Autocorrelation

EC 421, Set 8

Edward Rubin
05 February 2019

Prologue

Schedule

Last Time

Introduction to time series

Today

Autocorrelation

Upcoming

1. **Survey** due at 11:59pm on Wednesday (02/06).
2. **Assignment** *extended* to noon on Saturday (02/09). Don't wait.
3. **Midterm** next Tuesday (02/11).

R showcase

ggplot2

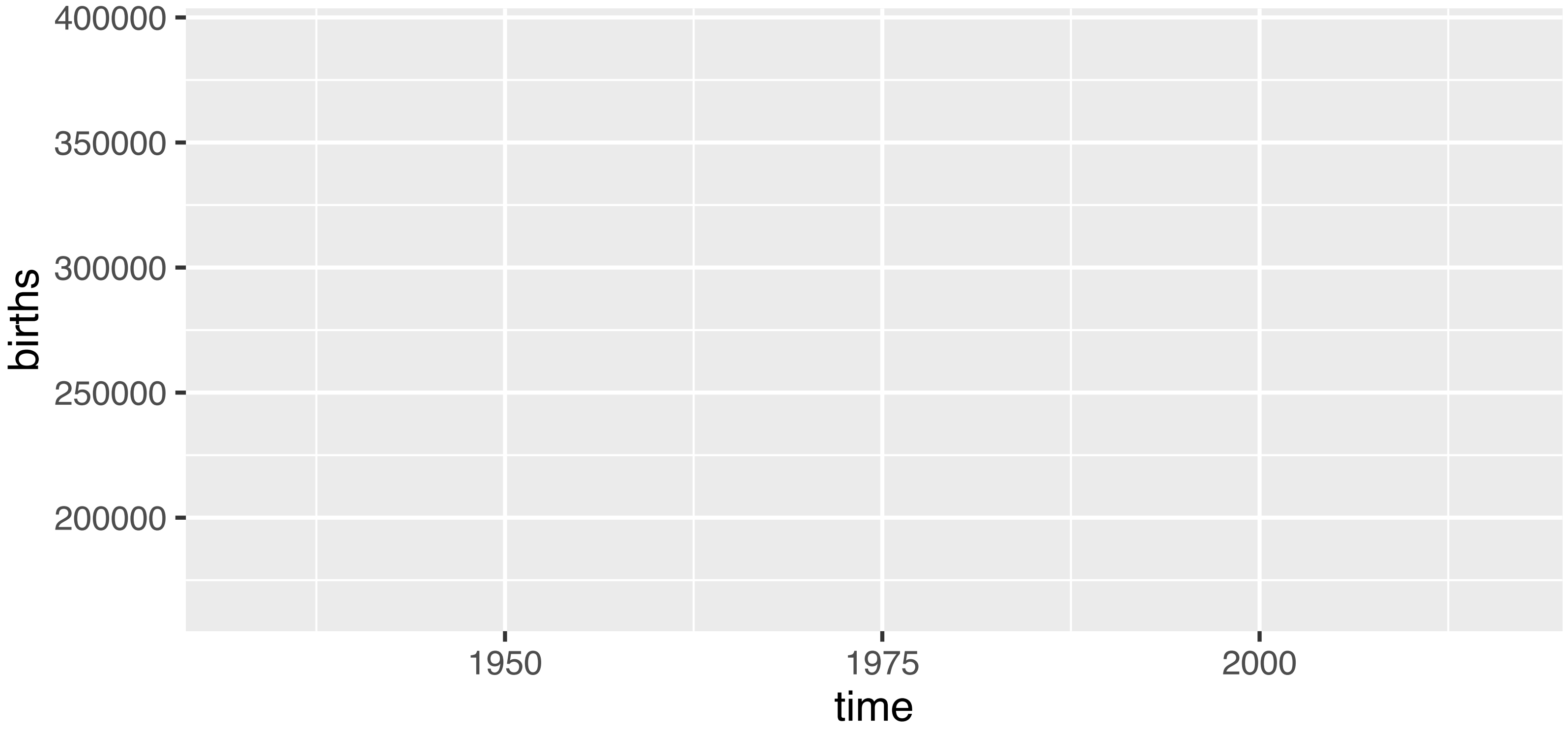
I previously mentioned the R package `ggplot2`.

Today, I'm going to show you a bit of the basics of `ggplot2`.

ggplot2

The `ggplot` function `aes` arguments define variables from `data`.

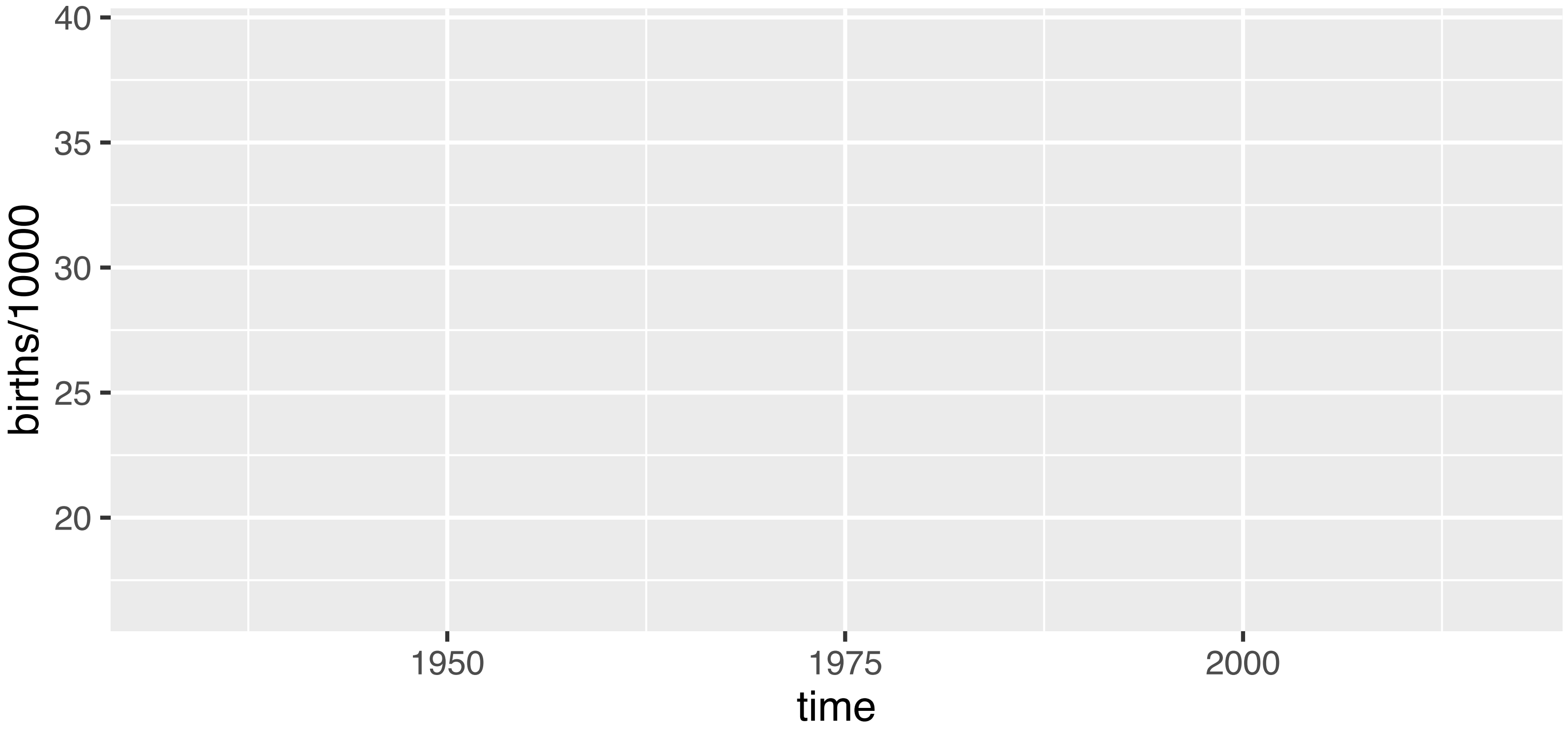
```
ggplot(data = birth_df, aes(x = time, y = births))
```



ggplot2

You can apply mathematical operators to the variables.

