Modeling Change over Time

Day 1
Foundations of Growth Curve Models
June 22, 2022
Michael Bader

Objectives

1. Describe how latent growth, latent growth trajectory analysis, and growth mixture models are related to, and different from, one another

2. Interpret results all three types of growth models

3. Identify the proper modeling technique for analytical questions regarding change

4. Clearly articulate the benefits, assumptions, and shortcomings of different models of change

Models

A **simplified or idealized description or conception** of a particular system, situation, or process, often in mathematical terms, that is **put forward as a basis for theoretical or empirical understanding**, or for calculations, predictions, etc.; a conceptual or mental representation of something.

Oxford English Dictionary

It has been said that "all models are wrong but some models are useful." In other words, any model is at best a **useful fiction**—there never was, or ever will be, an exactly normal distribution or an exact linear relationship.

George E. P. Box

The Mean as Model

Mean as Model

$$\bar{y} = \frac{1}{N} \sum_{t=1}^{T} (y_t)$$

Mean as Model

Single error

$$y_t = \bar{y} - \epsilon_t$$

$$\epsilon_t = y_t - \bar{y}$$

Distribution across all of the errors

$$Var(y) = \frac{1}{N}\sum_{t=1}^{T}\epsilon_t^2 = \frac{1}{N}\sum_{t=1}^{T}(y_t - \bar{y})^2$$

$$\sigma_y = \sqrt{Var(y)} = \sqrt{\frac{1}{N}\sum_{t=1}^{T}\epsilon_t^2} = \sqrt{\frac{\sum_{t=1}^{T}(y_t - \bar{y})^2}{N}}$$

The Mean as Model

$$\underbrace{y_t}_{\text{outcome}} = \underbrace{\bar{y}}_{\text{deterministic}} \underbrace{+\epsilon_t,\ \epsilon_t \sim \mathcal{N}(0, \sigma^2)}_{\text{stochastic}}$$

- Outcome (y_i): the value observed on *y* for datum (observation) *i*

- **Deterministic** (ybar): is the best guess (by minimizing the error) of the data as a whole

- **Stochastic** (e_t): is the deviation of the mean from the observed outcome for datum (observation) *t*

Estimates describe **the data**
while the error describes **a datum**

Models

	Population	Sample
Deterministic	True relationship	Standard error (uncertainty of sample)
Stochastic	Natural variation	Estimate of natural variation (after accounting for uncertainty of sample)

Linear Regression of a Trend

Describe: Linear Trend Model

- Our outcome of interest changes at a constant rate over time and depends only on the amount of time that has elapsed

$$\underbrace{Y_t}_{\text{outcome}} = \underbrace{\beta_0 + \beta_1 \text{time}_t}_{\text{deterministic}} + \underbrace{\epsilon_t}_{\text{stochastic}}$$

Simulate: Linear Trend Model

(File: `01b_lineartrend_simulation.R`)

$$\underbrace{Y_t}_{\text{outcome}} = \underbrace{\beta_0 + \beta_1 \text{time}_t}_{\text{deterministic}} + \underbrace{e_t}_{\text{stochastic}}$$

N The number of time periods that we will generate

β_0 Initial (logged) value of home prices

β_1 Monthly increase in the (logged) value of home prices (equals the percentage increase per month)

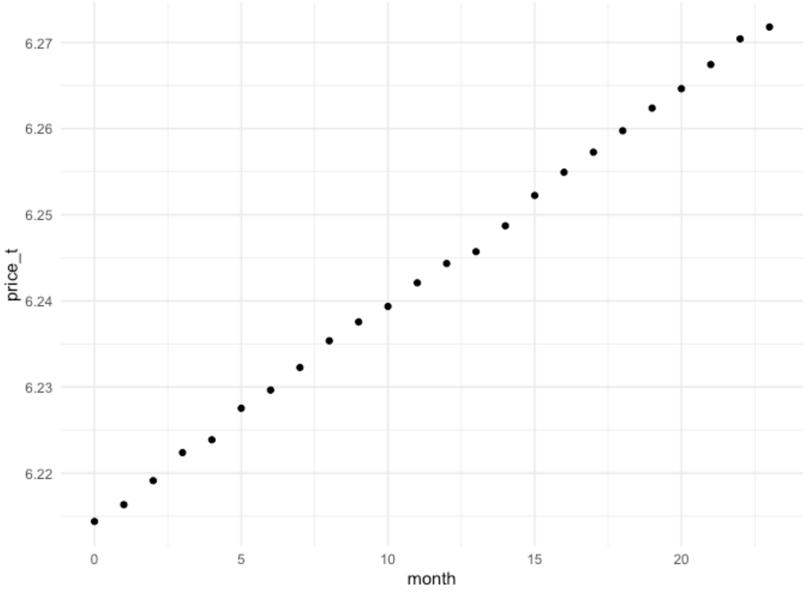
ϵ_t Error off of the predicted trend for an single month. We will assume that these are normally distributed with a mean of zero (definition of best estimate) and a standard deviation of σ. That means that we also need to define:

σ The standard deviation of the error off of the linear trend

Simulation of a Linear Trend

Simulation data:

```
## PLAN POPULATION
month <- c(0:23)
beta0 <- log(500)      ## $500/square foot
beta1 <- 0.03/12       ## 3% annual increase
sigma <- 0.03/48       ## 0.25% fluctuation off of
                       ##   trend in given month

## CONJURE POPULATION
epsilon_t = rnorm(24,0,sigma)
price_t <- beta0 + beta1*month + epsilon_t
```



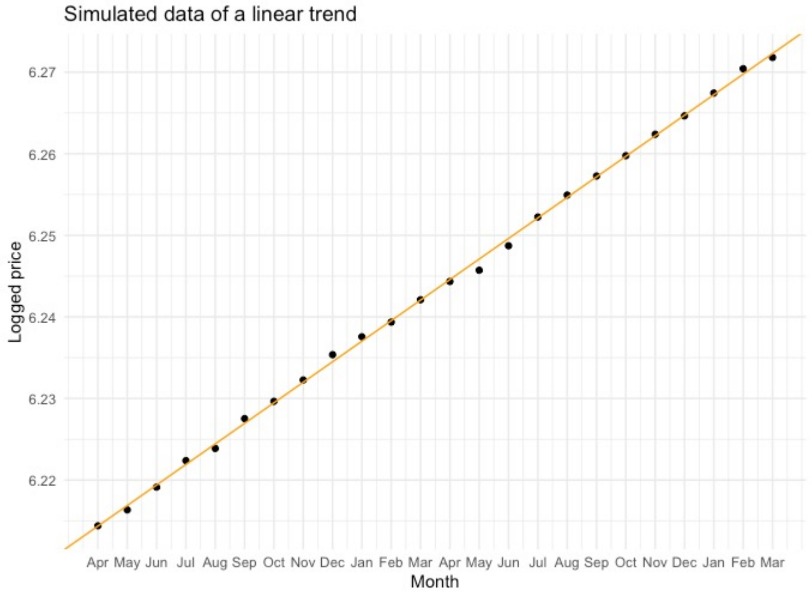
Simulation of a Linear Trend

Simulation data:

```
## PLAN POPULATION
month <- c(0:23)
beta0 <- log(500)       ## $500/square foot
beta1 <- 0.03/12        ## 3% annual increase
sigma <- 0.03/48        ## 0.25% fluctuation off of
                        ##   trend in given month

## CONJURE POPULATION
epsilon_t = rnorm(24,0,sigma)
price_t <- beta0 + beta1*month + epsilon_t
```

Simulation estimates:

```
Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept) 6.214e+00  2.068e-04 30044.6   <2e-16 ***
month       2.518e-03  1.541e-05   163.4   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.0005226 on 22 degrees of freedom
```



Analyze: Linear Trend Model

(File: `01b_lineartrend_analysis.R`)

Original data (wide):

	RegionID	RegionName	val.201803	val.201804	val.201805	val.201806	val.201807
2	394913	New York, NY	739	743	746	748	750

Reshaped data (long):

	RegionID	RegionName	yyyymm	value_t	month
1	394913	New York, NY	201803	739	0
2	394913	New York, NY	201804	743	1
3	394913	New York, NY	201805	746	2
4	394913	New York, NY	201806	748	3
5	394913	New York, NY	201807	750	4
6	394913	New York, NY	201808	752	5
7	394913	New York, NY	201809	754	6

Reshape data to be in long format and make sure to index time variable to something meaningful

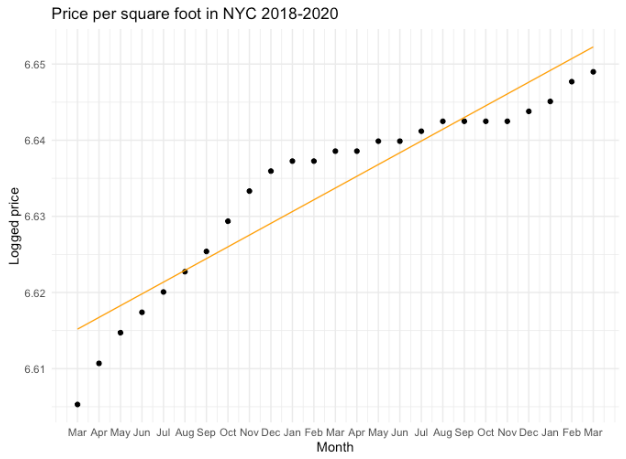
(intercept occurs where $t = 0$!)

