Fixed effect models and repeat sales method

DENG KUANGKUANG

Panel Data

- Panel data, also known as cross-sectional time-series data, is a dataset in which the behavior of entities are repeated observed across time.

- Theses entities could be states, companies, individuals, countries, etc.

- Panel data looks like this =>

country	year	y	x1	x2	x3
A	1990	1.34	0.28	-1.11	0.28
A	1991	-1.90	0.32	-0.95	0.49
A	1992	-0.01	0.36	-0.79	0.70
A	1993	2.65	0.25	-0.89	-0.09
A	1994	3.01	0.42	-0.73	0.95
A	1995	3.23	0.48	-0.72	1.03
B	1990	-5.94	-0.08	1.43	0.02
B	1991	-0.71	0.11	1.65	0.26
B	1992	-1.93	0.35	1.59	-0.23
B	1993	3.07	0.73	1.69	0.26
B	1994	3.77	0.72	1.74	0.41
B	1995	2.84	0.67	1.71	0.54
C	1990	-1.29	1.31	-1.29	0.20
C	1991	-3.42	1.18	-1.34	0.28
C	1992	-0.36	1.26	-1.26	0.37
C	1993	1.23	1.42	-1.31	-0.38
C	1994	3.80	1.11	-1.28	0.56

Panel data

- Panel data allows you to control for variables you cannot observe or measure, like cultural factors or differences in business practices across companies;
- Panel data allows you to control variables that change over time but not across entities, e.g. national policies, federal regulations, international agreements, etc.
- => It accounts for individual heterogeneity.

- With panel data, you can include levels of analysis (i.e. students, schools; districts, states) suitable for multilevel or hierarchical modeling.

- Drawbacks: data collection, non-response in the case of micro panels or cross-country dependency in the case of macro panels.

Panel Data

In this section, we focus on one technique used to analyze panel data:

- - Fixed effects

The data used for demonstration can be downloaded here:

- http://dss.princeton.edu/training/Panel101.dta

Setting panel data

- The Stata command to run fixed/random effects is *xtreg*
- Before using *xtreg*, you need to set Stata to handle panel data by using the command *xtset*. Type:
  - *xtset country year*

```
. xtset country year
       panel variable:   country (strongly balanced)
        time variable:   year, 1990 to 1999
                delta:   1 unit
```

- => “country” represents the entities/panels (i); “year” represents the time variable (t)
- Strongly balanced: all countries have data for all years. If, for example, one country does not have data for one year, then the data is unbalanced. Ideally you would wan to have a balanced dataset but this is not always the case. Even if so, you can still run the model

NOTE: if you get the following error after using xtset:

varlist: country: string variable not allowed

Convert ‘country’ to numeric, type:

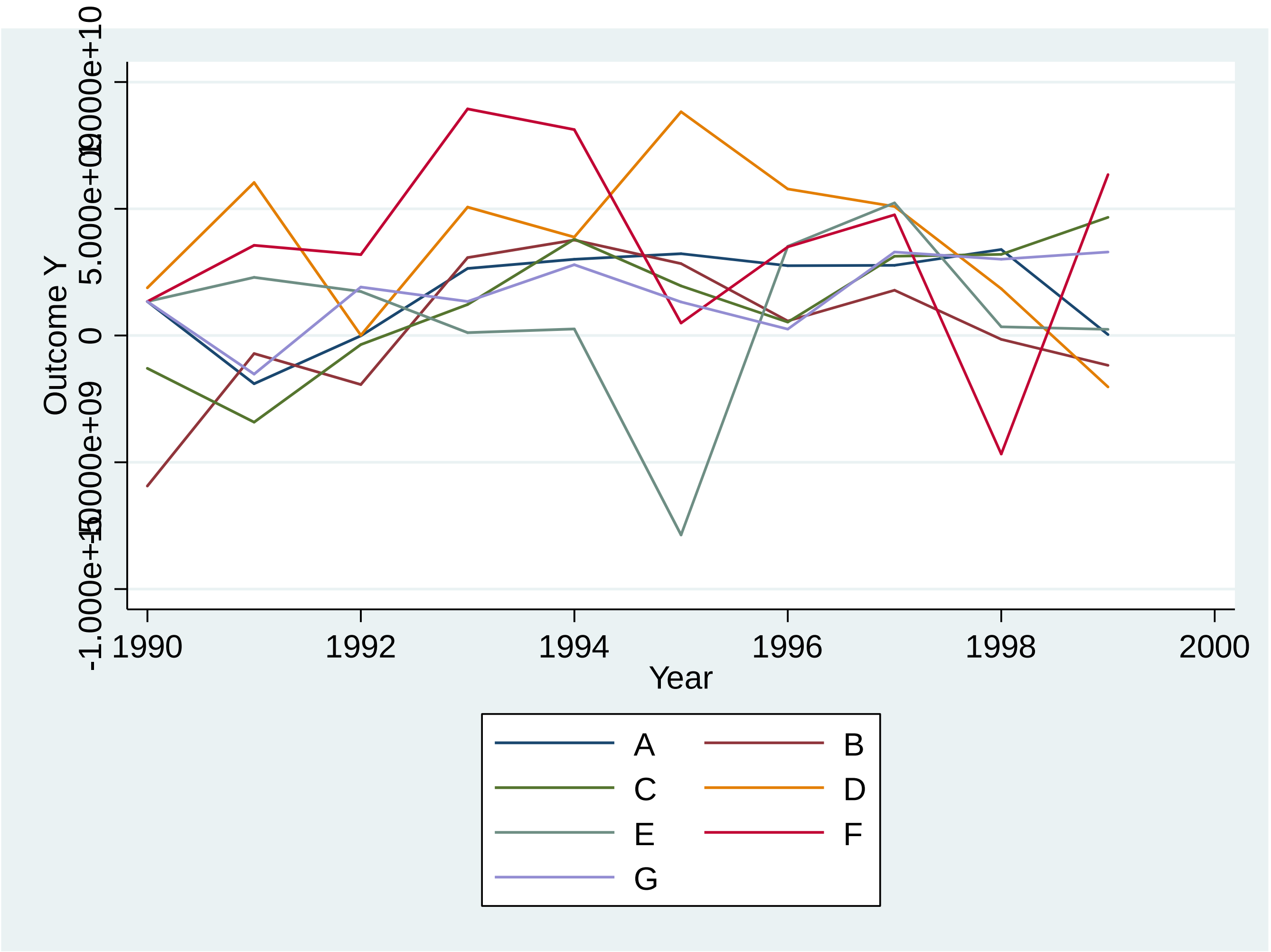
encode country, gen(country1)