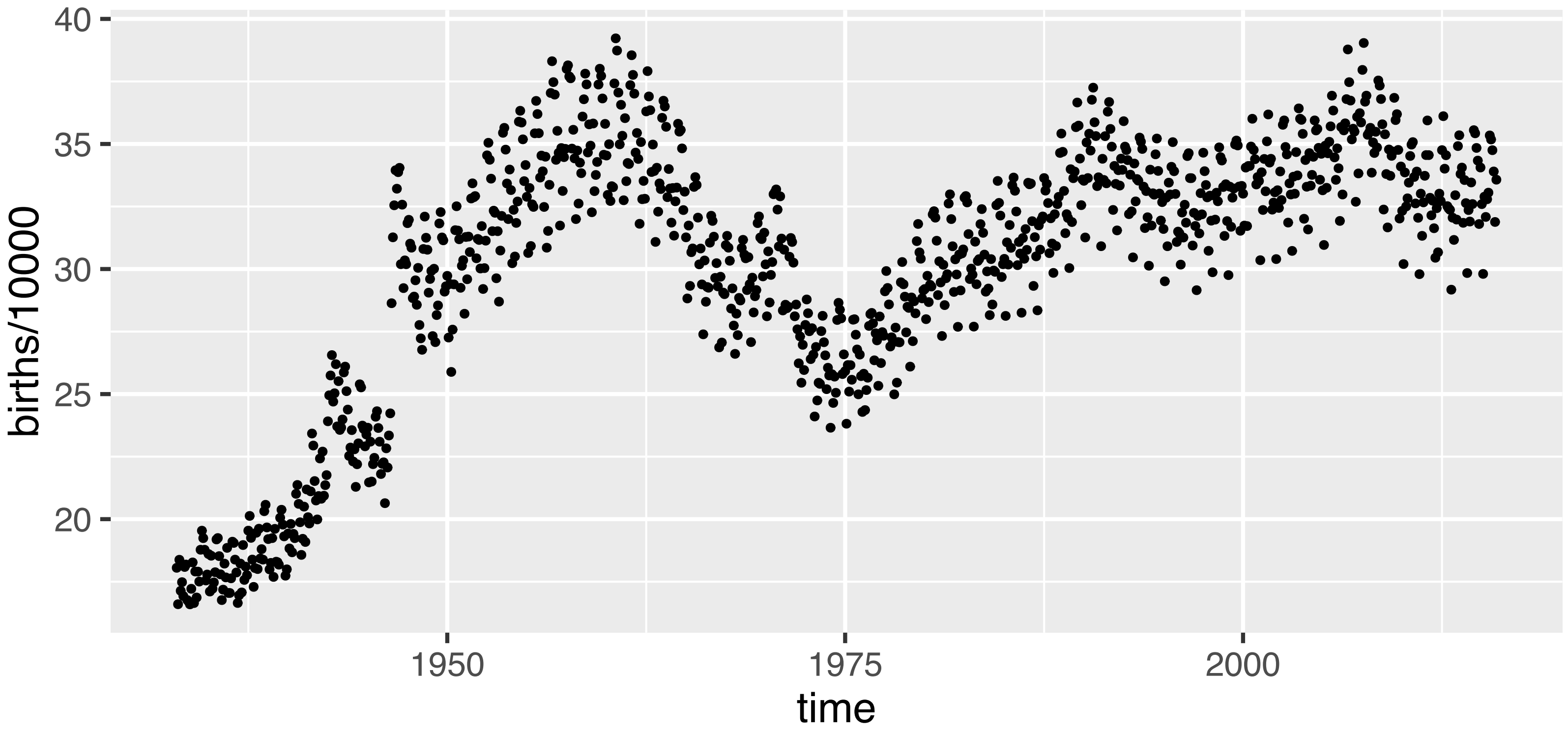
```
ggplot(data = birth_df, aes(x = time, y = births/10000))
```



ggplot2

You add *geometries* (points, lines, *etc.*) layer by layer.

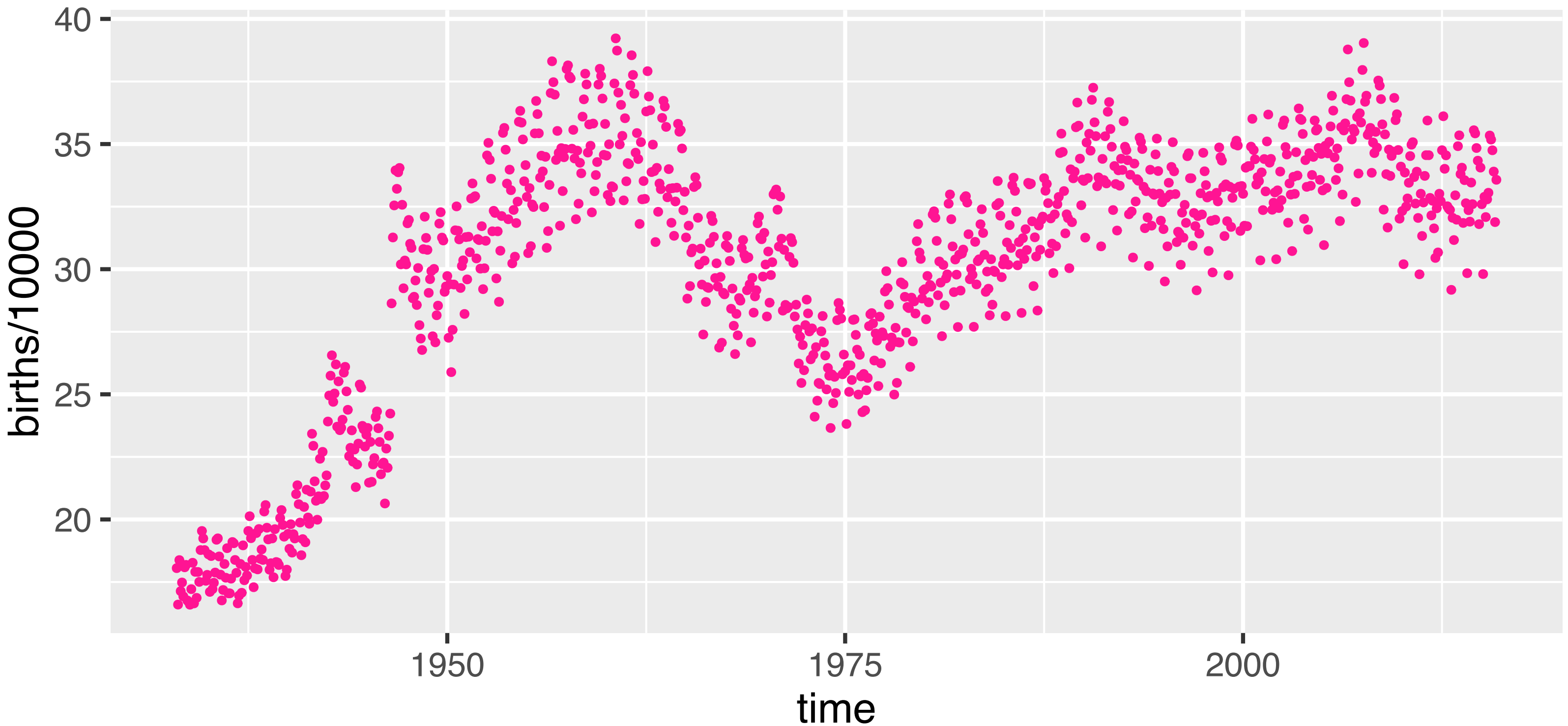
```
ggplot(data = birth_df, aes(x = time, y = births/10000)) +
  geom_point()
```



ggplot2

Color is easy.

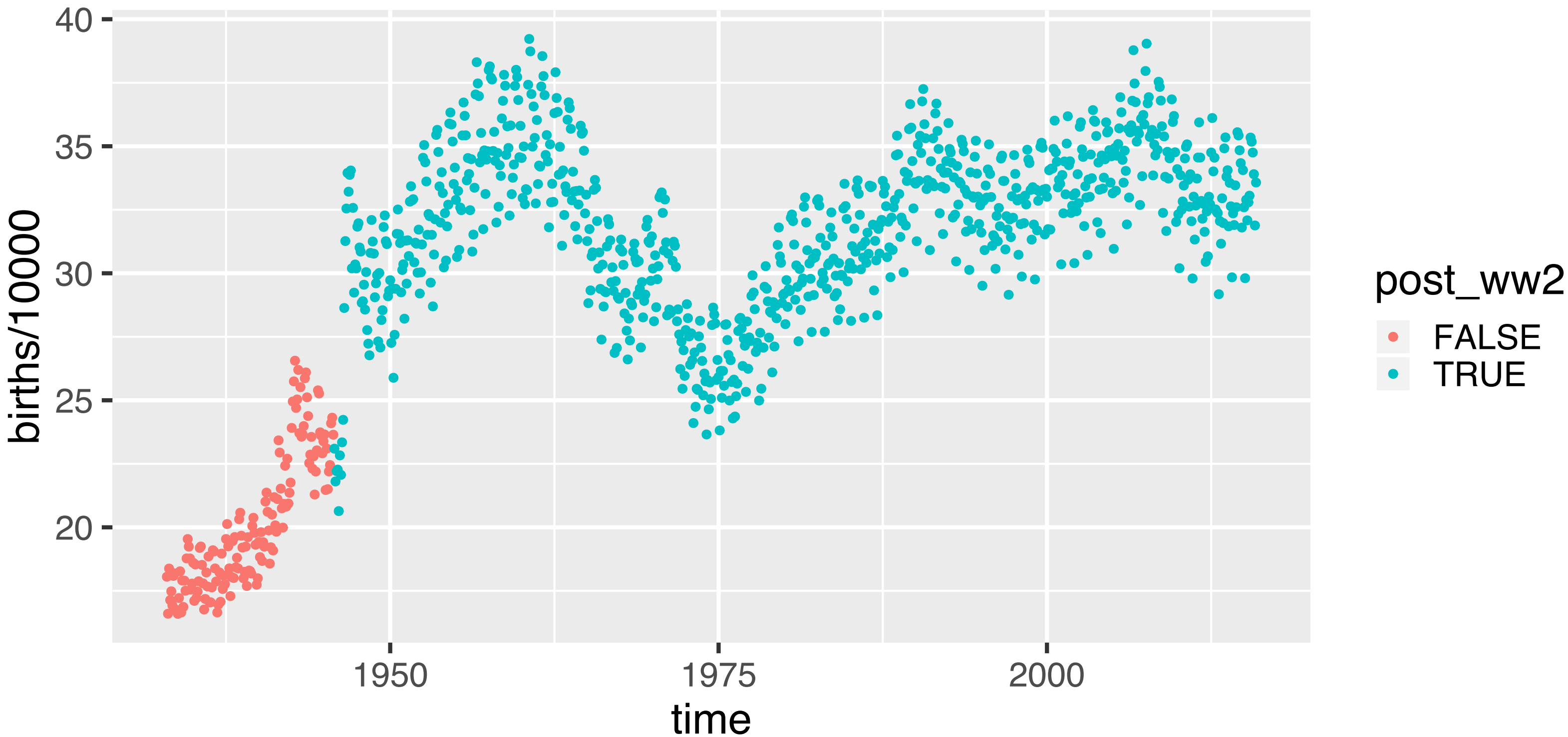
```r
ggplot(data = birth_df, aes(x = time, y = births/10000)) +
  geom_point(color = "deeppink")
```



ggplot2

You can even use variables to color.