Analyze: Linear Trend Model

(File: `01b_lineartrend_analysis.R`)

Let's interpret the results:

```
Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 6.6151878  0.0016867 3922.03   < 2e-16 ***
month       0.0015445  0.0001205   12.82 5.85e-12 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.004344 on 23 degrees of freedom
```



`(Intercept)` represents β_0 in our formal model. Its value of 6.615 means that the logged median home value of New York homes in March two years ago was 6.615, or homes sold for \$746.35 per square foot

`month` represents the β_1 in our formal model. Its value of 0.0015 means that home prices increased by about 0.15% per month from March two years ago to this March.

QUESTIONS?

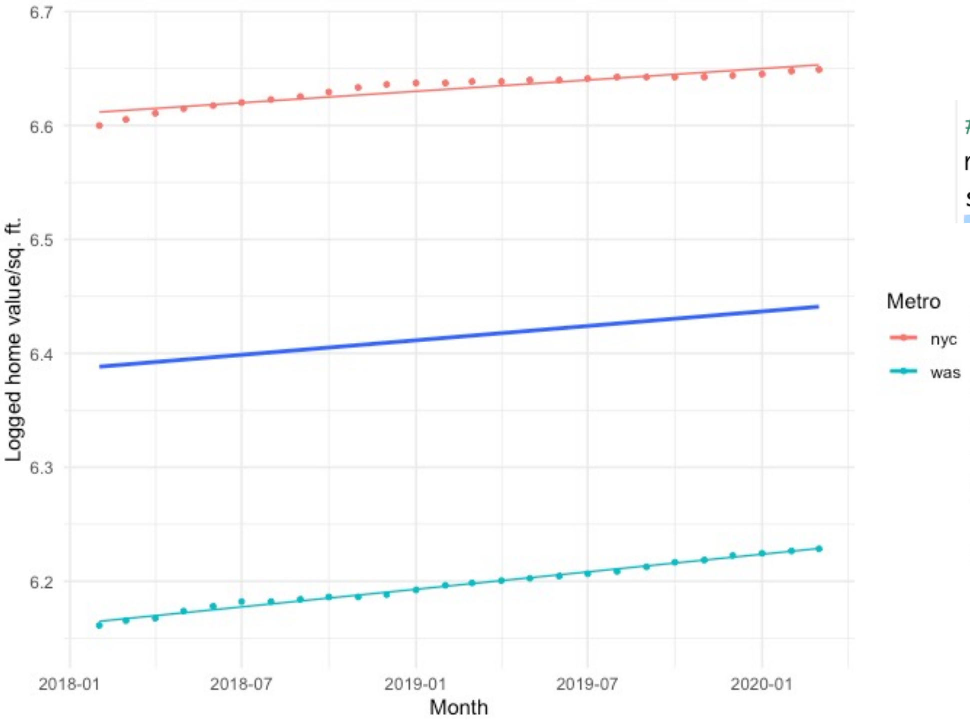
Random Intercept Models

Linear Trend Model

$$\ln(\text{value})_t = \beta_0 + \beta_1 (\text{month})_t + \epsilon_t$$

Analyze Two Trends

(File: `01b_twocity_analysis.R`)

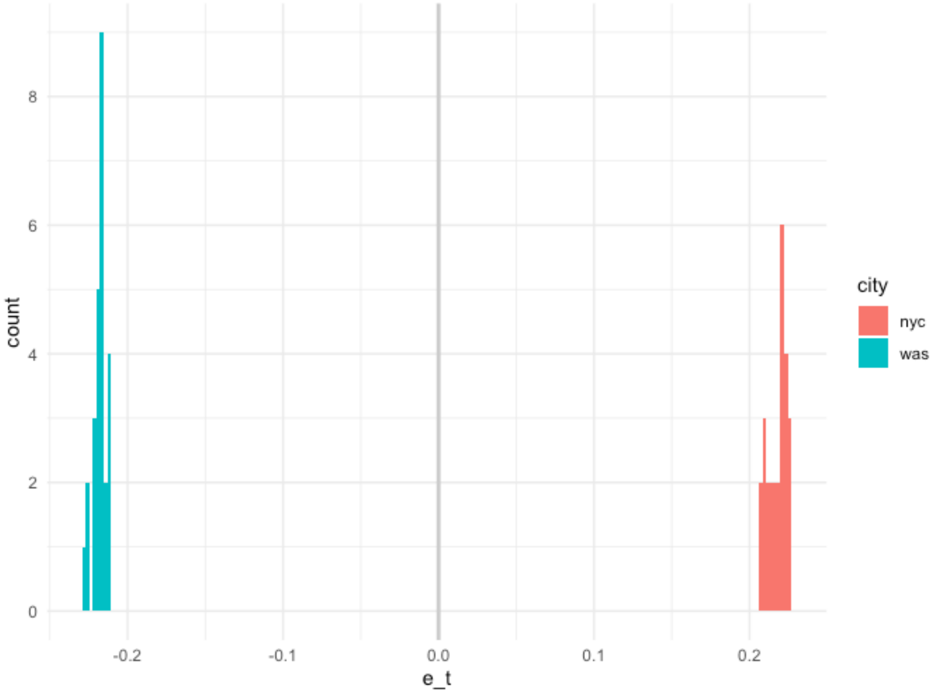



```r
## ANALYZE THE DATA
m.ana <- lm(lnvalue_t ~ datec, data=d.ana)
summary(m.ana)
```

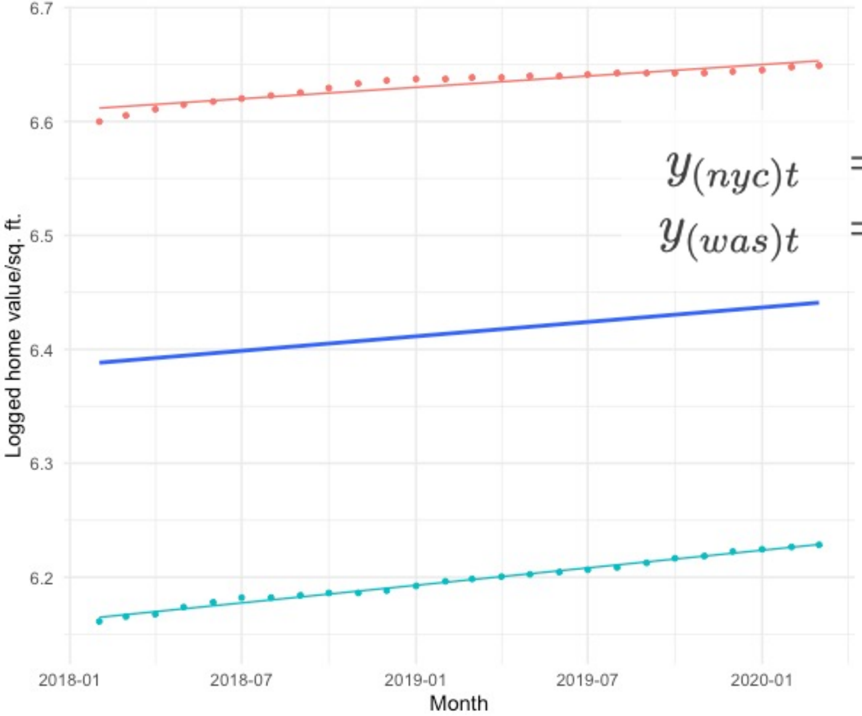
```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 6.388165   0.059896 106.655   <2e-16 ***
datec       0.002113   0.004109   0.514    0.609
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1

Residual standard error: 0.2222 on 50 degrees of freedom
```