Exploring panel data

```
use http://dss.princeton.edu/training/Panel101.dta
xtset country year
xtline y
```

Exploring panel data

```
xtline y, overlay
```



Fixed effects model

- Various types of fixed effects model:
  - Covariance model
  - Within estimator
  - Individual dummy variable model
  - Least squares dummy variable model

Fixed effects

- Use fixed-effects (FE) whenever you are only interested in analyzing the impact of variables that vary over time.

- FE explore the relationship between predictor and outcome variables within an entity (country, person, company, etc.). Each entity has its own individual characteristics that may or may not influence the predictor variables. For example:
  - being a male or female could influence the opinion toward certain issue
  - the political system of a particular country could have some effect on trade or GDP
  - the business practices of a company may influence its stock price

Fixed effects

- Assumption 1:
  - Something within the individual may impact or bias the predictor or the outcome variables and we need to control for this. This is the rationale behind the assumption of the correlation between entity's error term and predictor variables. FE remove the effect of those time-invariant characteristics so we can assess the net effect of the predictors on the outcome variable.

- Assumption 2:
  - Time-invariant characteristics are unique to the individual and should not be correlated with other individuals' characteristics. Each entity is different therefore the entity's error term and the constant (which captures individual characteristics) should not be correlated with the others'.

What if there is group-wise correlation?

Fixed effects

- The equation for the fixed effect model is:

$$Y_{it} = \beta_1 X_{it} + \alpha_i + u_{it}$$

There is no constant term.

Where

- α_i (i=1….n) is the unknown intercept for each entity (n entity-specific intercepts).
- Y_{it} is the dependent variable (DV) where i = entity and t = time.
- X_{it} represents one independent variable
- β_1 is the coefficient for that independent variable
- u_{it} is the error term

- The key insight is that if the unobserved variable does not change over time, then any changes in the dependent variable must be due to influences other than these fixed characteristics.
- In the case of time-series cross-sectional data the interpretation of the beta coefficients would be "…*for a given country*, as X varies *across time* by one unit, Y increases or decreases by β units"
- Fixed-effects will not work well with data for which within-cluster variation is minimal or for slow changing variables over time.

Fixed effects

- Another way to see the fixed effects model is by using binary variables. So the equation for the fixed effects model is:

$$Y_{it} = \beta_0 + \beta_1 X_{1,it} + \ldots + \beta_k X_{k,it} + \gamma_2 E_2 + \ldots + \gamma_n E_n + u_{it}$$

Where

–Y_{it} is the dependent variable (DV) where i = entity and t = time.

–$X_{k,it}$ represents independent variables

–β_k is the coefficient for the independent variable

–u_{it} is the error term

–E_n is the entity n. Since they are binary (dummies) you have n-1 entities included in the model.

–γ_2 Is the coefficient for the binary repressors (entities)

- The two equations are equivalent to each other
  - The slope coefficient on X is the same from one entity to the next. The <u>entity-specific intercepts</u> in [eq.1] and the <u>binary regressors</u> in [eq.2] have the same source: the unobserved variable Zi that varies across states but not over time."

Fixed effects

- You could add time effects to the entity effects model to have a time fixed effects regression model:

$$Y_{it} = \beta_0 + \beta_1 X_{1,it} + \ldots + \beta_k X_{k,it} + \gamma_2 E_2 + \ldots + \gamma_n E_n + \delta_2 T_2 + \ldots + \delta_t T_t + u_{it} \quad \text{[eq.3]}$$