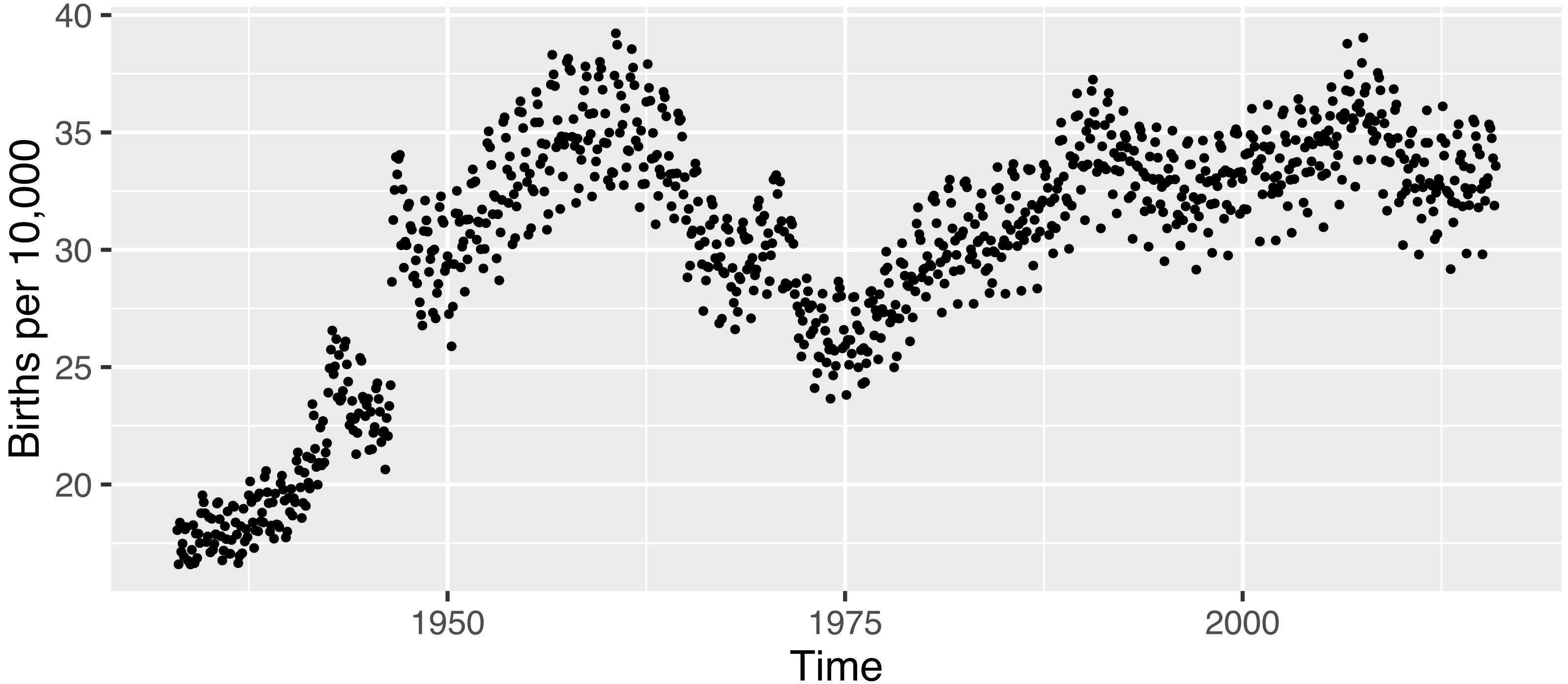
```r
ggplot(data = birth_df, aes(x = time, y = births/10000)) +
  geom_point(aes(color = post_ww2))
```



ggplot2

Add labels

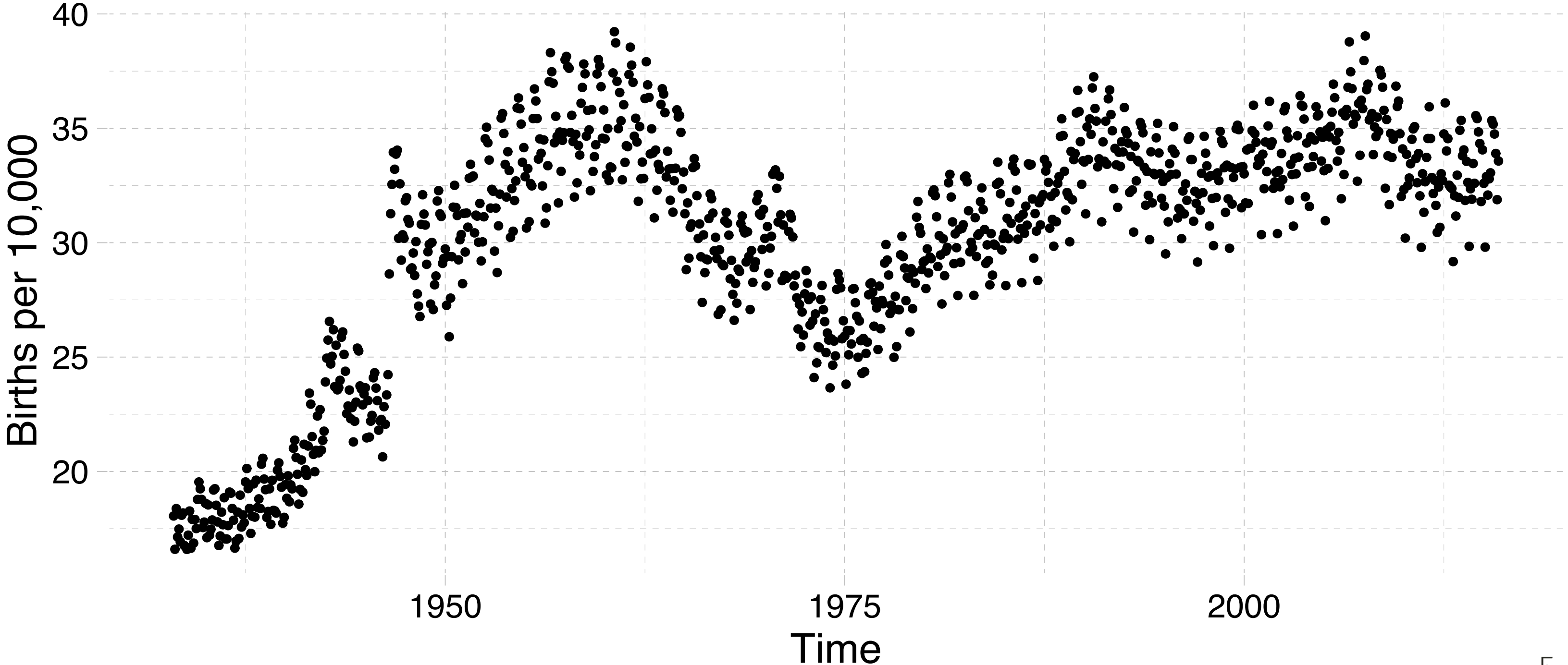
```r
ggplot(data = birth_df, aes(x = time, y = births/10000)) +
  geom_point() +
  xlab("Time") + ylab("Births per 10,000")
```



ggplot2

Change the theme...

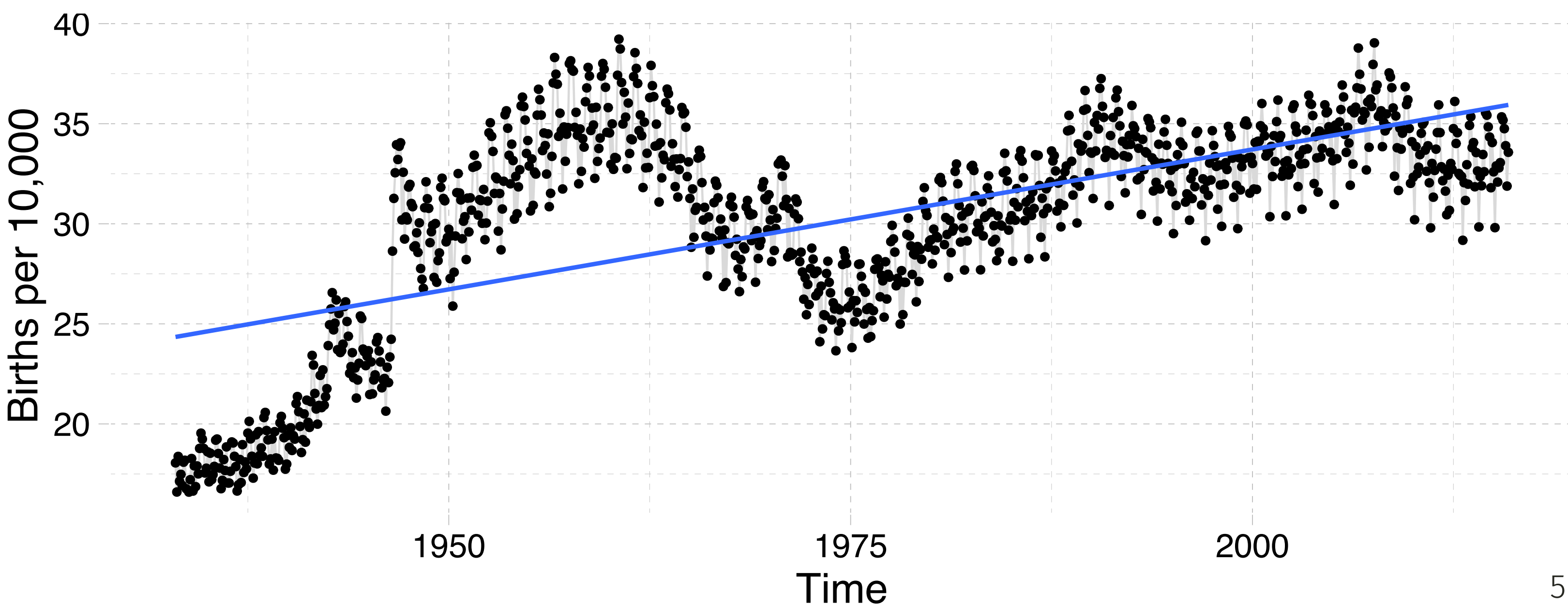
```
ggplot(data = birth_df, aes(x = time, y = births/10000)) +
  geom_point() +
  xlab("Time") + ylab("Births per 10,000") +
  theme_pander(base_size = 20)
```



ggplot2

Add other geometries—*e.g.*, connect the dots (`line`) and a regression line

```
ggplot(data = birth_df, aes(x = time, y = births/10000)) +
  geom_line(color = "grey85") +
  geom_point() +
  geom_smooth(se = F, method = lm) +
  xlab("Time") + ylab("Births per 10,000") +
  theme_pander(base_size = 20)
```



Time series

Review

Review

Changes to our model/framework.