Analyze Two Trends

Residuals from the combined model:



Two Linear Trends



$$y_{(nyc)t} = \beta_{0(nyc)} + \beta_{1(nyc)}(\text{month})_t + \epsilon_{(nyc)t}$$
$$y_{(was)t} = \beta_{0(was)} + \beta_{1(was)}(\text{month})_t + \epsilon_{(was)t}$$

Two Linear Trends

$$
\begin{aligned}
y_{(nyc)t} &= \beta_{0(nyc)} &+\beta_{1(nyc)}(\text{month})_t &+\epsilon_{(nyc)t} \\
y_{(was)t} &= \beta_{0(was)} &+\beta_{1(was)}(\text{month})_t &+\epsilon_{(was)t}
\end{aligned}
$$

Mean intercept model: $\beta_{0i} = \overline{\beta_0} + \rho_{0i}$

Mean intercept value: $\overline{\beta_0} = \dfrac{\beta_{0(ny)} + \beta_{0(was)}}{2} = \dfrac{\sum_{i=1}^{N} \beta_{0i}}{N}$

Random intercept model: $y_{it} = \overline{\beta_0} + \beta_1(\text{month})_t + \rho_{0i} + \epsilon_{it}$

$$\underbrace{y_{it}}_{\text{outcome}} = \underbrace{\overline{\beta_0} + \beta_1(\text{month})_t}_{\text{deterministic}} + \underbrace{\rho_{0i} + \epsilon_{it}}_{\text{stochastic}}$$

Let's define each of the terms more carefully:

y_{it} The outcome (median home value) for metropolitan area i (either NYC or WAS) at time t can be predicted by:

$\overline{\beta_0}$ the mean home value across metropolitan areas at time zero, plus

β_1 the mean increase per month in home values (assumed to be the same across metropolitan areas), **and**

ρ_{0i} the *individual* variation of initial values in metropolitan area i (either NYC or WAS) from the average initial value, and

ϵ_{it} the individual deviation of month t from the trend due to normal month-to-month variation *within* each metropolitan area i

Random Intercept Model

Single-line equation:
(from above)

$$y_{it} = \overline{\beta_0} + \beta_1(\textsf{month})_t + \rho_{0i} + \epsilon_{it}$$

Multiple-line equation:

$$\begin{aligned} y_{it} &= \beta_{0i} + \beta_1(\textsf{month})_t + \epsilon_i t \\ \beta_{0i} &= \gamma_{00} + \rho_{0i} \end{aligned}$$

Multiple-line equation:
(with errors)

$$\begin{aligned} y_{it} &= \beta_{0i} + \beta_1(\textsf{month})_t + \epsilon_{it}, & \epsilon_{it} &\sim \mathcal{N}(0, \sigma) \\ \beta_{0i} &= \gamma_{00} + \rho_{0i}, & \rho_{0i} &\sim \mathcal{N}(0, \tau_0) \end{aligned}$$

Random Intercept Model

$$\underbrace{y_{it}}_{\text{outcome}} = \underbrace{\gamma_{00} + \beta_1(\text{month})_t}_{\text{deterministic}} + \underbrace{\rho_{0i} + \epsilon_{it}}_{\text{stochastic}}$$

γ_{00} , the grand mean initial value when we start the study (mean of the intercepts)

β_1 , the change in home value, which we assume to be the same across all metros

τ_0 , the standard deviation of intercepts away from the mean intercept across metros

σ , the deviation of each month's observation away from the line of best fit for that metropolitan area

Random Intercept Model

Useful for describing change across entities (e.g., participants) that **start at different values of the outcome** from one another and then **change at the same rate** as one another.

A note on terminology

You will often see the term “fixed effects” and “random effects.” In the context of these models:

“**fixed effects**” refer to the **deterministic** parameters
“**random effects**” refer to the **stochastic** parameters

BUT, “fixed effects” can often be confused with econometric models that model the first-difference across groups (i.e., add a dummy variable for *N-1* groups)

Random Intercept Model

Simulation Example

1. Plan the Population
   - Number of repeated observations
   - Initial level of outcome
   - Variation in initial level of outcome across cities
   - Change per unit time
   - Time-to-time fluctuation off of the trend
2. Conjure the population into existence
3. Analyze the population we created

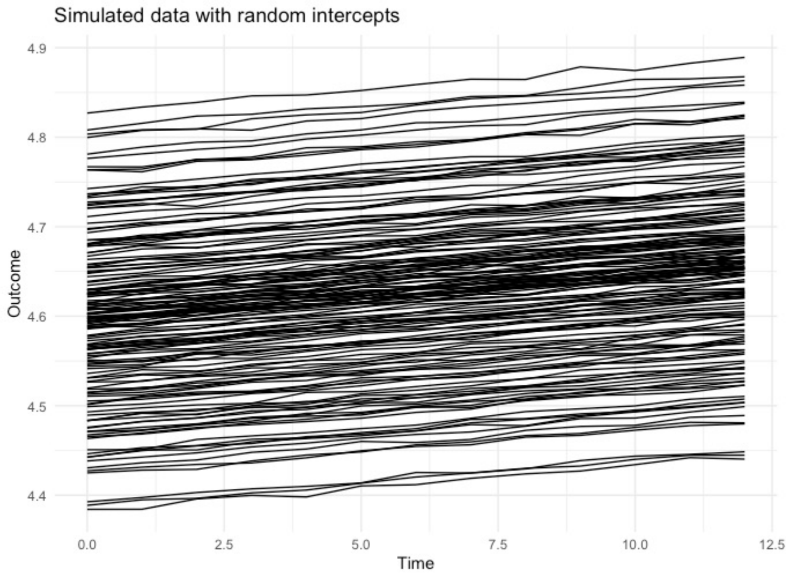
File: `01d_random_intercepts_simulation.R`

Random Intercept Model

Plan simulated population parameters & conjure population

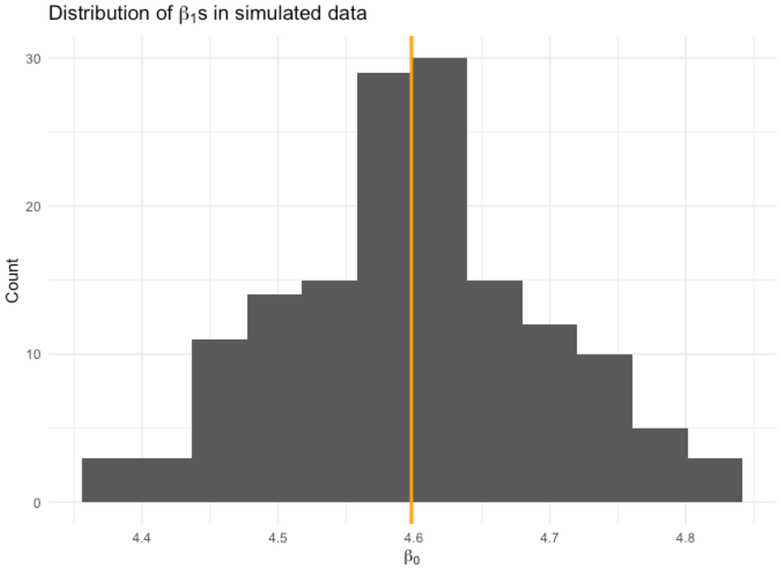
```r
## PLAN POPULATION
month <- c(0:12)
N <- 150
N_t <- length(month)
t <- rep(month, N)
## Create an indicator for each metro area
i <- as.factor(rep(c(1:N), each=N_t))

sigma_it <- 0.002
tau_0i <- 0.10

beta_0  <- log(100) + rnorm(N,0,tau_0i)
beta_1  <- 0.005

## CONJURE POPULATION
beta_0i <- rep(beta_0,each=N_t)
lnvalue_it <- beta_0i + beta_1*t + rnorm(N*N_t,0,sigma_it)
d.sim <- data.frame(i,t,lnvalue_it)
```