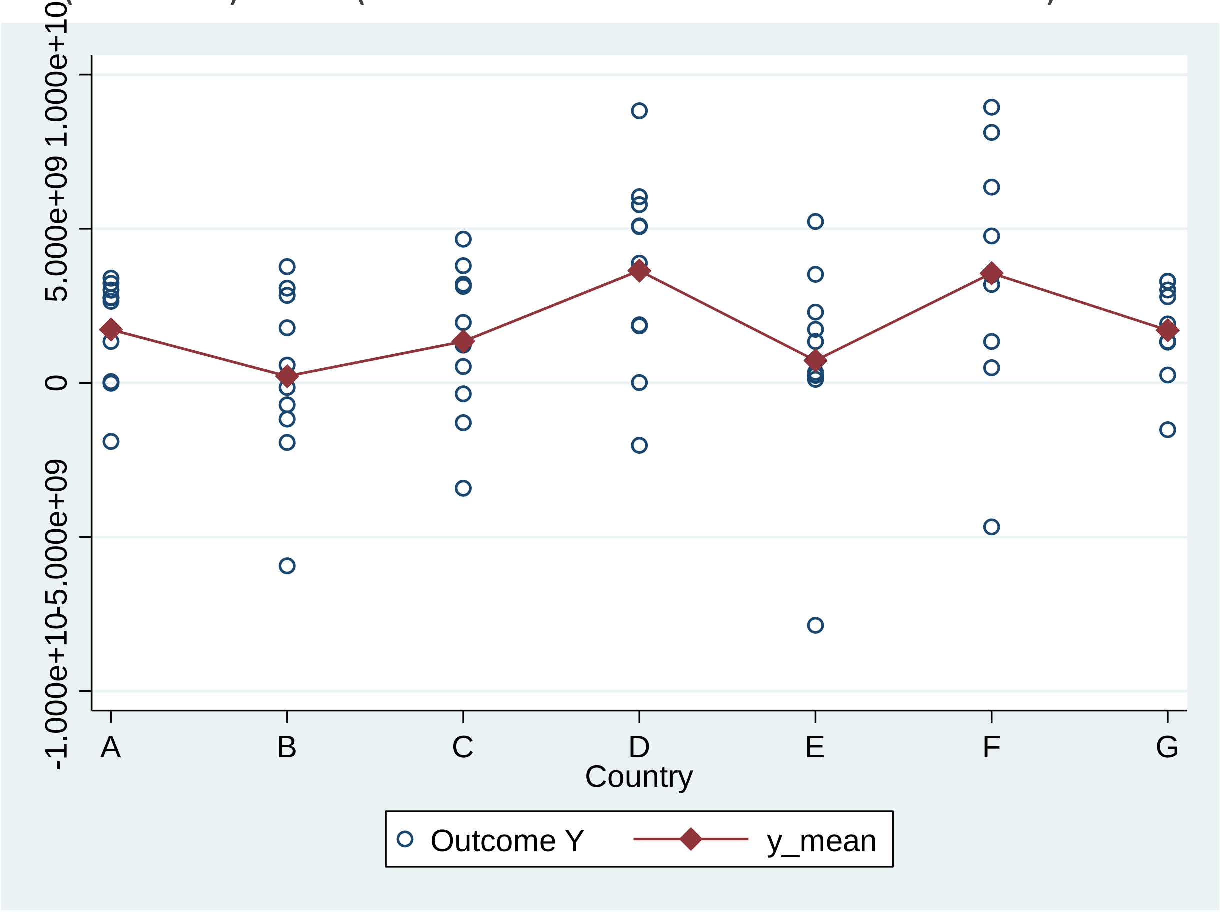
Where

- Y_{it} is the dependent variable (DV) where i = entity and t = time.
- $X_{k,it}$ represents independent variables (IV),
- β_k is the coefficient for the IVs,
- u_{it} is the error term
- E_n is the entity n. Since they are binary (dummies) you have n-1 entities included in the model.
- γ_2 is the coefficient for the binary regressors (entities).
- T_t is time as binary variable (dummy), so we have t-1 time periods.
- δ_t is the coefficient for the binary time regressors .

- Control for time effects whenever unexpected variation or special events may affect the outcome variable.

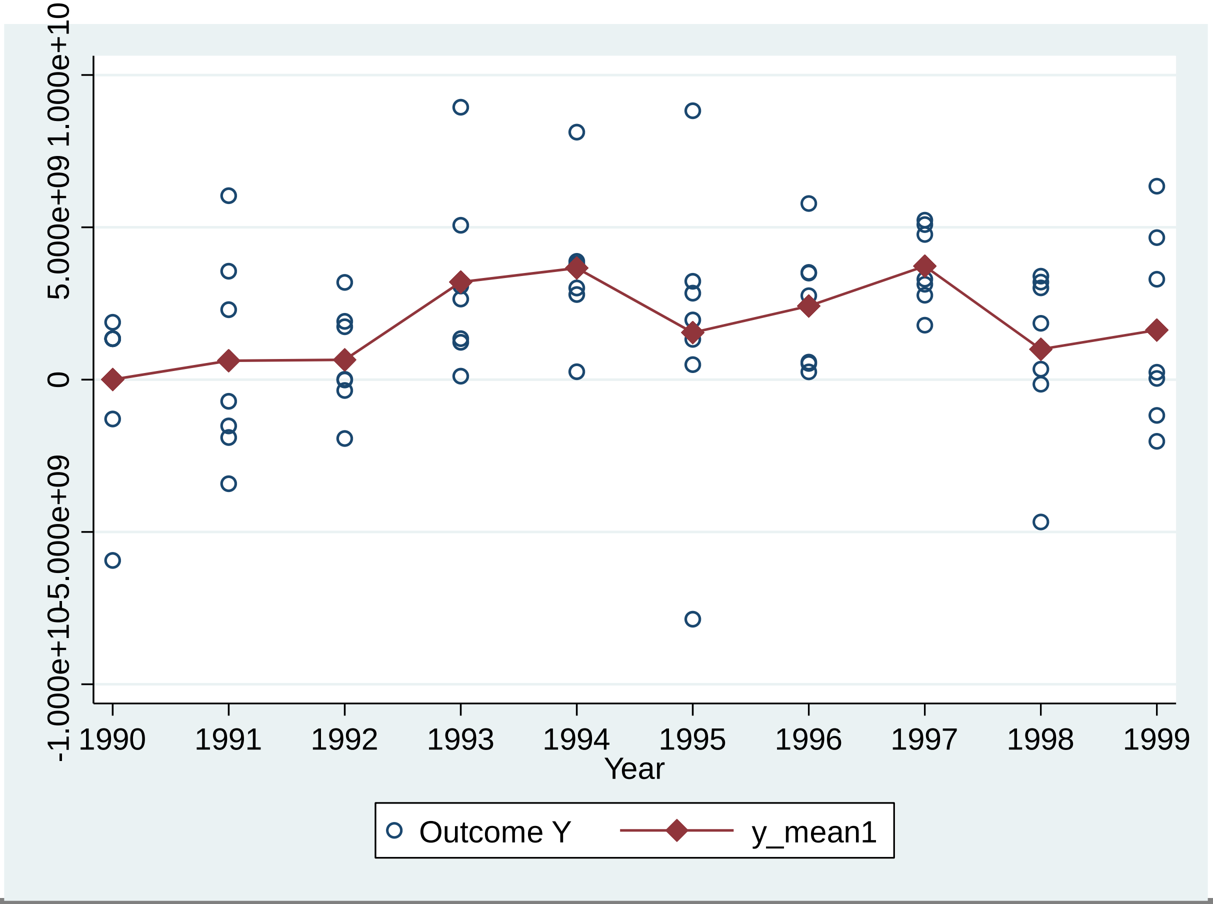
Fixed effects: heterogeneity across countries

- bysort country: egen y_mean=mean(y)
- twoway scatter y country, msymbol(circle_hollow) || connected y_mean country, msymbol(diamond) xlabel(1 "A" 2 "B" 3 "C" 4 "D" 5 "E" 6 "F" 7 "G")