- Our model now has t subscripts for **time periods**.
- **Dynamic models** allow **lags** of explanatory and/or outcome variables.
- We changed our **exogeneity** assumption to **contemporaneous exogeneity**, *i.e.*, $\boldsymbol{E}[u_t|X_t] = 0$
- Including **lags of outcome variables** causes **biased/inconsistent coefficient estimates** for OLS.
- **Lagged explanatory variables** make **OLS inefficient**.

Autocorrelation

Autocorrelation

What is it?

Autocorrelation occurs when our disturbances are correlated over time, *i.e.*, $\text{Cov}(u_t,\ u_s) \neq 0$ for $t \neq s$.

Autocorrelation

What is it?

Autocorrelation occurs when our disturbances are correlated over time, *i.e.*, $\mathrm{Cov}(u_t,\, u_s) \neq 0$ for $t \neq s$.

Another way to think about: If the *shock* from disturbance t correlates with "nearby" shocks in $t-1$ and $t+1$.

Autocorrelation

What is it?

Autocorrelation occurs when our disturbances are correlated over time, *i.e.*, $\text{Cov}(u_t,\, u_s) \neq 0$ for $t \neq s$.

Another way to think about: If the *shock* from disturbance t correlates with "nearby" shocks in $t-1$ and $t+1$.

Note: **Serial correlation** and **autocorrelation** are the same thing.

Autocorrelation

What is it?

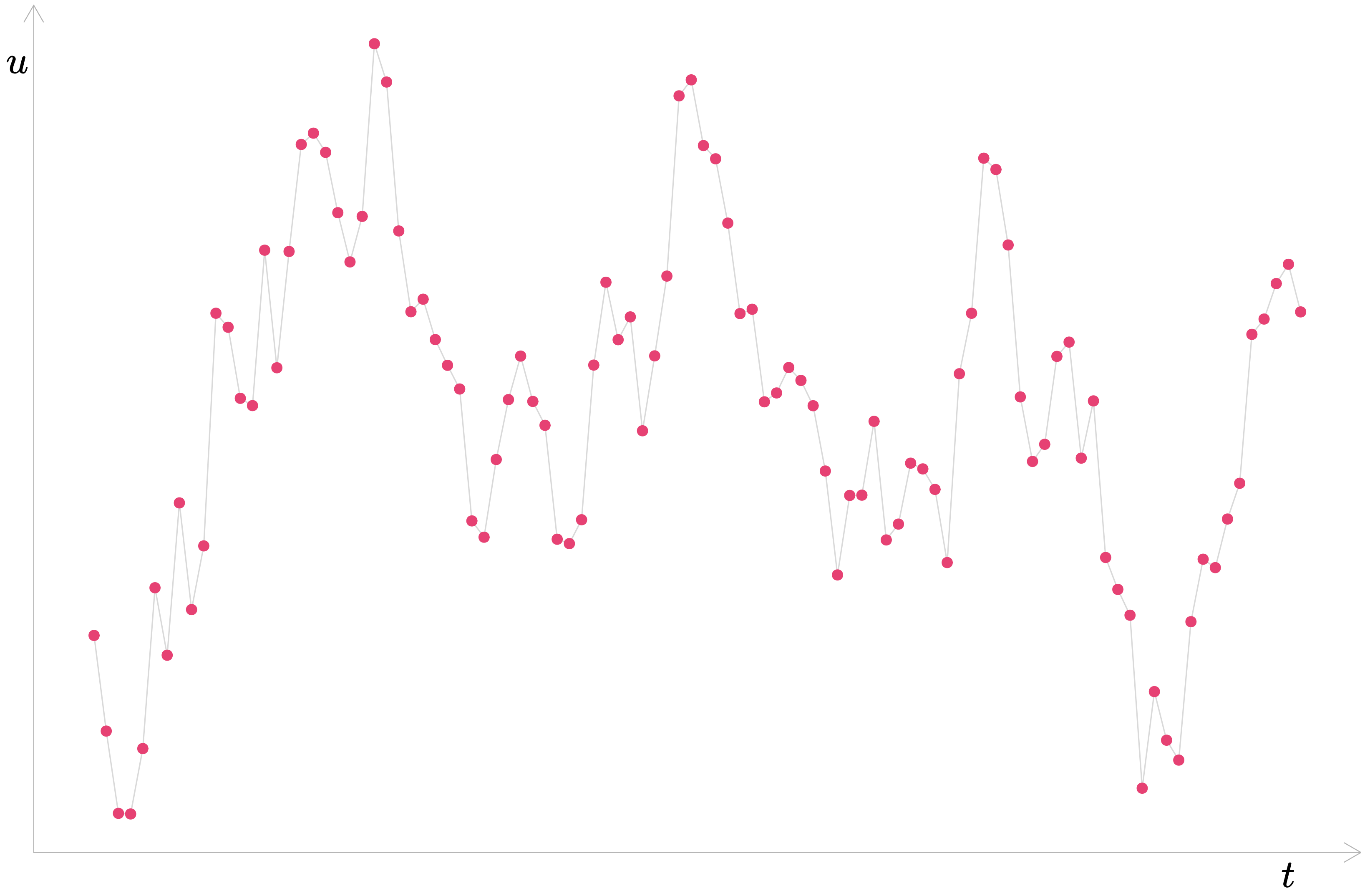
Autocorrelation occurs when our disturbances are correlated over time, *i.e.*, $\mathrm{Cov}(u_t,\, u_s) \neq 0$ for $t \neq s$.

Another way to think about: If the *shock* from disturbance t correlates with "nearby" shocks in $t-1$ and $t+1$.

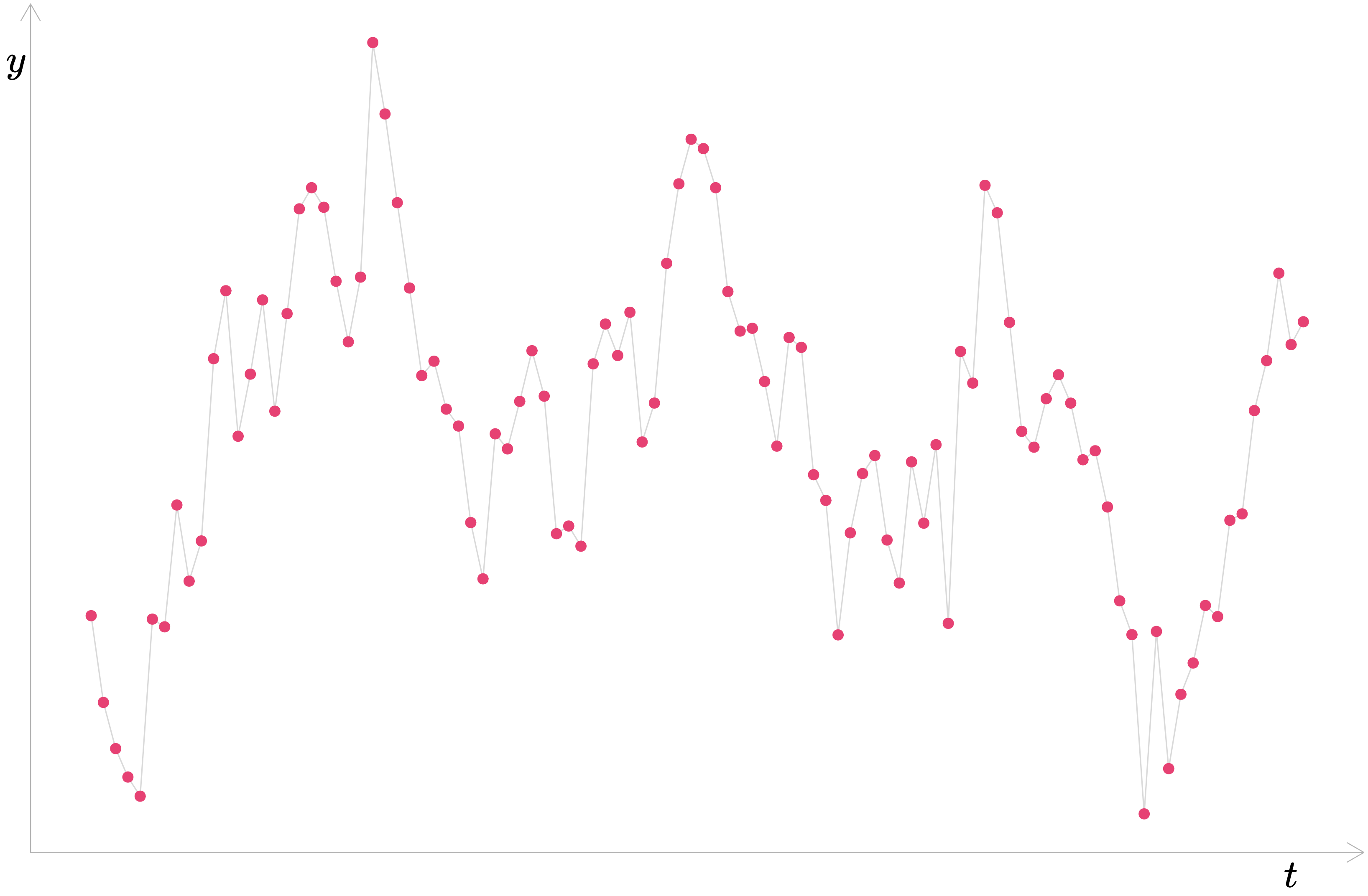
Note: **Serial correlation** and **autocorrelation** are the same thing.