Simulated data of 150 trends



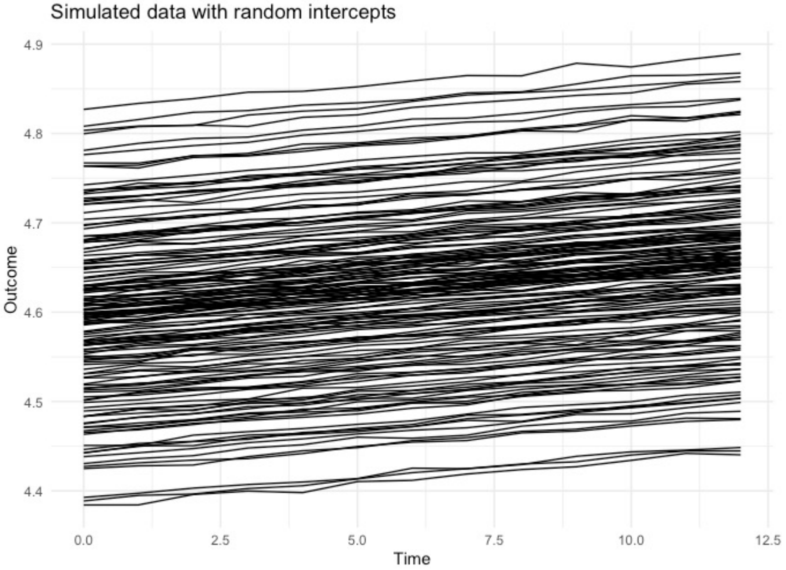
Random Intercept Model

Estimate intercept for all 150 simulated metros (150 regressions)



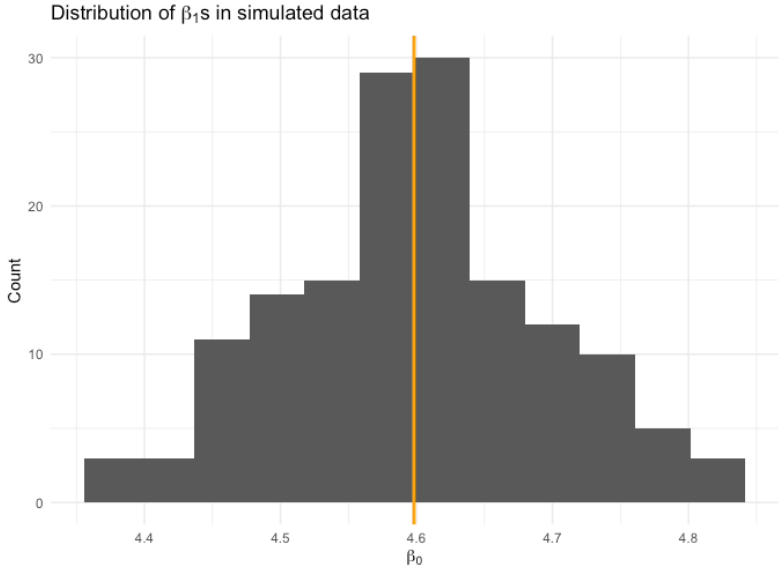
Mean = 4.6; σ = 9.4

Simulated data of 150 trends

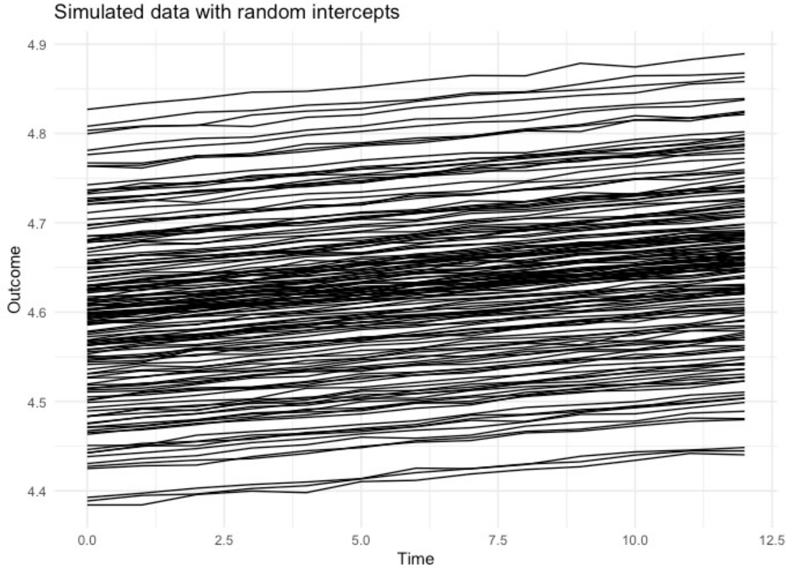


Random Intercept Model

Estimate intercept for all 150 simulated metros (150 regressions)



Simulated data of 150 trends



Mean = 4.6; $\sigma = 9.4$ } Assumes β_{0i} is "Truth" for each i

Random Intercept Model

$$\underbrace{y_{it}}_{\text{outcome}} = \underbrace{\gamma_{00} + \beta_1 (\text{month})_t}_{\text{deterministic}} + \underbrace{\rho_{0i} + \epsilon_{it}}_{\text{stochastic}}$$

γ_{00} , the grand mean initial value when we start the study (mean of the intercepts)

β_1 , the change in home value, which we assume to be the same across all metros

τ_0 , the standard deviation of intercepts away from the mean intercept across metros

σ , the deviation of each month's observation away from the line of best fit for that metropolitan area

We need to *jointly* estimate parameters!!!

Random Intercept Model

Analysis Example

1. Gather the data
2. Describe the data
3. Estimate the model using our data
4. Interpret the estimates

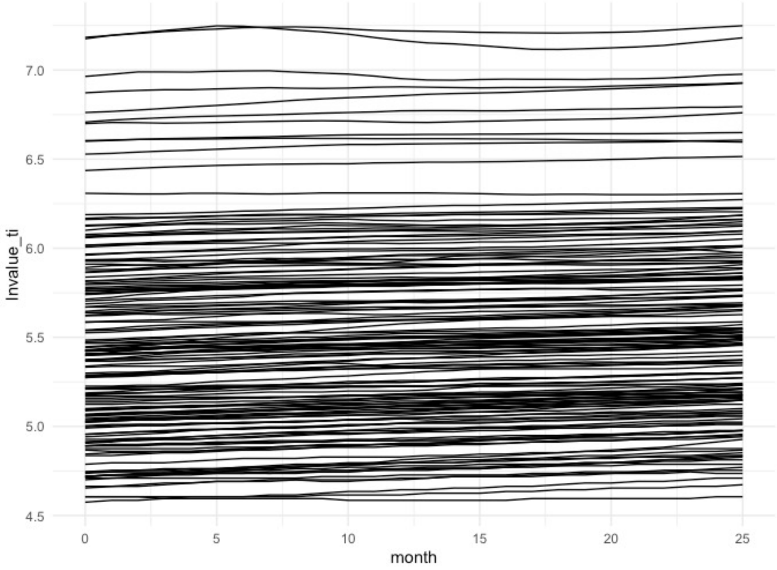
File: `01d_randomintercepts_analysis.R`

Random Intercept Model

Gather the data

```r
## GATHER THE DATA
load('../Data/zillow_long.RData')

## Subtract minimum value of month to set intercept to t=0
zillow.long <- zillow.long %>%
    mutate(
        month = month - min(month),
        date = ymd(paste0(yyyymm, "01")),
        lnvalue_ti = lnvalue_t, ## Helps us remember that
                                ## value is per month/per metro
        )
```

Describe the data



Random Intercept Model

Estimate the model

```
## ANALYZE THE DATA
m.ana <- lmer(
    lnvalue_ti ~ month + (1 | RegionID),
    data=zillow.long
)
summary(m.ana)
```

```
Random effects:
 Groups   Name        Variance  Std.Dev.
 RegionID (Intercept) 0.3088825 0.55577
 Residual             0.0002164 0.01471
Number of obs: 3900, groups:  RegionID, 150

Fixed effects:
             Estimate Std. Error t value
(Intercept) 5.484e+00  4.538e-02   120.8
month       3.380e-03  3.141e-05   107.6
```

Interpret the results

$$y_{it} = \gamma_{00} + \beta_1(\text{month})_t + \rho_{0i} + \epsilon_{it}$$

$\gamma_{00} = 5.484$ The average (logged) median home value in the largest 150 metro areas in March 2018 ago was 5.484, or \$241 (i.e., $e^{5.484}$)

$\beta_1 = 0.0034$ The average metropolitan area home values increased by 0.0034 per month, or 0.34% per month

$\tau_0 = 0.556$ Home values in March of 2018 varied with a standard deviation of 55.60%

$\sigma = 0.015$ Home values from month to month fluctuated off of the metropolitan specific trend by about 1.50%

QUESTIONS?