Fixed effects: heterogeneity over time

- bysort year: egen y_mean1=mean(y)
- twoway scatter y year, msymbol(circle_hollow) || connected y_mean1 year, msymbol(diamond) xlabel(1990(1)1999)

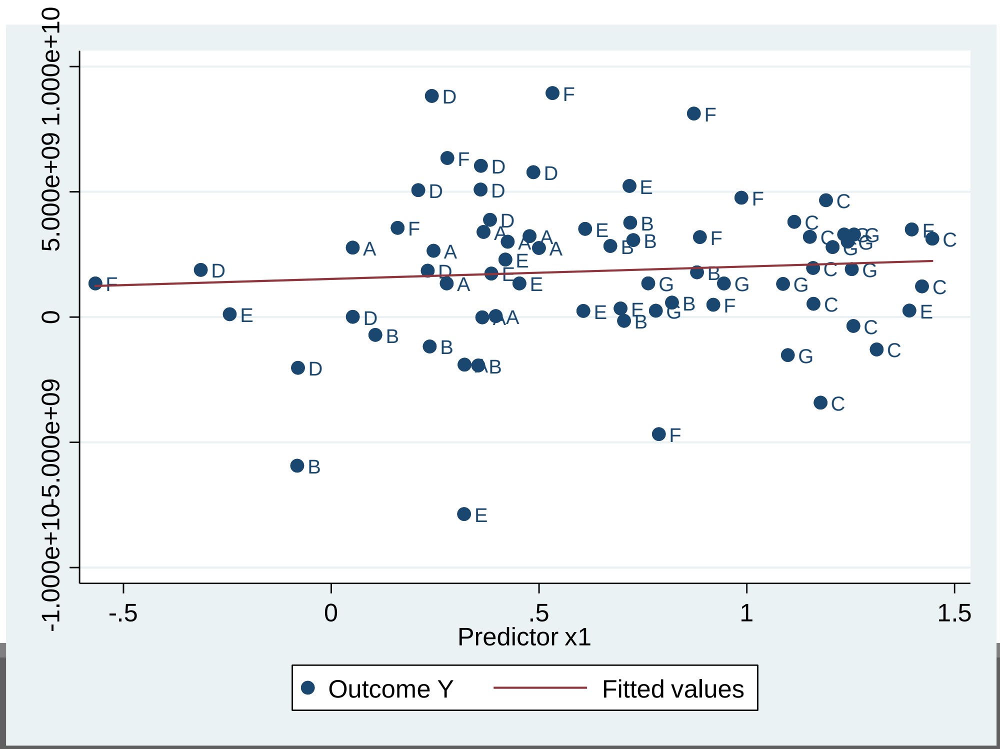


OLS regression

```
. regress y x1

      Source |       SS           df       MS      Number of obs   =        70
-------------+----------------------------------   F(1, 68)        =      0.40
       Model |  3.7039e+18         1  3.7039e+18   Prob > F        =    0.5272
    Residual |  6.2359e+20        68  9.1705e+18   R-squared       =    0.0059
-------------+----------------------------------   Adj R-squared   =   -0.0087
       Total |  6.2729e+20        69  9.0912e+18   Root MSE        =   3.0e+09

------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          x1 |   4.95e+08   7.79e+08     0.64   0.527    -1.06e+09    2.05e+09
       _cons |   1.52e+09   6.21e+08     2.45   0.017     2.85e+08    2.76e+09
------------------------------------------------------------------------------
```



FE using least square dummy variable model

```
. regress y x1 i.country
```

Source	SS	df	MS
Model	1.4276e+20	7	2.0394e+19
Residual	4.8454e+20	62	7.8151e+18
Total	6.2729e+20	69	9.0912e+18

Number of obs	=	70
F(7, 62)	=	2.61
Prob > F	=	0.0199
R-squared	=	0.2276
Adj R-squared	=	0.1404
Root MSE	=	2.8e+09

y	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
x1	2.48e+09	1.11e+09	2.24	0.029	2.63e+08	4.69e+09
country						
B	-1.94e+09	1.26e+09	-1.53	0.130	-4.47e+09	5.89e+08
C	-2.60e+09	1.60e+09	-1.63	0.108	-5.79e+09	5.87e+08
D	2.28e+09	1.26e+09	1.81	0.075	-2.39e+08	4.80e+09
E	-1.48e+09	1.27e+09	-1.17	0.247	-4.02e+09	1.05e+09
F	1.13e+09	1.29e+09	0.88	0.384	-1.45e+09	3.71e+09
G	-1.87e+09	1.50e+09	-1.25	0.218	-4.86e+09	1.13e+09
_cons	8.81e+08	9.62e+08	0.92	0.363	-1.04e+09	2.80e+09

```
regress y x1
estimates store ols
regress y x1 i.country
estimates store ols_dum
estimates table ols ols_dum,
star stats(N)
```

Variable	ols	ols_dum
x1	4.950e+08	2.476e+09*
country		
B		-1.938e+09
C		-2.603e+09
D		2.282e+09
E		-1.483e+09
F		1.130e+09
G		-1.865e+09
_cons	1.524e+09*	8.805e+08
N	70	70

legend: * p<0.05; ** p<0.01; *** p<0.001

Fixed effects using xtreg

Comparing the fixed effects using dummies with xtreg we get the same results.

