, habit persistence). Instead, consumption depends upon **current desired level** and **previous consumption level**

$$\text{Cig}_t = \lambda \widetilde{\text{Cig}}_t + (1 - \lambda) \text{Cig}_{t-1} \quad (\text{B})$$

# Autoregressive distributed-lag models

## The partial-adjustment model

*Example, continued*

$$\widetilde{\text{Cig}}_t = \beta_0 + \beta_1 \text{Tax}_t + u_t \quad (\text{A})$$

$$\text{Cig}_t = \lambda \widetilde{\text{Cig}}_t + (1 - \lambda) \text{Cig}_{t-1} \quad (\text{B})$$

Substituting  $\widetilde{\text{Cig}}_t$  from (A) into (B) yields

$$\begin{aligned} \text{Cig}_t &= \lambda (\beta_0 + \beta_1 \text{Tax}_t + u_t) + (1 - \lambda) \text{Cig}_{t-1} \\ &= \lambda \beta_0 + \lambda \beta_1 \text{Tax}_t + (1 - \lambda) \text{Cig}_{t-1} + \lambda u_t \end{aligned} \quad (\text{C})$$

The equation in (C) is ADL(1, 0).

We can also estimate/recover the speed-of-adjustment coefficient  $\lambda$ .



# OLS in time series

# OLS in time series

## Unbiased coefficients

As before, the unbiased-ness of OLS is going to depend upon our exogeneity assumption, *i.e.*,  $\mathbf{E}[u_t|X] = 0$ .

We can split this assumption into two parts.

1. The disturbance  $u_t$  is independent of the explanatory variables in the **same period** (*i.e.*,  $X_t$ ).
2. The disturbance  $u_t$  is independent of the explanatory variables in the **other periods** (*i.e.*,  $X_s$  for  $s \neq t$ ).

We need both of these parts to be true for OLS to be unbiased.

# OLS in time series

## Unbiased coefficients

We need both parts of our exogeneity assumption for OLS to be unbiased:

$$\mathbf{E}[\hat{\beta}_1 | X] = \beta_1 + \mathbf{E}\left[\frac{\sum_t (x_t - \bar{x}) u_t}{\sum_t (x_t - \bar{x})^2} \middle| X\right]$$

*i.e.*, to guarantee the numerator equals zero, we need  $\mathbf{E}[u_t | X] = 0$ —for both  $\mathbf{E}[u_t | X_t] = 0$  and  $\mathbf{E}[u_t | X_s] = 0$  ( $s \neq t$ ).

The second part of our exogeneity assumption—requiring that  $u_t$  is independent of all regressors in other periods—fails with dynamic models with lagged outcome variables.

Thus, **OLS is biased for dynamic models with lagged outcome variables.**

# OLS in time series

## Unbiased coefficients

To see why dynamic models with lagged outcome variables violate our exogeneity assumption, consider two periods of our simple ADL(1, 0) model.

$$\text{Births}_t = \beta_0 + \beta_1 \text{Income}_t + \beta_2 \text{Births}_{t-1} + u_t \quad (1)$$

$$\text{Births}_{t+1} = \beta_0 + \beta_1 \text{Income}_{t+1} + \beta_2 \text{Births}_t + u_{t+1} \quad (2)$$

In (1),  $u_t$  clearly correlates with  $\text{Births}_t$ .

However,  $\text{Births}_t$  is a regressor in (2) (lagged dependent variable).

$\therefore$  The disturbance in  $t$  ( $u_t$ ) correlates with a regressor in  $t + 1$  ( $\text{Births}_t$ ).

This correlation violates the second part of our exogeneity requirement.

# OLS in time series

## Consistent coefficients

All is not lost.

For OLS to be **consistent**, we only need **contemporaneous exogeneity**.

**Contemporaneous exogeneity:** each disturbance is uncorrelated with the explanatory variables **in the same period**, *i.e.*,

$$\mathbf{E}[u_t | X_t] = 0$$

With contemporaneous exogeneity, OLS estimates for the coefficients in a time series model are consistent.

# OLS in time series

## Consistent coefficients

To see why OLS is consistent with contemporaneous exogeneity, consider the OLS estimate for  $\beta_1$  in

$$\text{Births}_t = \beta_0 + \beta_1 \text{Births}_{t-1} + u_t$$

which we've shown (a few times) can be written

$$\hat{\beta}_1 = \beta_1 + \frac{\sum_t \left( \text{Births}_{t-1} - \overline{\text{Births}} \right) u_t}{\sum_t \left( \text{Births}_{t-1} - \overline{\text{Births}} \right)^2}$$