Random Slope Model

Random Slope Model

Single-line equation:

$$y_{ti} = \beta_0 + (\gamma_{10} + \rho_{1i})\,(\textsf{month})_t + \epsilon_{ti}$$

Multiple-line equation:

$$y_{ti} = \beta_0 + \beta_{1i}(\textsf{month})_t + e_{ti}$$
$$\beta_{1i} = \gamma_{10} + \rho_{1i}$$

Multiple-line equation:
(with errors)

$$\begin{aligned} y_{ti} &= \beta_0 + \beta_{1i}(\textsf{month})_t + \epsilon_{ti}, & \epsilon_{it} \sim \mathcal{N}(0, \sigma) \\ \beta_{1i} &= \gamma_{10} + \rho_{1i}, & \rho_{1i} \sim \mathcal{N}(0, \tau_1) \end{aligned}$$

Random Slope Model

$$y_{ti} = \beta_0 + \gamma_{10}(\text{month})_t + \rho_{1i}(\text{month})_t + \epsilon_{ti}$$

y_{ti} The outcome (median home value) for metropolitan area i at time t, which can be estimated by:

β_0 the mean home value at time 0 (assumed to be the same across metropolitan areas),

γ_{10} the mean increase per month in home values in metropolitan area i over the period $t = \{0, 1, \ldots, T\}$, **and**

ρ_{1i} the *individual* variation of the slope in metropolitan area i from the average metropolitan rate of change, and

ϵ_{it} the individual deviation of month t from the trend due to normal month-to-month variation *within* each metropolitan area i

Random Slope Model

Draw three lines, one each under the **outcome**, **deterministic**, and **stochastic** components of the model below

$$y_{ti} = \beta_0 + \gamma_{10}(\text{month})_t + \rho_{1i}(\text{month})_t + \epsilon_{ti}$$

Random Slope Model

Useful for describing change across entities (e.g., participants) that **start at similar values of the outcome** to one another and then **change at the different rates** from one another.

Random Slope Model

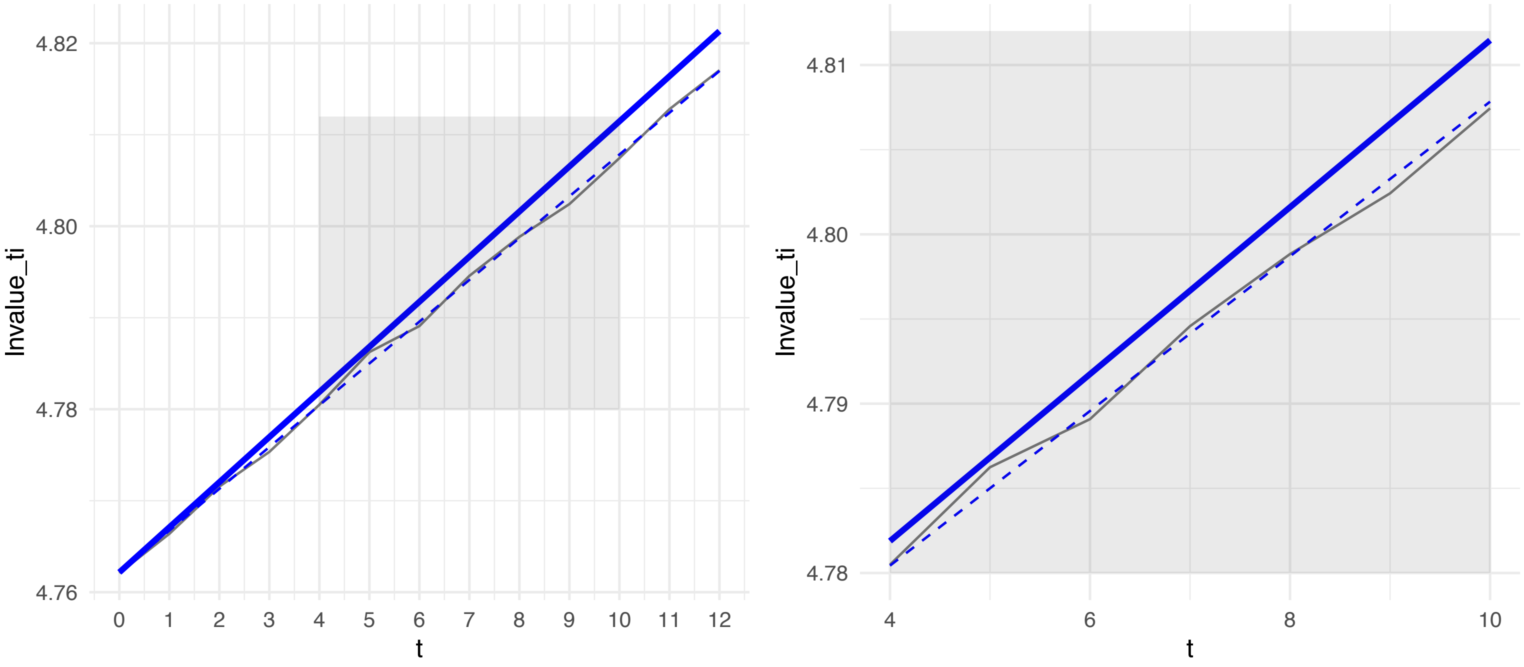
Simulation Example

1. Plan the Population
    - Number of repeated observations
    - Initial level of outcome
    - Change per unit time
    - Variation in the change per unit time
    - Time-to-time fluctuation off of the trend
2. Conjure the population into existence
3. Analyze the population we created

File: `01e_random_slope_simulation.R`

Random Slope Model

Sources of error in simulated data



QUESTIONS?

Random Intercept & Slope

Random Intercept & Slope Model

$$
\begin{aligned}
y_{ti} &= \beta_{0i} + \beta_{1i}(\text{month})_t + \epsilon_{ti} \\
\beta_{0i} &= \gamma_{00} + \rho_{0i} \\
\beta_{1i} &= \gamma_{10} + \rho_{1i}
\end{aligned}
$$

- Write the outcome, deterministic, and stochastic components of this model

$$
\underbrace{y_{ti}}_{\text{outcome}} = \underbrace{\gamma_{00} + \gamma_{10}(\text{month})_t}_{\text{deterministic}} + \underbrace{\rho_{0i} + \rho_{1i}(\text{month})_t + \epsilon_{ti}}_{\text{stochastic}}
$$

Random Intercept and Slope Model

Useful for describing change across entities (e.g., participants) that **start at different values of the outcome** from one another ***and*** then **change at the different rates** from one another.

How many parameters does this model have?

$$\underbrace{y_{ti}}_{\text{outcome}} = \underbrace{\gamma_{00} + \gamma_{10}(\text{month})_t}_{\text{deterministic}} + \underbrace{\rho_{0i} + \rho_{1i}(\text{month})_t + \epsilon_{ti}}_{\text{stochastic}}$$

Random Intercept & Slope Model

$$y_{ti} = \beta_{0i} + \beta_{1i}(\text{month})_t + \epsilon_{ti}$$
$$\beta_{0i} = \gamma_{00} + \rho_{0i}$$
$$\beta_{1i} = \gamma_{10} + \rho_{1i}$$

where $e_{ti} \sim \mathcal{N}(0, \sigma)$

and $\mathbf{T} \sim \mathcal{MVN}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \tau_{00} & \\ \tau_{01} & \tau_{11} \end{bmatrix}\right)$

Unstructured:

$$\mathbf{T} = \begin{bmatrix} \tau_{00} & \tau_{01} \\ \tau_{10} & \tau_{11} \end{bmatrix}$$