Problem?

```
. xtreg y x1, fe

Fixed-effects (within) regression               Number of obs      =        70
Group variable: country                         Number of groups   =         7

R-sq:                                           Obs per group:
     within  = 0.0747                                         min =        10
     between = 0.0763                                         avg =      10.0
     overall = 0.0059                                         max =        10

                                                F(1,62)            =      5.00
corr(u_i, Xb)  = -0.5468                        Prob > F           =    0.0289

------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          x1 |   2.48e+09   1.11e+09     2.24   0.029     2.63e+08    4.69e+09
       _cons |   2.41e+08   7.91e+08     0.30   0.762    -1.34e+09    1.82e+09
-------------+----------------------------------------------------------------
     sigma_u |  1.818e+09
     sigma_e |  2.796e+09
         rho |  .29726926   (fraction of variance due to u_i)
------------------------------------------------------------------------------
F test that all u_i=0:     F(6, 62) =     2.97               Prob > F = 0.0131
```

Fixed effects using areg

- If you want to hide the binary variables for each entity (country), use *areg* with *absorb*
- This is particularly useful when you have many levels of fixed effects, and you want to generate nice and succinct output tables

```
. areg y x1, absorb(country)

Linear regression, absorbing indicators         Number of obs     =         70
                                                F(   1,      62)  =       5.00
                                                Prob > F          =     0.0289
                                                R-squared         =     0.2276
                                                Adj R-squared     =     0.1404
                                                Root MSE          =  2.796e+09
```

y	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
x1	2.48e+09	1.11e+09	2.24	0.029	2.63e+08	4.69e+09
_cons	2.41e+08	7.91e+08	0.30	0.762	-1.34e+09	1.82e+09
country	F(6, 62) =		2.965	0.013	(7 categories)	

Fixed effects: compare the three approaches

- comparing xtreg (with fe), regress (OLS with dummies) and areg

- `xtreg y x1 x2 x3, fe`
- `estimates store fixed`
- `regress y x1 x2 x3 i.country`
- `estimates store ols`
- `areg y x1 x2 x3, absorb(country)`
- `estimates store areg`
- `estimates table fixed ols areg, star stats(N r2 r2_a)`

Variable	fixed	ols	areg
x1	2.425e+09*	2.425e+09*	2.425e+09*
x2	1.823e+09	1.823e+09	1.823e+09
x3	3.097e+08	3.097e+08	3.097e+08
country			
2		-5.961e+09	
3		-1.598e+09	
4		-2.091e+09	
5		-5.732e+09	
6		8.026e+08	
7		-1.375e+09	
_cons	-2.060e+08	2.073e+09	-2.060e+08
N	70	70	70
r2	.10092442	.24948198	.24948198
r2_a	-.03393692	.13690428	.13690428

legend: * p<0.05; ** p<0.01; *** p<0.001

- The alternative approach: demean each variable from its group average across time

A note on fixed effects

- "…The fixed-effects model controls for all time-invariant differences between the individuals, so the estimated coefficients of the fixed-effects models cannot be biased because of omitted time-invariant characteristics…[like culture, religion, gender, race, etc]

- One side effect of the features of fixed-effects models is that they cannot be used to investigate time-invariant causes of the dependent variables. Technically, time-invariant characteristics of the individuals are perfectly collinear with the person [or entity] dummies. Substantively, fixed-effects models are designed to study the causes of changes within a person [or entity]. A time-invariant characteristic cannot cause such a change, because it is constant for each person.

Testing for time-fixed effects

i.month versus i.yrmon

```
. xtreg y x1 i.year, fe

Fixed-effects (within) regression               Number of obs      =        70
Group variable: country                         Number of groups   =         7

R-sq:                                           Obs per group:
     within  = 0.2323                                         min =        10
     between = 0.0763                                         avg =      10.0
     overall = 0.1395                                         max =        10

                                                F(10,53)           =      1.60
corr(u_i, Xb)  = -0.2014                        Prob > F           =    0.1311

------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          x1 |   1.39e+09   1.32e+09     1.05   0.297    -1.26e+09    4.04e+09
             |
        year |
       1991  |   2.96e+08   1.50e+09     0.20   0.844    -2.72e+09    3.31e+09
       1992  |   1.45e+08   1.55e+09     0.09   0.925    -2.96e+09    3.25e+09
       1993  |   2.87e+09   1.50e+09     1.91   0.061    -1.42e+08    5.89e+09
       1994  |   2.85e+09   1.66e+09     1.71   0.092    -4.84e+08    6.18e+09
       1995  |   9.74e+08   1.57e+09     0.62   0.537    -2.17e+09    4.12e+09
       1996  |   1.67e+09   1.63e+09     1.03   0.310    -1.60e+09    4.95e+09
       1997  |   2.99e+09   1.63e+09     1.84   0.072    -2.72e+08    6.26e+09
       1998  |   3.67e+08   1.59e+09     0.23   0.818    -2.82e+09    3.55e+09
       1999  |   1.26e+09   1.51e+09     0.83   0.409    -1.77e+09    4.29e+09
             |
       _cons |  -3.98e+08   1.11e+09    -0.36   0.721    -2.62e+09    1.83e+09
-------------+----------------------------------------------------------------
     sigma_u |  1.547e+09
     sigma_e |  2.754e+09
         rho |  .23985725   (fraction of variance due to u_i)
------------------------------------------------------------------------------
F test that all u_i=0:     F(6, 53) =     2.45               Prob > F = 0.0362
```

To see if time fixed effects are needed when running a FE model use the command **`testparm`**. It is a joint test to see if the dummies for all years are equal to 0, if they are then no time fixed effects are needed

```
. testparm i.year

 ( 1)  1991.year = 0
 ( 2)  1992.year = 0
 ( 3)  1993.year = 0
 ( 4)  1994.year = 0
 ( 5)  1995.year = 0
 ( 6)  1996.year = 0
 ( 7)  1997.year = 0
 ( 8)  1998.year = 0
 ( 9)  1999.year = 0

       F(  9,    53) =    1.21
            Prob > F =    0.3094
```

Time FE unnecessary

Other types of dataset

- Fixed effect models can be used even if you are not dealing with standard panel data
- For example:
  - Real estate resale data
  - Cross-sectional feature: geographical location
  - Time-series feature: each transaction occurs at a different time, although not at a fixed time interval

  - Running models on this dataset: e.g. how household income affects housing prices:
    - District/block/estate-level fixed effects are often used
    - When household income can be measured at a level that is more specific than just time-varying, time fixed effects are also often included

Repeat sales method for housing resales data

Suppose we want to construct a housing price index:

How would you do it?