Why is autocorrelation prevalent in time-series analyses?

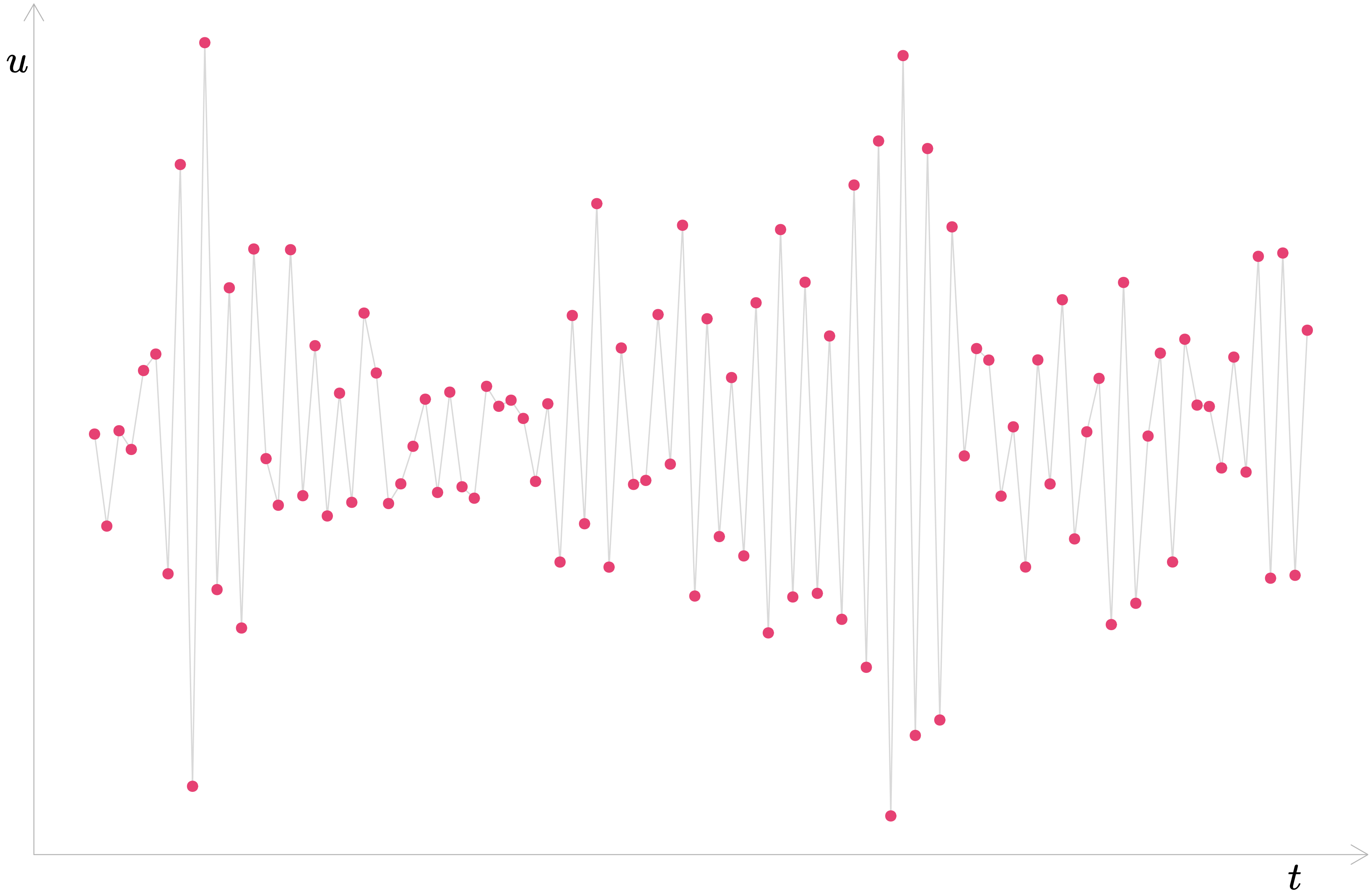
Positive autocorrelation: Disturbances (u_t) over time



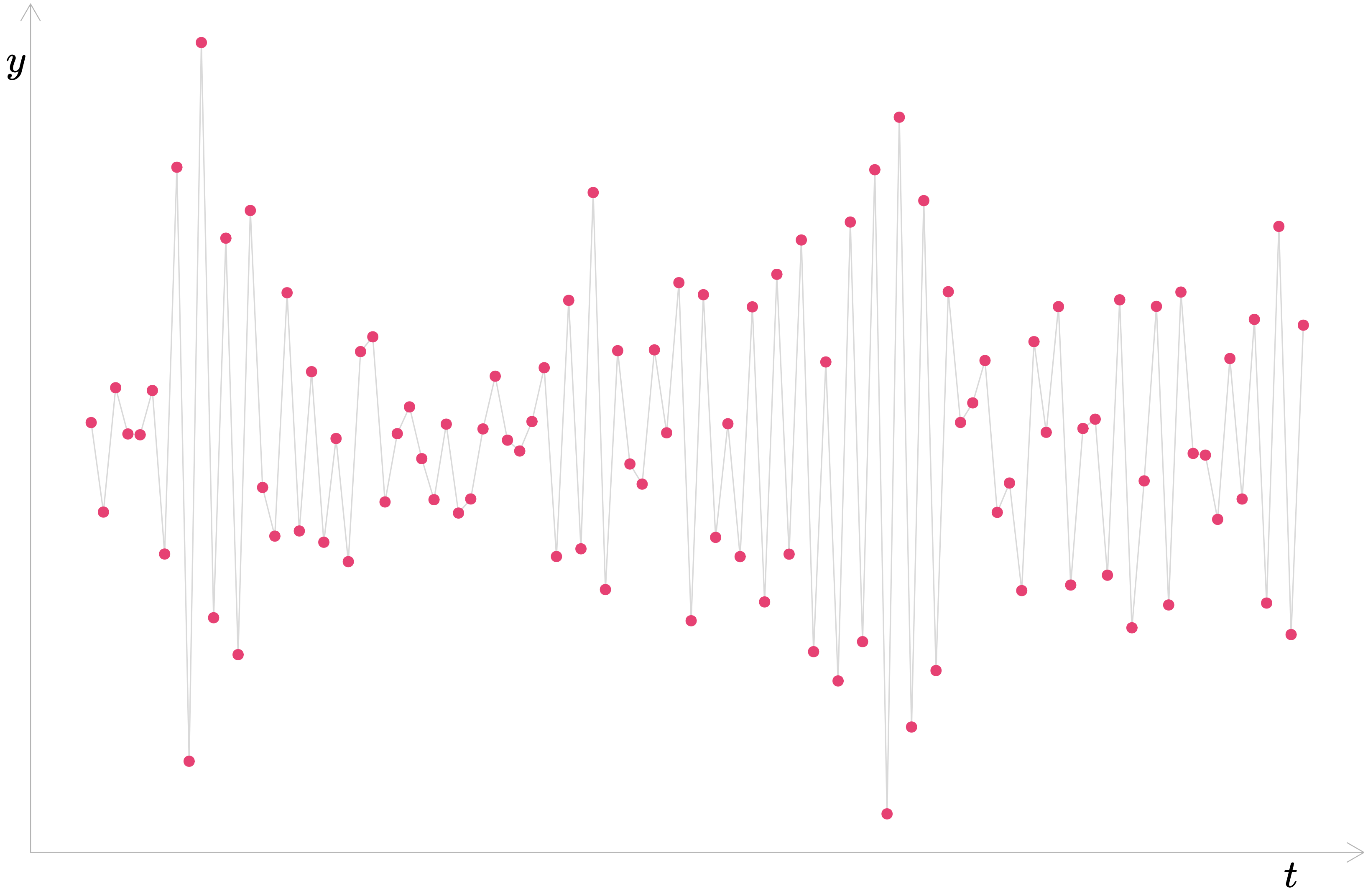
Positive autocorrelation: Outcomes (y_t) over time



Negative autocorrelation: Disturbances (u_t) over time



Negative autocorrelation: Outcomes (y_t) over time



Autocorrelation

In static time-series models

Let's start with a very common model: a static time-series model whose disturbances exhibit **first-order autocorrelation**, *a.k.a.* AR(1):

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

where

$$u_t = \rho \, u_{t-1} + \varepsilon_t$$

and the ε_t are independently and identically distributed (*i.i.d.*).

Autocorrelation

In static time-series models

Let's start with a very common model: a static time-series model whose disturbances exhibit **first-order autocorrelation**, *a.k.a.* AR(1):

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

where

$$u_t = \rho \, u_{t-1} + \varepsilon_t$$

and the ε_t are independently and identically distributed (*i.i.d.*).

Second-order autocorrelation, or AR(2), would be

$$u_t = \rho_1 u_{t-1} + \rho_2 u_{t-2} + \varepsilon_t$$

Autocorrelation

In static time-series models

An AR(p) model/process has a disturbance structure of

$$u_t = \sum_{j=1}^{p} \rho_j u_{t-j} + \varepsilon_t$$

allowing the current disturbance to correlated with up to p of its lags.

OLS

For **static models** or **dynamic models with lagged explanatory variables**, in the presence of autocorrelation

1. OLS provides **unbiased** estimates for the coefficients.
2. OLS creates **biased** estimates for the standard errors.
3. OLS is **inefficient**.

Recall: Same implications as heteroskedasticity.

Autocorrelation get trickier with lagged outcome variables.

OLS and lagged outcome variables

Consider a model with one lag of the outcome variable—ADL(1, 0)—model with AR(1) disturbances

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

where

$$u_t = \rho u_{t-1} + \varepsilon_t$$

OLS and lagged outcome variables

Consider a model with one lag of the outcome variable—ADL(1, 0)—model with AR(1) disturbances

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

where

$$u_t = \rho u_{t-1} + \varepsilon_t$$

Problem:

OLS and lagged outcome variables

Consider a model with one lag of the outcome variable—ADL(1, 0)—model with AR(1) disturbances

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

where

$$u_t = \rho u_{t-1} + \varepsilon_t$$

Problem: Both Births_{t-1} (a regressor in the model for time t) and u_t (the disturbance for time t) depend upon u_{t-1}. *I.e.*, a regressor is correlated with its contemporaneous disturbance.