Diagonal:

$$\mathbf{T} = \begin{bmatrix} \tau_{00} & 0 \\ 0 & \tau_{11} \end{bmatrix}$$

Random Intercept & Slope Model

Analysis Example

1. Gather our data
2. Describe our data
3. Analyze our data
4. Interpret the data

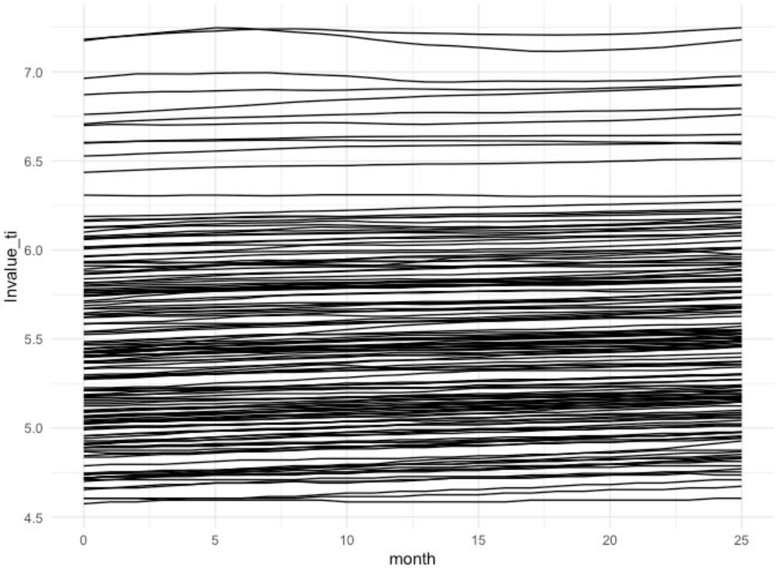
File: `01f_randominterceptslope_analysis.R`

Random Intercept & Slope Model

Gather the data

```r
## GATHER THE DATA
load('../Data/zillow_long.RData')

## Subtract minimum value of month to set intercept to t=0
zillow.long <- zillow.long %>%
    mutate(
        month = month - min(month),
        date = ymd(paste0(yyyymm, "01")),
        lnvalue_ti = lnvalue_t, ## Helps us remember that
                                ## value is per month/per metro
        )
```

Describe the data



Random Intercept & Slope Model

Estimate the model

```
## ANALYZE THE DATA
m.ana <- lmer(
    lnvalue_ti ~ month + (1 + month | RegionID),
    data=zillow.long
)
summary(m.ana)
```

```
Warning messages:
1: In checkConv(attr(opt, "derivs"), opt$par, ctrl = control$checkConv,  :
  unable to evaluate scaled gradient
2: In checkConv(attr(opt, "derivs"), opt$par, ctrl = control$checkConv,  :
  Model failed to converge: degenerate  Hessian with 1 negative eigenvalues
```

Random Intercept & Slope Model

Estimate the model

```r
## ANALYZE THE DATA
m.ana <- lmer(
    lnvalue_ti ~ month + (1 + month | RegionID),
    data=zillow.long
)
summary(m.ana)
```

```
 Groups   Name        Variance  Std.Dev. Corr
 RegionID (Intercept) 3.189e-01 0.564728
          month       3.030e-06 0.001741 -0.43
 Residual             4.026e-05 0.006345
Number of obs: 3900, groups:  RegionID, 150

Fixed effects:
             Estimate Std. Error t value
(Intercept) 5.4841554  0.0461103  118.94
month       0.0033800  0.0001428   23.67
```

- **Fixed effects:**
  - **(Intercept)***: Reports the value of the model intercept, γ_{00}
  - **t***: Reports the value of the change per unit time, γ_{10}
- **Random effects:**
  - **Groups**: This specifies how observations were grouped within units; in our case, we have random components based on unit
  - **(Intercept)***: Reports the variance of metropolitan-specific intercepts around the mean intercept (τ_{00})
  - **t***: Reports the variance of the metropolitan-specific slopes around the mean slope (τ_{11})
  - **Residual***: Reports the variation off of the metropolitan-specific trend for month t in metropolitan area i
  - **Corr***: Reports the correlation between the two metropolitan-specific stochastic terms (related to τ_{01})

Random Intercept & Slope

Write the estimated values of the parameters:

γ_{00}

γ_{10}

τ_{00}

τ_{11}

τ_{01}

σ

Random Intercept & Slope

Interpreting the results

Across the largest 150 metropolitan areas, the median home price per square foot in March 2018 was $241, but logged prices varied across metro areas with a standard deviation of 0.56. During the next two years, median home prices increased, on average, by 0.338% per month. Monthly appreciation in home prices varied around the average increase by 0.17% percent. The negative correlation between the random terms of the intercept and slope means that increases in home values were smaller in metropolitan areas that started with higher median values.

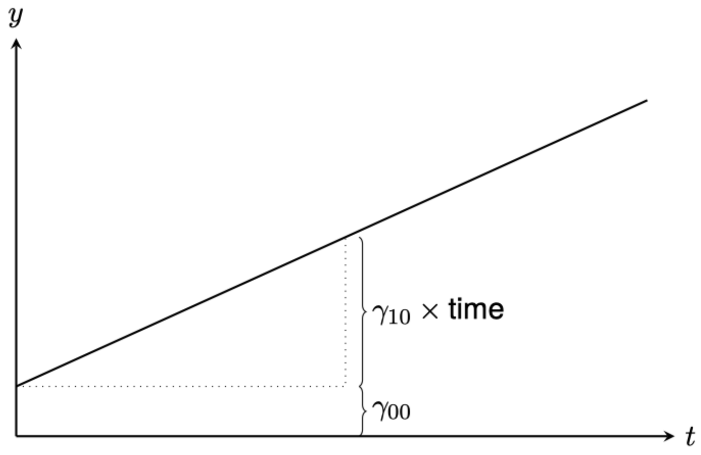
QUESTIONS?

Modeling Trends & Measuring Time

Measuring Time

- Demographic Time
  - Age
  - Period
  - Cohort (Panel)
- Event Time (“Synthetic cohorts”)
- Centering time on meaningful measure

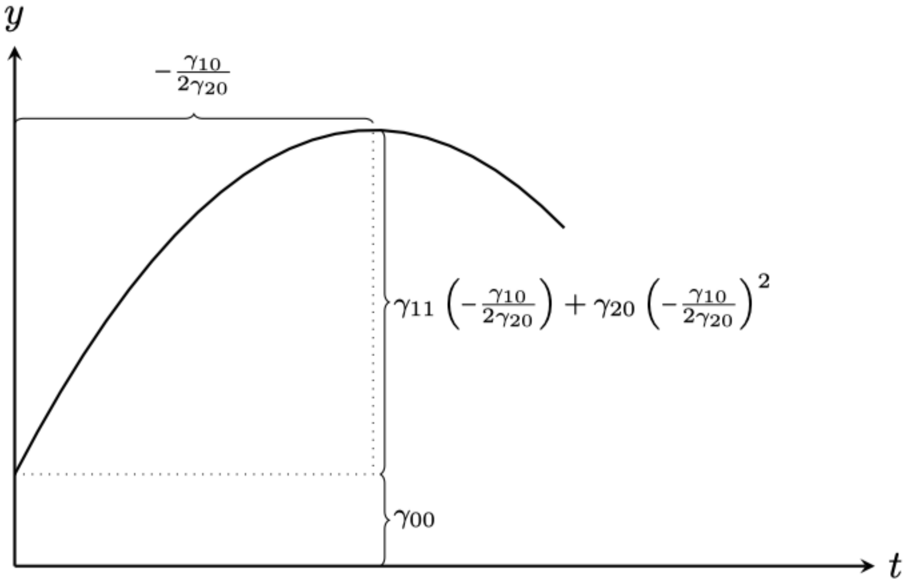
Linear Trend



This is the model that we have constructed to date:

$$\underbrace{y_{ti}}_{\text{outcome}} = \underbrace{\gamma_{00} + \gamma_{10}(\text{time})_{ti}}_{\text{deterministic}} + \underbrace{\rho_{0i} + \rho_{1i}(\text{time})_{ti} + \epsilon_{ti}}_{\text{stochastic}}$$

Quadratic Trend



$$\underbrace{y_{ti}}_{\text{outcome}} = \underbrace{\gamma_{00} + \gamma_{10}(\text{time})_{ti} + \gamma_{20}(\text{time}^2_{ti})}_{\text{deterministic}} + \underbrace{\rho_{0i} + \rho_{1i}(\text{time}_{ti}) + \rho_{2i}(\text{time}^2_{ti}) + \epsilon_{ti}}_{\text{stochastic}}$$

Quadratic Trend

$$\underbrace{y_{ti}}_{\text{outcome}} = \underbrace{\gamma_{00} + \gamma_{10}(\text{time})_{ti} + \gamma_{20}(\text{time}_{ti}^2)}_{\text{deterministic}} + \underbrace{\rho_{0i} + \rho_{1i}(\text{time}_{ti}) + \rho_{2i}(\text{time}_{ti}^2) + \epsilon_{ti}}_{\text{stochastic}}$$

- How many parameters does the squared term add to the model?