Compare with stock price index

- Simplest method – average price
  - Collect all property transactions every month (quarter / year)
  - Average the price in each month
  - Last month: $100,000/sq.m (say)
  - This month: $101,000/sq.m (say)
  - Home price increases by $1,000/sq.m (+1%)
  - Any problem?

Repeat sales method


Same for:
Stock price index,
Consumer price index,
…

Period	A	B	Average
t	$8	$12	$10
t+1	$10	$16	$13
Change			$3 (+30%)

Properties are not frequently transacted…

Period	A	B	Average
t	$8	Not sold	$8
t+1	Not sold	$16	$16
Change			$8 (+100%) !!!

Not *pure* price change because of quality differences

Repeat sales method

Period	A	B	C
t	$8	-	$10
t+1	-	$16	$12
t+2	$6	$11	-

Average only when there are repeat sales

- Repeat-sales method
- Case & Shiller

Transaction-based methods

Adjust A, B, C (if sold) to same quality before taking average

- Hedonic pricing method

Fill in the blanks (unsold property) before taking average

- Valuation-based method
- Commonly used by consultants

Repeat sales method

%change from t to t+1:
(12-10)/10 = +20%

Not used

Period	A	B	C	Average
t	$8	-	$10	$9
t+1	-	$16	$12	$14 (+56%)
t+2	$6		-	$6 (-57%)

%change from t to t+2:
(6-8)/8 = -25%

No quality differences for temporal comparison of each property

	Index
t	100 (arbitrary)
t+1	120 = 100*(1+20%)
t+2	75 = 100*(1-25%)

%change from t+1 to t+2:
(75-120)/120 = -38%

The ‘averaging’ process

t to t+2: -25%

t to t+1: +20%

Period	A	B	C
t	$8	-	$10
t+1	-	$16	$12
t+2	$6	$11	-

t+1 to t+2: -32%

	Index
t	100 (arbitrary)
t+1	$115 = 100 * \left[1 + \frac{2 * 20\% + 1 * (-25\% + 32\%)}{3}\right]$
t+2	$80 = 100 * \left[1 + \frac{2 * (-25\%) + 1 * (20\% - 32\%)}{3}\right]$

t to t+1: 15%

t+1 to t+2: -31%

Repeat sales method

Suppose you collect a lot of repeat sales in 3 periods: t, t+1, t+2

A property was transacted twice, first at price $P(t_1)$ and then at price $P(t_2)$

Continuously compounded return

D_1: =1 if t_2=t+1; =-1 if t_1=t+1; 0= otherwise

D_2: =1 if t_2=t+2; =-1 if t_1=t+2; 0= otherwise

$$ln\frac{P(t_2)}{P(t_1)} = a_1 \times D_1 + a_2 \times D_2 + \varepsilon$$

	$P(t_1)$	$P(t_2)$	D_1	D_2
A	8	6	0	1
B	16	11	-1	1
C	10	12	1	0

What is the 'average' return from t to t+1?

Index at t=100 (arbitrary)

Index in t+1 = 100*exp(a_1)

% change from t to t+1: exp(a_1)-1

from t to t+2?

Index in t+2 = 100*exp(a_2)

% change from t to t+1: exp(a_2)-1

Repeat sales method

- An alternative way is to control for property attributes through quantified variables
- Pros of RS
  - No need to measure and control housing attributes, some of which are difficult to observe or hard to measure
- Cons of RS
  - Repeat-sales index uses only a sub-sample of sales but does not required knowledge about X as long as quality remains constant between the two sales
  - => limited sample size induces relatively big estimation errors => suitable for highly liquid market
  - => building age matters
  - => fix-ups

Repeat sales method – fix 'age'

- Depreciation of properties between sales is ignored in repeat sales method
  - Returns underestimated from the traditional RS method
  - Depreciation increases with the time interval between sales
- Fix 1: select repeat sales with limited time interval between sales, e.g. 10 years
- Fix 2: Control depreciation

Continuously compounded return

$=1$ if $t_2=t+1$
$=-1$ if $t_1=t+1$
0= otherwise

$=1$ if $t_2=t+2$
$=-1$ if $t_1=t+2$
0= otherwise

The age difference between t_1 and t_2 used to control depreciation

$$ln\frac{P(t_2)}{P(t_1)} = a_1 \times D_1 + a_2 \times D_2 + a_3(Age_{t2}^{\lambda_2} - Age_{t1}^{\lambda_1}) + \varepsilon$$

Chau KW, Wong SK & Yiu CY (2005) Adjusting for non-linear age effects in the repeat sales index, *Journal of Real Estate Finance & Economics*, 31(2), 137-153.

Repeat sales method – fix ‘fix-ups’

- when a home is transacted, the new owner-occupier tend to fix the property before moving in
- Flippers who aims to buy and quickly sell the property may also make cosmetic renovation in order to ask for a higher price
- Control the fix-up effect
  - Fix-ups happen to ‘all’ transactions, likely with equal chance