OLS and lagged outcome variables

Consider a model with one lag of the outcome variable—ADL(1, 0)—model with AR(1) disturbances

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

where

$$u_t = \rho u_{t-1} + \varepsilon_t$$

Problem: Both Births_{t-1} (a regressor in the model for time t) and u_t (the disturbance for time t) depend upon u_{t-1}. *I.e.*, a regressor is correlated with its contemporaneous disturbance.

Q: Why is this a problem?

Autocorrelation

OLS and lagged outcome variables

Consider a model with one lag of the outcome variable—ADL(1, 0)—model with AR(1) disturbances

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

where

$$u_t = \rho u_{t-1} + \varepsilon_t$$

Problem: Both Births_{t-1} (a regressor in the model for time t) and u_t (the disturbance for time t) depend upon u_{t-1}. *I.e.*, a regressor is correlated with its contemporaneous disturbance.

Q: Why is this a problem?

A: It violates contemporaneous exogeneity

OLS and lagged outcome variables

Consider a model with one lag of the outcome variable—ADL(1, 0)—model with AR(1) disturbances

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

where

$$u_t = \rho u_{t-1} + \varepsilon_t$$

Problem: Both Births_{t-1} (a regressor in the model for time t) and u_t (the disturbance for time t) depend upon u_{t-1}. *I.e.*, a regressor is correlated with its contemporaneous disturbance.

Q: Why is this a problem?

A: It violates contemporaneous exogeneity, *i.e.*, $\text{Cov}(x_t,\, u_t) = 0$.

Autocorrelation

OLS and lagged outcome variables

To see this problem, first write out the model for t and $t-1$:

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

and now note that $u_t = \rho u_{t-1} + \varepsilon_t$. Substituting...

$$\text{Births}_t = \beta_0 + \beta_1 \text{Income}_t + \beta_2 \text{Births}_{t-1} + \overbrace{\left(\rho u_{t-1} + \varepsilon_t\right)}^{u_t} \quad (1)$$
$$\text{Births}_{t-1} = \beta_0 + \beta_1 \text{Income}_{t-1} + \beta_2 \text{Births}_{t-2} + u_{t-1} \quad (2)$$

In (1), we can see that u_t depends upon (covaries with) u_{t-1}.

In (2), we can see that Births_{t-1}, a regressor in (1), also covaries with u_{t-1}.

OLS and lagged outcome variables

To see this problem, first write out the model for t and $t-1$:

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

and now note that $u_t = \rho u_{t-1} + \varepsilon_t$. Substituting...

$$\text{Births}_t = \beta_0 + \beta_1 \text{Income}_t + \beta_2 \text{Births}_{t-1} + \overbrace{(\rho u_{t-1} + \varepsilon_t)}^{u_t} \tag{1}$$

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

In (1), we can see that u_t depends upon (covaries with) u_{t-1}.

In (2), we can see that Births_{t-1}, a regressor in (1), also covaries with u_{t-1}.

$\therefore$ This model violates our contemporaneous exogeneity requirement.

OLS and lagged outcome variables

Implications: For models with **lagged outcome variables** and **autocorrelated disturbances**

1. The models **violate contemporaneous exogeneity**.

2. OLS is **biased and inconsistent** for the coefficients.

OLS and lagged outcome variables

Intuition? Why is OLS inconsistent and biased when we violate exogeneity?

Think back to omitted-variable bias...

$$y_t = \beta_0 + \beta_1 x_t + u_t$$

When $\mathrm{Cov}(x_t,\, u_t) \neq 0$, we cannot separate the effect of u_t on y_t from the effect of x_t on y_t. Thus, we get inconsistent estimates for β_1.

OLS and lagged outcome variables

Intuition? Why is OLS inconsistent and biased when we violate exogeneity?

Think back to omitted-variable bias...

$$y_t = \beta_0 + \beta_1 x_t + u_t$$

When $\text{Cov}(x_t,\, u_t) \neq 0$, we cannot separate the effect of u_t on y_t from the effect of x_t on y_t. Thus, we get inconsistent estimates for β_1. Similarly,

$$\text{Births}_t = \beta_0 + \beta_1 \text{Income}_t + \beta_2 \text{Births}_{t-1} + \overbrace{(\rho u_{t-1} + \varepsilon_t)}^{u_t} \tag{1}$$

we cannot separate the effects of u_t on Births_t from Births_{t-1} on Births_t, because both u_t and Births_{t-1} depend upon u_{t-1}.

OLS and lagged outcome variables

Intuition? Why is OLS inconsistent and biased when we violate exogeneity?

Think back to omitted-variable bias...

$$y_t = \beta_0 + \beta_1 x_t + u_t$$

When $\text{Cov}(x_t,\, u_t) \neq 0$, we cannot separate the effect of u_t on y_t from the effect of x_t on y_t. Thus, we get inconsistent estimates for β_1. Similarly,

$$\text{Births}_t = \beta_0 + \beta_1 \text{Income}_t + \beta_2 \text{Births}_{t-1} + \overbrace{\left(\rho u_{t-1} + \varepsilon_t\right)}^{u_t} \tag{1}$$

we cannot separate the effects of u_t on Births_t from Births_{t-1} on Births_t, because both u_t and Births_{t-1} depend upon u_{t-1}. $\hat{\beta}_2$ is **biased** (w/ OLS).

Simulation

To see how this bias can look, let's run a simulation.

$$y_t = 1 + 2x_t + 0.5y_{t-1} + u_t$$
$$u_t = 0.9u_{t-1} + \varepsilon_t$$

To generate 100 disturbances from AR(1), with $\rho = 0.9$:

```
arima.sim(model = list(ar = c(0.9)), n = 100)
```

Simulation

To see how this bias can look, let's run a simulation.

$$
\begin{aligned}
y_t &= 1 + 2x_t + 0.5y_{t-1} + u_t \\
u_t &= 0.9u_{t-1} + \varepsilon_t
\end{aligned}
$$

To generate 100 disturbances from AR(1), with $\rho = 0.9$:

```
arima.sim(model = list(ar = c(0.9)), n = 100)
```

We are going to run 10,000 iterations with $T = 100$.

Autocorrelation and bias

Simulation

To see how this bias can look, let's run a simulation.

$$y_t = 1 + 2x_t + 0.5y_{t-1} + u_t$$
$$u_t = 0.9u_{t-1} + \varepsilon_t$$

To generate 100 disturbances from AR(1), with $\rho = 0.9$:

```
arima.sim(model = list(ar = c(0.9)), n = 100)
```

We are going to run 10,000 iterations with $T = 100$.

Q: Will this simulation tell us about *bias* or *consistency*?

Simulation

To see how this bias can look, let's run a simulation.

$$y_t = 1 + 2x_t + 0.5y_{t-1} + u_t$$
$$u_t = 0.9u_{t-1} + \varepsilon_t$$

To generate 100 disturbances from AR(1), with $\rho = 0.9$:

```
arima.sim(model = list(ar = c(0.9)), n = 100)
```

We are going to run 10,000 iterations with $T = 100$.

Q: Will this simulation tell us about *bias* or *consistency*?

A: Bias. We would need to let $T \to \infty$ to consider consistency.