→ Four:

The ρ components are multivariate normally distributed, i.e., $(\rho_{0i}, \rho_{1i}, \rho_{2i}) \sim \mathcal{MVN}(\mathbf{0}, \mathbf{T})$, where:

$$\mathbf{T} = \begin{bmatrix} \tau_{00} & & \\ \tau_{10} & \tau_{11} & \\ \tau_{20} & \tau_{21} & \tau_{22} \end{bmatrix}$$

Other Parametric Trends

- **Other polynomials.** You can also model other polynomials of change by adding the appropriate terms. Using basic algebra, you can solve for inflection points, local and global maxima/mimima, and other values that might be of interest. In order to keep things

- **Exponential growth.** You can also model exponential growth by modeling the change in values (this is, in fact, what we are doing by using the natural log of the median home value).

Independent Variables

Independent Variables

- **Time invariant variables** that remain constant within an individual over time
- **Time varying variables** that can differ over time within an individual

Independent Variables: Time Invariant

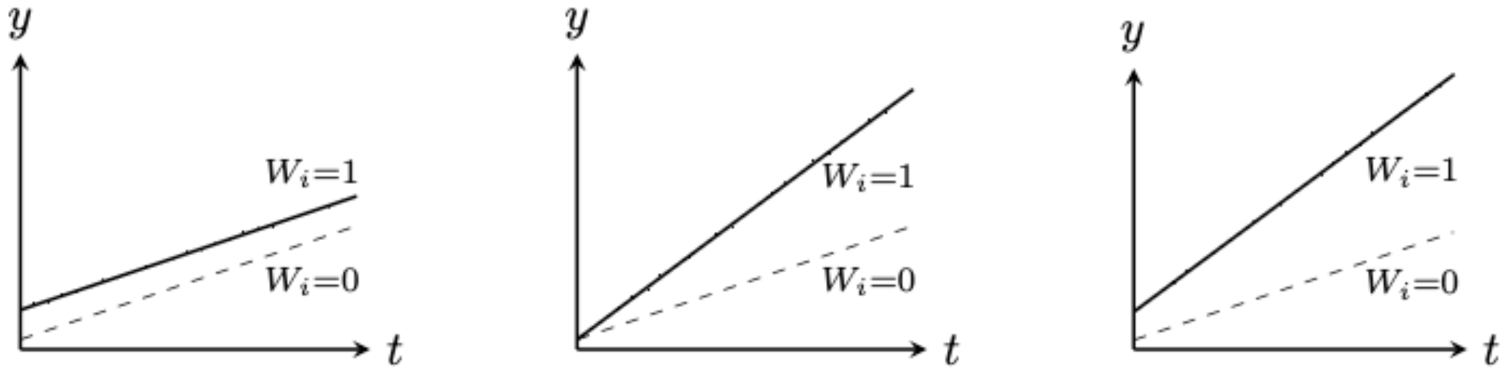
- Does the severity of depression over time differ between males and females?
  - **Outcome**: Depression severity
  - **Independent variable**: Sex (categorical)
- Does change in pulmonary function differ between never-smokers and ever-smokers?
  - **Outcome**: Pulmonary function
  - **Independent variable**: Ever-smoker (categorical)
- Does the probability of recidivism increase at different rates depending on age of first incarceration?
  - **Outcome**: Probability of recidivism (would use generalized linear model with a logit or probit link function)
  - **Independent variable**: Age of first incarceration (continous)

Independent Variables: Time Invariant

Do we think that:

- The trait leads some group to have higher values at the outset than another group, but both groups generally change at a similar rate?
- The trait leads groups to change at different rates after having no meaningful difference in where they started?
- The trait is associated with **both** different starting values and different rates of change?

Independent Variables: Time Invariant



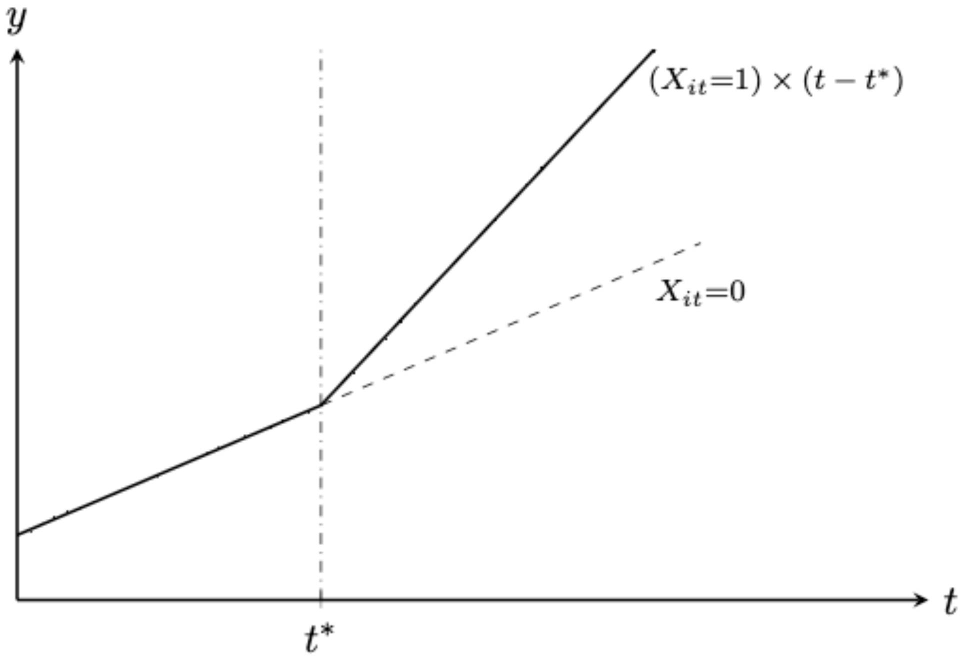
(a) Different starting values (b) Different rates of change (c) Different starting values **and** rates of change

Independent Variables: Time Varying

1. Altered rate of change
2. Deviation in level
3. Both altered rate of change **and** deviation in level

(we will go over *how* to model these independent variables tomorrow)