Continuously compounded return

Fix-ups controlled by a constant

D_1: =1 if t_2=t+1; =-1 if t_1=t+1; 0= otherwise

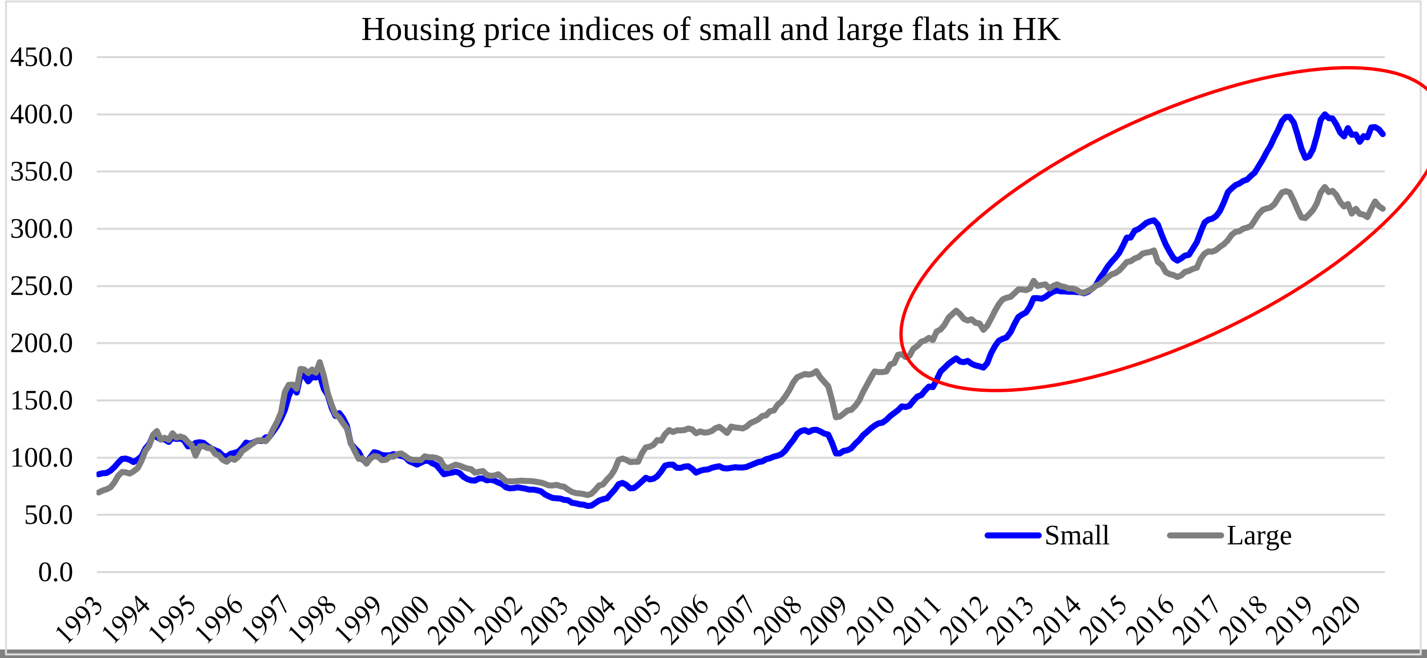
D_2: =1 if t_2=t+2; =-1 if t_1=t+2; 0= otherwise

$$ln\frac{P(t_2)}{P(t_1)} = a_0 + a_1 \times D_1 + a_2 \times D_2 + \varepsilon$$

But in practice, a_0 overestimated the fix-up effects, leading to underestimation of returns.

An example

- Initial observation:
  - In the housing market, the price per square meter of smaller/less expensive housing units (starter homes) is usually higher than larger/more expensive housing units.
  - When the market price grows, the prices of smaller/less expensive properties grows faster.



Wong SK, Deng KK & Cheng KS (forthcoming) Starter home premium and housing affordability. Journal of Real Estate Finance and Economics

Research question: why

- Our answer: capital constraint (资本约束)
  - Buyers pay less for smaller homes, which relaxes the capital constraint. Being less expensive attracts buyers to pay a 'premium'!
  - The effect of capital constraint should be stronger when
    - Some tools help fruther relax the constraint for less expensive homes, i.e. eligibility to initiate a mortgage loan at lower rate (处女贷)
    - Housing is particularly unaffordable

- Research design:
  - Difficulty:
    - Housing heterogeneity: any difference in price/price growth could be simply due to quality difference
    - The higher unit price of smaller flats could be due to diminishing marginal utility
    - Can we single out the effect of capital constraint
  - Our unique case
    - The HOS (共有产权房) secondary market in HK

Institutional background

- The shared-equity homes in Shanghai follows the institutional design of HOS in Hong Kong
- The homeownership scheme is subsidized housing plan in Hong Kong. It is essentially a shared-equity scheme.
- The gov't constructs HOS housing units, and sells them at a discount to eligible low-to-middle income households. The discount (d) varies from 6% to 63%, averaged at 40%.
- Shared-equity arrangement: If a buyer buys at a 40% discount to the market price, the owner owns 60% of the flat, the gov't owns the other 40%.
- The market of new HOS units is highly competitive due to oversubscription. A lottery is used to determine which applicants could purchase HOS directly from the gov't.

Institutional background

- After a minimum holding period, the owners of HOS could sell the units in the secondary market through either of two approaches:
- S2P sales: purchase the 40% ownership from the government by paying the land premium (privatize the unit). Then the owner could sell the flat to anyone.

$$P_{i,t}^{S2P} = V_{i,t}$$

- S2S sales: sell to other eligible HOS buyers who are not lucky enough to purchase new HOS from gov't. The transaction price is negotiated between the seller and buyer without any restriction.
  - In this case: the buyer purchases only 60% of ownership. Gov't keeps the 40% of ownership.

$$P_{i,t}^{S2S} = V_{i,t}(1 - d_i) + V_{i,t} d_i \left(\sum_{t=1}^{H_i} y_t + f_{i,t} \right)$$

Research design

- The properties of S2S and S2P sales are drawn from the same pool. The quality of the two groups should be homogenous.
- The embedded gov't share of ownership relaxes the capital constraint for buyers of S2S but not for buyers of S2P.
- If capital constraint really matters, there should be a price premium in S2S sales:

$$prem_{i,t} = \frac{P_{i,t}^{S2S}}{1 - d_i} - P_{i,t}^{S2P}$$

=>

$$prem_{i,t} = V_{i,t} \frac{d_i}{(1 - d_i)} \left(\sum_{t=1}^{H_i} y_t + f_{i,t} \right)$$

Data

Statistic	N	Mean	St. Dev.	Min	Median	Max
Panel A All S2S transactions						
Price (million HK$)	27,508	1.425	0.761	0.220	1.220	6.000
Age (months)	27,508	121.719	54.252	33	115	383
Size (sq.ft.)	27,508	693.535	110.189	69	707	997
Floor	27,508	18.226	10.466	0	18	45
Discount	27,508	0.406	0.075	0.060	0.430	0.630
Rel.P	27,508	0.179	0.267	-2.255	0.202	1.101
yield	27,508	0.908	0.174	0.600	0.923	1.218
PI 1	27,508	0.516	0.194	0.283	0.449	0.997
PI 2	27,508	1.087	0.495	0.590	0.857	2.232
Panel B All S2P transactions						
Price (million HK$)	48,697	1.619	0.807	0.102	1.420	9.500
Age (months)	48,697	213.691	67.177	38	213	386
Size (sq.ft.)	48,697	581.102	103.919	69	564	997
Floor	48,697	16.412	10.099	0	16	46