Autocorrelation and bias

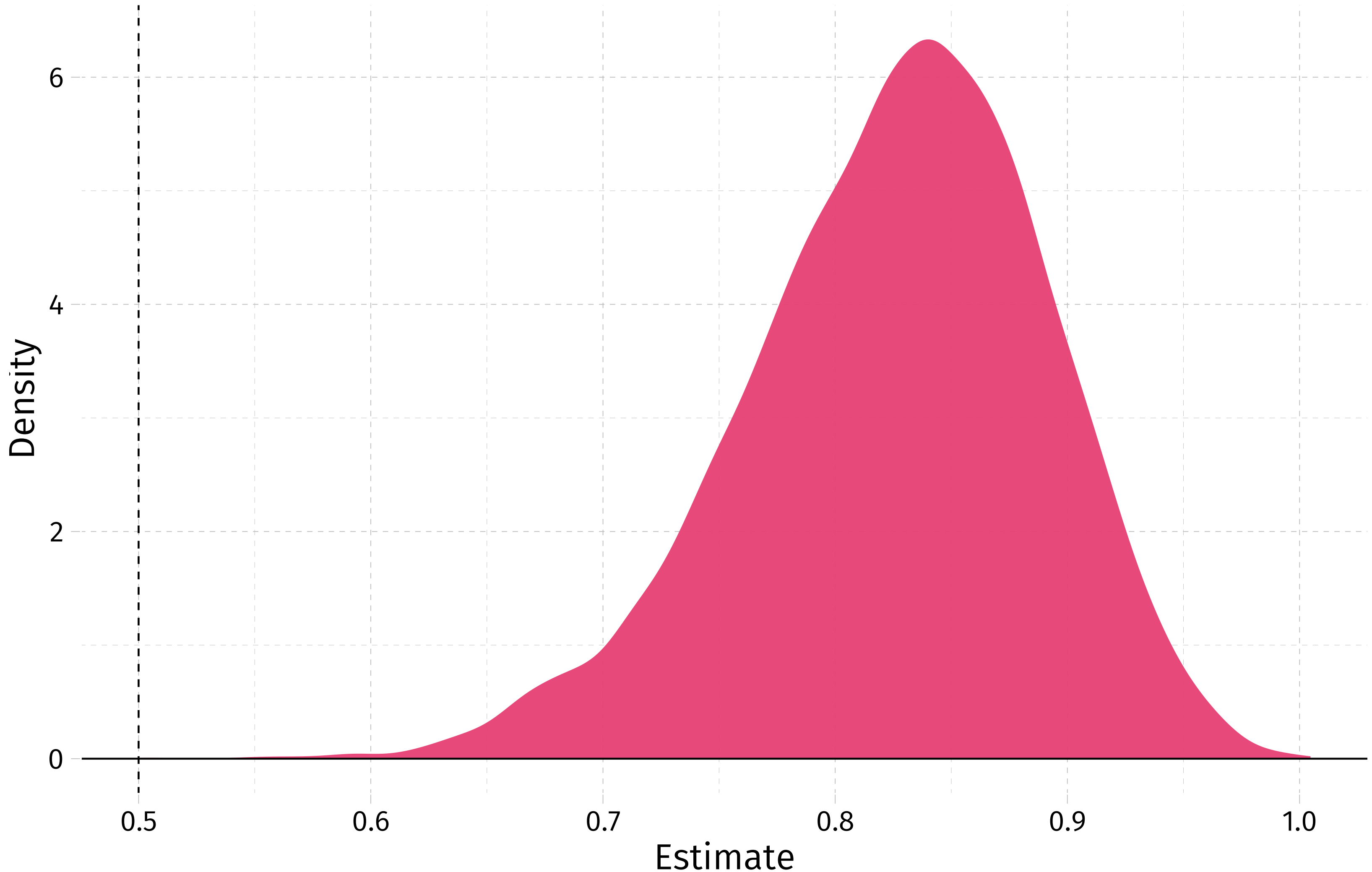
Simulation

Outline of our simulation:

```
1. Generate T=100 values of x

2. Generate T=100 values of u

    ∘ Generate T=100 values of ε

    ∘ Use ε and ρ=0.9 to calculate u_t = ρ u_t-1 + ε_t

3. Calculate y_t = β_0 + β_1 x_t + β_2 y_t-1 + u_t

4. Regress y on x; record estimates

Repeat 1–4 10,000 times
```

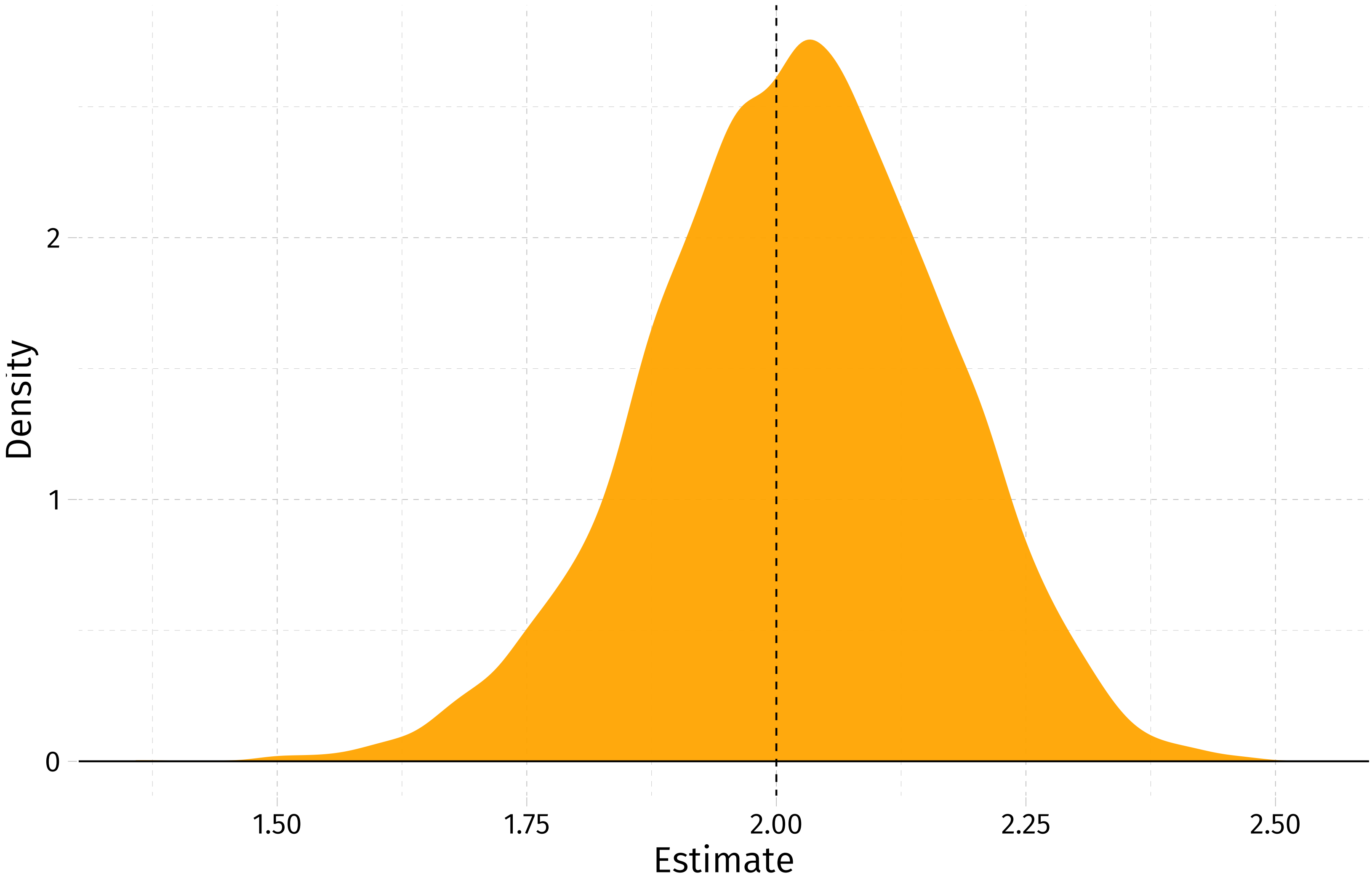
Distribution of OLS estimates, $\hat{\beta}_2$

$$y_t = 1 + 2x_t + 0.5y_{t-1} + u_t$$



Distribution of OLS estimates, $\hat{\beta}_1$

$$y_t = 1 + 2x_t + 0.5y_{t-1} + u_t$$



Testing for autocorrelation

Testing for autocorrelation

Static models

Suppose we have the **static model**,

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

and we want to test for an AR(1) process in our disturbances u_t.

Testing for autocorrelation

Static models

Suppose we have the **static model**,

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

and we want to test for an AR(1) process in our disturbances u_t.

Test for autocorrelation: Test for correlation in the lags of our residuals:

Testing for autocorrelation

Static models

Suppose we have the **static model**,

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

and we want to test for an AR(1) process in our disturbances u_t.

Test for autocorrelation: Test for correlation in the lags of our residuals:

$$e_t = \rho e_{t-1} + v_t$$

Testing for autocorrelation

Static models

Suppose we have the **static model**,

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

and we want to test for an AR(1) process in our disturbances u_t.

Test for autocorrelation: Test for correlation in the lags of our residuals:

$$e_t = \rho e_{t-1} + v_t$$

Does $\hat{\rho}$ differ significantly from zero?

Testing for autocorrelation

Static models

Specifically, to test for AR(1) disturbances in the static model

$$\textbf{Births}_t = \beta_0 + \beta_1 \textbf{Income}_t + u_t \tag{A}$$

1. Estimate $(\mathbf{A})$ via OLS.

2. Calculate residuals from the OLS regression in step 1.

3. Regress the residuals on their lags (without an intercept).

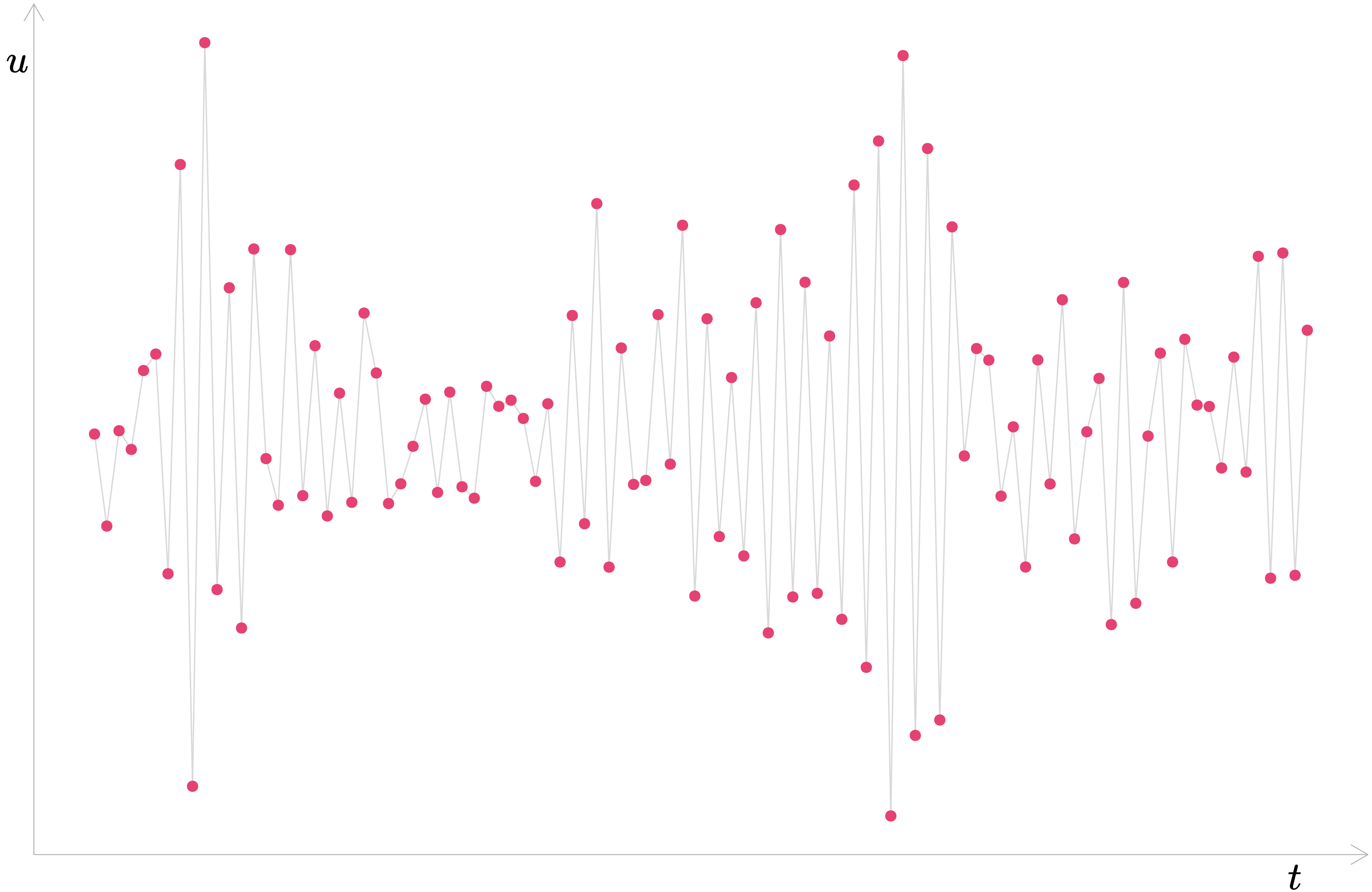
$$e_t = \rho\ e_{t-1} + v_t$$

4. Use a *t* test to determine whether there is statistically significant evidence that ρ differs from zero.

5. Rejecting H_0 implies significant evidence of autocorrelation.

For an example, let's return to our plot of negative autocorrelation.