# OLS in time series

## Consistent coefficients

$$\begin{aligned}\text{plim } \hat{\beta}_1 &= \text{plim} \left( \beta_1 + \frac{\sum_t \left( \text{Births}_{t-1} - \overline{\text{Births}} \right) u_t}{\sum_t \left( \text{Births}_{t-1} - \overline{\text{Births}} \right)^2} \right) \\&= \beta_1 + \frac{\text{plim} \left[ \sum_t \left( \text{Births}_{t-1} - \overline{\text{Births}} \right) u_t / T \right]}{\text{plim} \left[ \sum_t \left( \text{Births}_{t-1} - \overline{\text{Births}} \right)^2 / T \right]} \\&= \beta_1 + \frac{\text{Cov}(\text{Births}_{t-1}, u_t)}{\text{Var}(\text{Births}_t)} \\&= \beta_1 \quad \text{if } \text{Cov}(\text{Births}_{t-1}, u_t) = 0\end{aligned}$$

**Contemporaneous exogeneity** gives us  $\text{Cov}(\text{Births}_{t-1}, u_t) = 0$ .

# OLS in time series

## Consistent coefficients

Thus, if we assume **contemporaneous exogeneity**, **OLS is consistent** for the coefficients, *even for models with lagged dependent variables*.



The end.

# Table of contents

## Admin

1. Schedule
2. Example questions
3. Review: Asymptotics

## Time series

1. Introduction
2. Assumptions
3. Static vs. dynamic models
4. "ADL" models
  - Underlying complexity
  - Partial-adjustment models
5. Unbiasedness of OLS
6. Consistency of OLS
7. *Extra*: ADL in equilibrium

# Autoregressive distributed-lag models

## Equilibrium effects

ADL models also offer interesting insights for long-run/equilibrium effects.

$$\text{Births}_t = \beta_0 + \beta_1 \text{Income}_t + \beta_2 \text{Births}_{t-1} + u_t$$

In this ADL(1, 0) model,  $\beta_1$  gives the **short-run effect** of income on the number of births. *i.e.*, how income in time  $t$  affects births in time  $t$ .

# Autoregressive distributed-lag models

## Equilibrium effects

Starting with

$$\text{Births}_t = \beta_0 + \beta_1 \text{Income}_t + \beta_2 \text{Births}_{t-1} + u_t$$

we move into equilibrium, *i.e.*,  $\text{Births}_t = \text{Births}^*$ , *i.e.*,

$$\text{Births}^* = \beta_0 + \beta_1 \text{Income}^* + \beta_2 \text{Births}^*$$

Now rearrange...

$$\text{Births}^* - \beta_2 \text{Births}^* = \beta_0 + \beta_1 \text{Income}^*$$

$$(1 - \beta_2) \text{Births}^* = \beta_0 + \beta_1 \text{Income}^*$$

$$\text{Births}^* = \frac{\beta_0}{(1 - \beta_2)} + \frac{\beta_1}{(1 - \beta_2)} \text{Income}^*$$

# Autoregressive distributed-lag models

## Equilibrium effects

**Short-run** effect of income on births:

$$\text{Births}_t = \beta_0 + \beta_1 \text{Income}_t + \beta_2 \text{Births}_{t-1} + u_t$$

**Long-run** effect of income on births:

$$\text{Births}^* = \frac{\beta_0}{(1 - \beta_2)} + \frac{\beta_1}{(1 - \beta_2)} \text{Income}^*$$

# Autoregressive distributed-lag models

## Equilibrium effects

Another way to see this result:

We already showed

$$\text{Births}_t = \beta_0 + \beta_1 \text{Income}_t + \beta_2 \text{Births}_{t-1}$$

gives us

$$\begin{aligned} \text{Births}_t = & \beta_0 (1 + \beta_2 + \beta_2^2 + \beta_2^3 + \dots) + \\ & \beta_1 (\text{Income}_t + \beta_2 \text{Income}_{t-1} + \beta_2^2 \text{Income}_{t-2} + \dots) + \\ & u_t + \beta_2 u_{t-1} + \beta_2^2 u_{t-2} + \dots \end{aligned}$$

In equilibrium:  $\text{Income}_t = \text{Income}_{t-k} = \text{Income}^*$  for all  $k$ .

# Autoregressive distributed-lag models

## Equilibrium effects

Substituting  $\text{Income}_t = \text{Income}^*$  for all  $k$   
(and assuming no disturbances in equilibrium):

$$\begin{aligned}\text{Births}_t &= \beta_0 (1 + \beta_2 + \beta_2^2 + \beta_2^3 + \dots) + \\ &\quad \beta_1 (\text{Income}^* + \beta_2 \text{Income}^* + \beta_2^2 \text{Income}^* + \dots) + \\ &= \beta_0 \left( \frac{1}{1 - \beta_2} \right) + \\ &\quad \beta_1 \left( \frac{1}{1 - \beta_2} \right) \text{Income}^*\end{aligned}$$

So long as  $-1 < \beta_2 < 1$ .<sup>†</sup>

<sup>†</sup> This simplification comes from  $\sum_{k=0}^{\infty} p^k = \frac{1}{1-p}$  for  $-1 < p < 1$ .