Table 2 Empirical results based on full sample

Dept. var.	log(discount-adjusted S2S price) or log(S2P Price)			
S2S	0.152***	0.188***	0.187***	0.184***
	(0.002)	(0.002)	(0.002)	(0.002)
S2S × Discount		1.225***	1.226***	1.316***
		(0.016)	(0.016)	(0.016)
S2S × rel.P			-0.012**	-0.031***
			(0.005)	(0.005)
S2S × PI				0.438***
				(0.010)
S2S × yield		0.396***	0.399***	0.810***
		(0.007)	(0.007)	(0.012)
log(district index)	1.048***	1.095***	1.095***	1.064***
	(0.002)	(0.002)	(0.002)	(0.002)
log(SIZE)	1.051***	1.044***	1.048***	1.060***
	(0.004)	(0.004)	(0.004)	(0.004)
log(AGE)	-0.018***	0.009***	0.009***	0.021***
	(0.002)	(0.002)	(0.002)	(0.002)
log(FLOOR)	0.060***	0.060***	0.060***	0.060***
	(0.001)	(0.001)	(0.001)	(0.001)
Constant	-11.349***	-11.730***	-11.758***	-11.716***
	(0.031)	(0.031)	(0.033)	(0.032)
Estate fixed effects	Yes	Yes	Yes	Yes
Observations	76,205	76,205	76,205	76,205
Adjusted R^2	0.907	0.916	0.916	0.918

PSM – synchronize S2S and S2P sales

Statistic	N	Mean	St. Dev.	Min	Median	Max
Panel A All S2S transactions						
Price (million HK$)	11,325	1.308	0.663	0.220	1.150	5.230
Age (months)	11,325	144.117	54.060	36	136	383
Size (sq.ft.)	11,325	670.841	108.843	282	692	997
Floor	11,325	17.845	10.245	0	17	40
Discount	11,325	0.395	0.077	0.060	0.400	0.630
Rel.P	11,325	0.104	0.268	-0.862	0.116	0.965
yield	11,325	0.881	0.161	0.600	0.899	1.218
PI 1	11,325	0.524	0.182	0.283	0.469	0.997
PI 2	11,325	1.132	0.477	0.590	0.890	2.232
Panel B All S2P transactions						
Price (million HK$)	11,325	1.707	0.892	0.108	1.460	6.800
Age (months)	11,325	159.587	53.700	38	153	382
Size (sq.ft.)	11,325	634.782	107.710	282	601	988
Floor	11,325	16.437	10.532	0	16	42

Table 4 Empirical results based on matched sample

Dept. var.	log(discount-adjusted S2S price) or log(S2P Price)			
S2S	0.132***	0.140***	0.139***	0.138***
	(0.002)	(0.002)	(0.002)	(0.002)
S2S × Discount		1.016***	1.013***	1.135***
		(0.024)	(0.024)	(0.024)
S2S × rel.P			-0.035***	-0.041***
			(0.007)	(0.007)
S2S × PI				0.551***
				(0.017)
S2S × yield		0.200***	0.209***	0.705***
		(0.012)	(0.012)	(0.019)
log(district index)	1.057***	1.090***	1.094***	1.030***
	(0.004)	(0.004)	(0.004)	(0.005)
log(SIZE)	1.089***	1.086***	1.103***	1.120***
	(0.007)	(0.007)	(0.008)	(0.008)
log(AGE)	-0.016***	-0.010**	-0.012***	0.008*
	(0.004)	(0.004)	(0.004)	(0.004)
log(FLOOR)	0.062***	0.062***	0.063***	0.063***
	(0.001)	(0.001)	(0.001)	(0.001)
Constant	-11.705***	-11.906***	-12.018***	-11.892***
	(0.060)	(0.059)	(0.063)	(0.062)
Estate fixed effects	Yes	Yes	Yes	Yes
Observations	22,650	22,650	22,650	22,650
Adjusted R^2	0.909	0.916	0.916	0.920

Repeat sales method – quality control

Table 5 The repeat sales sample

Statistic	N	Mean	St. Dev.	Min	Median	Max
S2S Price (million HK$)	1,719	1.084	0.434	0.250	1.010	3.580
S2P Price (million HK$)	1,719	2.038	1.055	0.272	1.755	6.500
Index return between S2S and S2P sales	1,719	-0.308	0.475	-1.697	-0.302	1.102
Discount	1,719	0.362	0.083	0.130	0.360	0.530
Rel.P	1,719	-0.007	0.319	-1.680	0.000	0.873
yield	1,719	0.992	0.134	0.607	0.952	1.218
PI 1	1,719	0.449	0.119	0.283	0.424	0.997
PI 2	1,719	0.846	0.243	0.590	0.801	2.232

Repeat sales method – quality control

Table 6 Empirical results based on repeat sales sample

Dept. var.	**log(discount-adjusted S2S price / S2P Price)**			
Index return	0.875***	0.966***	0.969***	0.971***
	(0.014)	(0.013)	(0.013)	(0.013)
rel.P			-0.137***	-0.135***
			(0.018)	(0.018)
PI				0.140***
				(0.041)
Discount		1.864***	1.826***	1.822***
		(0.072)	(0.071)	(0.071)
yield		0.245***	0.258***	0.468***
		(0.044)	(0.044)	(0.075)
Constant	0.155***	0.183***	0.184***	0.184***
	(0.008)	(0.007)	(0.007)	(0.007)
Observations	1,719			1,719
Adjusted R^2	0.689	0.776	0.783	0.784
Residual Std. Error	0.279	0.237	0.233	0.233