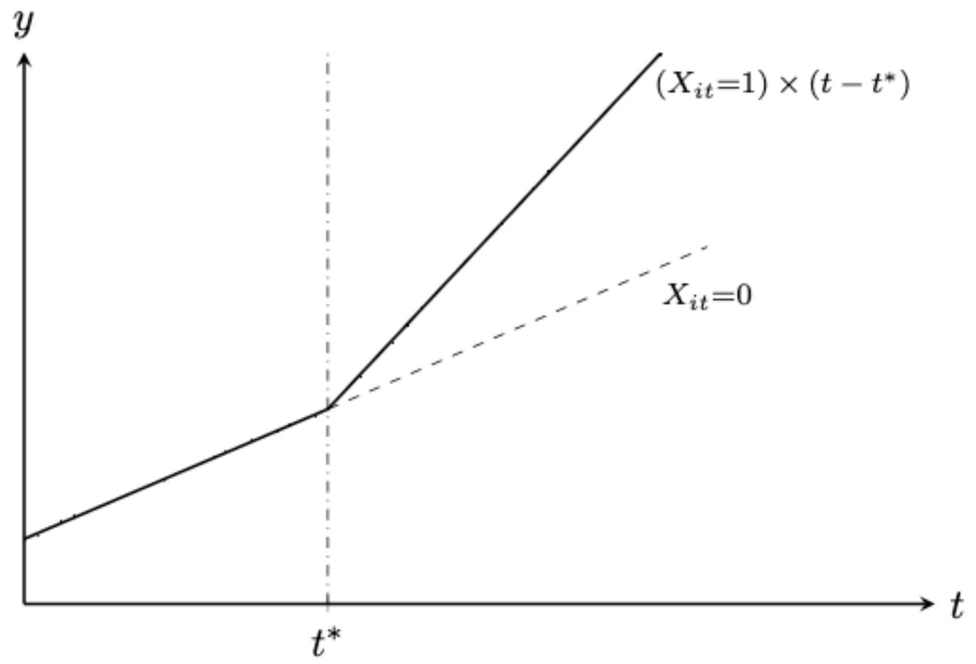
Altered Rate of Change



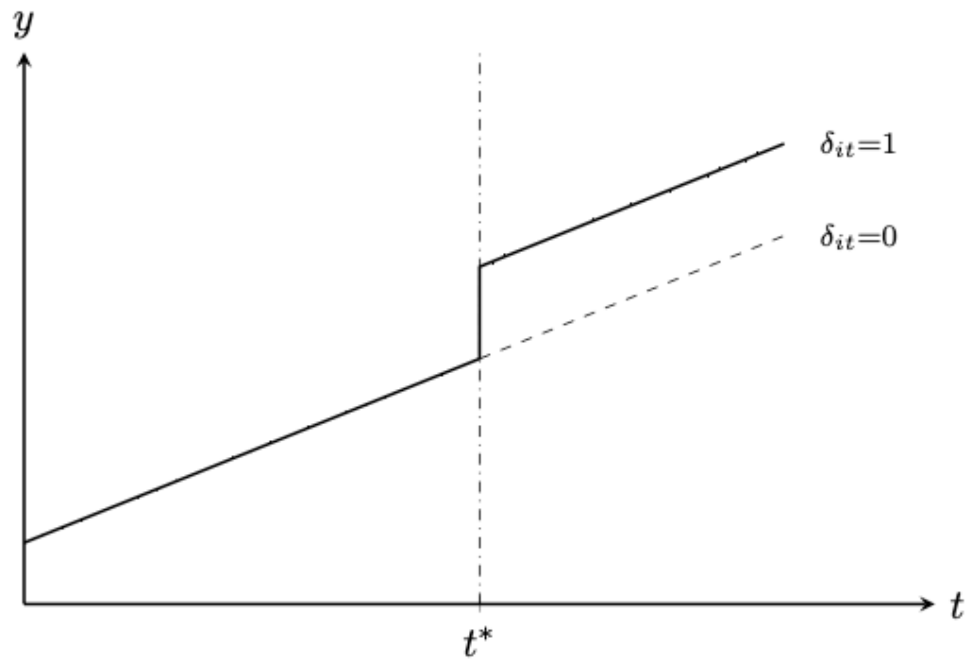
Altered Rate of Change

- Does a physical therapy intervention increase mobility improvements of leg amputees compared to a group that does not receive the intervention?
  - **Outcome**: Mobility
  - **Independent variable**: Time since physical therapy intervention
- Does a new drug lower depressive symptoms faster than an existing drug?
  - **Outcome**: Depressive symptoms
  - **Independent variable**: Time since onset of medication
- Does the rate of language acquisition increase for children after they enter early childhood programs?
  - **Outcome**: Language acquisition
  - **Independent variable**: Time since entering early childhood program

Altered Rate of Change



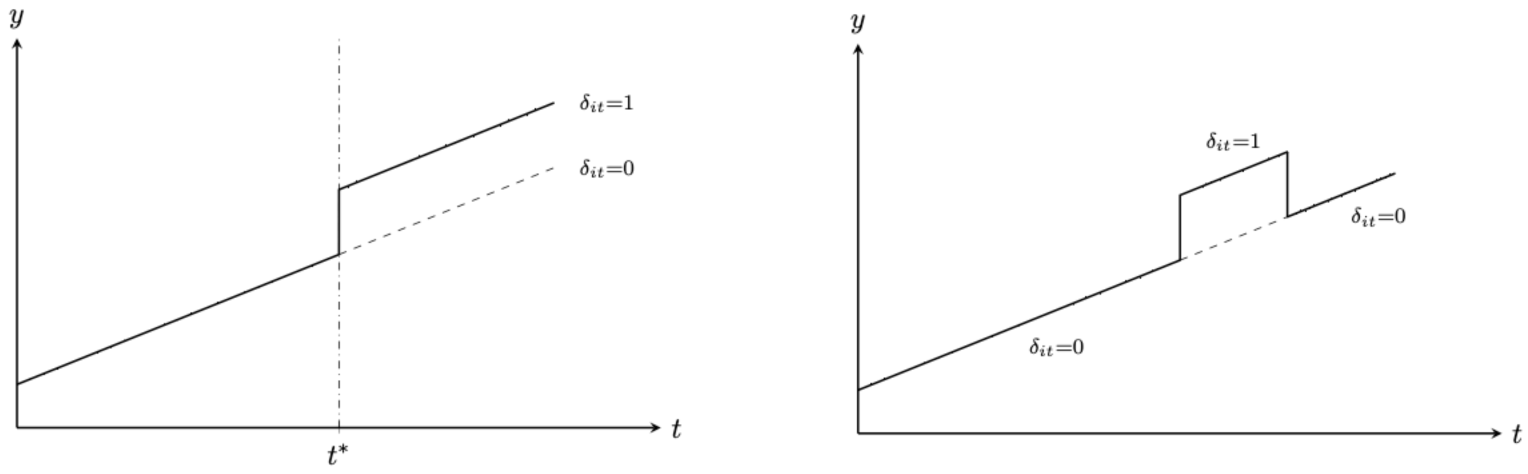
Deviation in Level



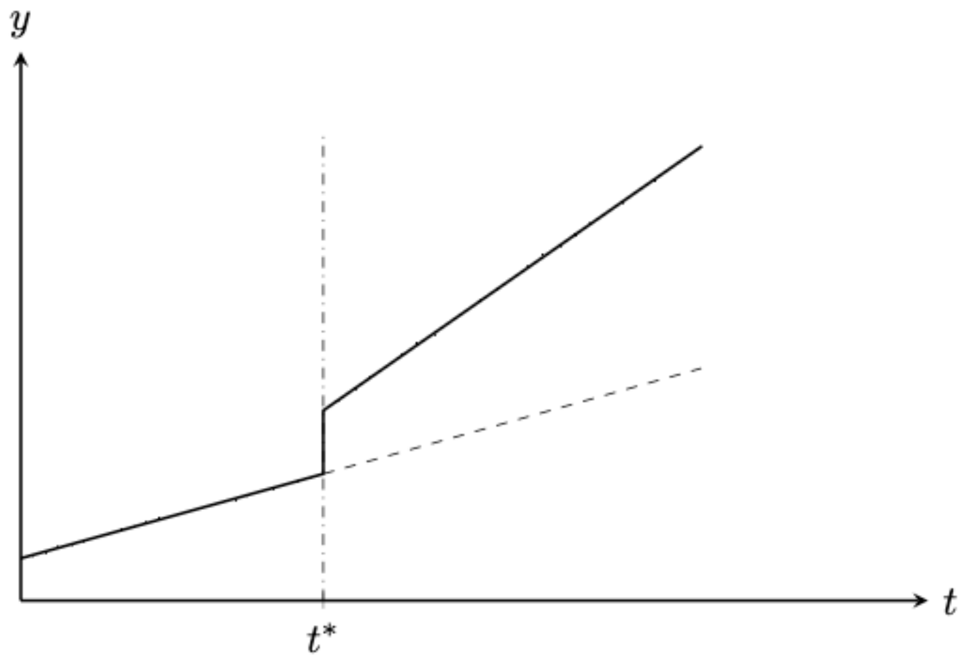
Deviation in Level

- Does hypertension decrease when a patient is compliant with treatment?
  - **Outcome**: Hypertension (blood pressure)
  - **Independent variable**: Patient compliant with antihypertensive medication
- The probability of moving might be higher in years where a member of the household changes jobs
  - **Outcome**: Probability of moving
  - **Independent variable**: Change of jobs
- A mental health care coordination program offered by an insurance provider reduces medical care expenditures on program participants
  - **Outcome**: Medical care expenditure
  - **Independent variable**: Participation in a coordinated care program

Deviation in Level



Both Altered Rate of Change and Deviation in Level



QUESTIONS?

Exercise: Describe Basic Model

1. Characterize your sample *relative to your population of interest*
   - Determine the group population (the values corresponding to *i*)
   - Determine how observations are sampled within each unit (the values corresponding to *t* in each group)
2. Generate a hypothesis that predicts how time relates to your outcome of interest
   - Determine your outcome and your *primary* independent variable (choose only one for right now)
   - Write a hypothesis for the relationship between your outcome and primary independent variable
   - Determine whether the measurement of the variable matches your hypothesis (i.e., time invariant or time-variant)
   - Explain where you expect *group-level variation* to exist in your data (intercept or slope)
   - Identify how time is measured and indexed in your analysis
3. Draw a figure representing they hypothesized relationship between the outcome and primary independent variable

Review

- Models distill important information about the world and have
  - An outcome
  - Deterministic component (data-level)
  - Stochastic component (datum-level)
- All models attempt to minimize error (everything we do creates fancy means)
- Use theory to determine which aspects in the model vary across groups, examine that variation empirically, and then analyze the data

QUESTIONS?