Negative autocorrelation: Disturbances (u_t) over time



Example: Static model and AR(1)

Step 1: Estimate the static model $(y_t = \beta_0 + \beta_1 x_t + u_t)$ with OLS

```
reg_est ← lm(y ~ x, data = ar_df)
```

Testing for autocorrelation

Example: Static model and AR(1)

Step 1: Estimate the static model $(y_t = \beta_0 + \beta_1 x_t + u_t)$ with OLS

```
reg_est ← lm(y ~ x, data = ar_df)
```

Step 2: Add the residuals to our dataset

```
ar_df$e ← residuals(reg_est)
```

Testing for autocorrelation

Example: Static model and AR(1)

Step 1: Estimate the static model ($y_t = \beta_0 + \beta_1 x_t + u_t$) with OLS

```
reg_est ← lm(y ~ x, data = ar_df)
```

Step 2: Add the residuals to our dataset

```
ar_df$e ← residuals(reg_est)
```

Step 3: Regress the residual on its lag (no intercept)

```
reg_resid ← lm(e ~ -1 + lag(e), data = ar_df)
```

Testing for autocorrelation

Example: Static model and AR(1)

Step 4: t test for the estimated $(\hat{\rho})$ coefficient in step 3.

```
tidy(reg_resid)
```

```
#> # A tibble: 1 x 5
#>   term   estimate std.error statistic  p.value
#>   <chr>     <dbl>     <dbl>     <dbl>    <dbl>
#> 1 lag(e)   -0.851    0.0535     -15.9 6.88e-29
```

Example: Static model and AR(1)

Step 4: t test for the estimated $(\hat{\rho})$ coefficient in step 3.

```
tidy(reg_resid)
```

```
#> # A tibble: 1 x 5
#>   term   estimate std.error statistic  p.value
#>   <chr>     <dbl>     <dbl>     <dbl>    <dbl>
#> 1 lag(e)   -0.851    0.0535     -15.9 6.88e-29
```

That's a very small p-value—much smaller than 0.05.

Example: Static model and AR(1)

Step 4: t test for the estimated $(\hat{\rho})$ coefficient in step 3.

```
tidy(reg_resid)
```

```
#> # A tibble: 1 x 5
#>   term    estimate std.error statistic  p.value
#>   <chr>      <dbl>     <dbl>     <dbl>    <dbl>
#> 1 lag(e)    -0.851    0.0535     -15.9 6.88e-29
```

That's a very small *p*-value—much smaller than 0.05.

Reject H_0 (H_0 was $\rho = 0$, *i.e.*, no autocorrelation).

Testing for autocorrelation

Example: Static model and AR(1)

Step 4: t test for the estimated $(\hat{\rho})$ coefficient in step 3.

```
tidy(reg_resid)
```

```
#> # A tibble: 1 x 5
#>   term   estimate std.error statistic  p.value
#>   <chr>     <dbl>     <dbl>     <dbl>    <dbl>
#> 1 lag(e)   -0.851    0.0535     -15.9 6.88e-29
```

That's a very small *p*-value—much smaller than 0.05.

Reject H_0 (H_0 was $\rho = 0$, *i.e.*, no autocorrelation).

Step 5: Conclude. Statistically significant evidence of autocorrelation.

Testing for autocorrelation

Example: Static model and AR(3)

What if we wanted to test for AR(3)?

- We add more lags of residuals to the regression in *Step 3*.

- We **jointly** test the significance of the coefficients (*i.e.*, LM or $\boldsymbol{F}$).

Let's do it.

Example: Static model and AR(3)

Step 1: Estimate the static model ($y_t = \beta_0 + \beta_1 x_t + u_t$) with OLS

```
reg_est ← lm(y ~ x, data = ar_df)
```

Testing for autocorrelation

Example: Static model and AR(3)

Step 1: Estimate the static model $(y_t = \beta_0 + \beta_1 x_t + u_t)$ with OLS

```
reg_est ← lm(y ~ x, data = ar_df)
```

Step 2: Add the residuals to our dataset

```
ar_df$e ← residuals(reg_est)
```

Testing for autocorrelation

Example: Static model and AR(3)

Step 1: Estimate the static model $(y_t = \beta_0 + \beta_1 x_t + u_t)$ with OLS

```
reg_est ← lm(y ~ x, data = ar_df)
```

Step 2: Add the residuals to our dataset

```
ar_df$e ← residuals(reg_est)
```

Step 3: Regress the residual on its lag (no intercept)

```
reg_ar3 ← lm(e ~ -1 + lag(e) + lag(e, 2) + lag(e, 3), data = ar_df)
```

Testing for autocorrelation

Example: Static model and AR(3)

Step 1: Estimate the static model $(y_t = \beta_0 + \beta_1 x_t + u_t)$ with OLS

```
reg_est ← lm(y ~ x, data = ar_df)
```

Step 2: Add the residuals to our dataset

```
ar_df$e ← residuals(reg_est)
```

Step 3: Regress the residual on its lag (no intercept)

```
reg_ar3 ← lm(e ~ -1 + lag(e) + lag(e, 2) + lag(e, 3), data = ar_df)
```

Note: `lag(var, n)` from `dplyr` takes the `n`$^{\text{th}}$ lag of the variable `var`.

Testing for autocorrelation

Example: Static model and AR(3)

Step 4: Calculate the $\text{LM} = n \times R_e^2$ test statistic—distributed χ_k^2.
k is the number of regressors in the regression in *Step 3* (here, $k = 3$).

```
# Grab R squared
r2_e ← summary(reg_ar3)$r.squared
# Calculate the LM test statistic: n times r2_e
(lm_stat ← 100 * r2_e)
```

```
#> [1] 72.38204
```

```
# Calculate the p-value
(pchisq(q = lm_stat, df = 3, lower.tail = F))
```

```
#> [1] 1.318485e-15
```

Testing for autocorrelation

Example: Static model and AR(3)

Step 5: Conclude.

Example: Static model and AR(3)

Step 5: Conclude.

Recall: Our hypotheses consider the model

$$e_t = \rho_1 e_{t-1} + \rho_2 e_{t-2} + \rho_3 e_{t-3}$$

Example: Static model and AR(3)

Step 5: Conclude.

Recall: Our hypotheses consider the model

$$e_t = \rho_1 e_{t-1} + \rho_2 e_{t-2} + \rho_3 e_{t-3}$$

which we are actually using to learn about the model

$$u_t = \rho_1 u_{t-1} + \rho_2 u_{t-2} + \rho_3 u_{t-3}$$

Testing for autocorrelation

Example: Static model and AR(3)

Step 5: Conclude.

Recall: Our hypotheses consider the model

$$e_t = \rho_1 e_{t-1} + \rho_2 e_{t-2} + \rho_3 e_{t-3}$$

which we are actually using to learn about the model

$$u_t = \rho_1 u_{t-1} + \rho_2 u_{t-2} + \rho_3 u_{t-3}$$

H_0: $\rho_1 = \rho_2 = \rho_3 = 0$ *vs.* H_A: $\rho_j \neq 0$ for at least one j in $\{1,\ 2,\ 3\}$

Our p-value is less than 0.05.

Testing for autocorrelation

Example: Static model and AR(3)

Step 5: Conclude.

Recall: Our hypotheses consider the model

$$e_t = \rho_1 e_{t-1} + \rho_2 e_{t-2} + \rho_3 e_{t-3}$$

which we are actually using to learn about the model

$$u_t = \rho_1 u_{t-1} + \rho_2 u_{t-2} + \rho_3 u_{t-3}$$

H_0: $\rho_1 = \rho_2 = \rho_3 = 0$ *vs.* H_A: $\rho_j \neq 0$ for at least one j in $\{1,\ 2,\ 3\}$

Our p-value is less than 0.05. **Reject $\mathbf{H_0}$.**

Testing for autocorrelation

Example: Static model and AR(3)

Step 5: Conclude.

Recall: Our hypotheses consider the model

$$e_t = \rho_1 e_{t-1} + \rho_2 e_{t-2} + \rho_3 e_{t-3}$$

which we are actually using to learn about the model

$$u_t = \rho_1 u_{t-1} + \rho_2 u_{t-2} + \rho_3 u_{t-3}$$

H_0: $\rho_1 = \rho_2 = \rho_3 = 0$ *vs.* H_A: $\rho_j \neq 0$ for at least one j in $\{1, 2, 3\}$

Our p-value is less than 0.05. **Reject H_0.**

Conclude there is statistically significant evidence of autocorrelation.

Testing for autocorrelation

Dynamic models with lagged outcome variables

Table of contents