Lecture 2: Panel Data

Fabian Waldinger

Topics Covered in Lecture

1. Omitted variables and panel data models.
2. Random effects (RE) methods.
3. Fixed effects (FE) methods.
4. Measurement error in FE methods.
5. Application 1: Bertrand and Schoar (2003) - Manager FE.
6. Application 2: Card, Heining, Kline (2013) - Estimating firm and worker FE to understand changes in wage inequality.

Panel Data

- Panel data sets are very widely used in applied economics.
- Panal data include observations from:
  - multiple cross-sectional units (e.g. individuals, firms, countries,...)
  - that are observed for at least two time periods (e.g. years, months, days,...)
- Panel data methods (mostly fixed effects) are often used in combination with other applied micro techniques such as IV or Differences-in-Differences.

Panel Data - Some Definitions

- Panel data sets come in two forms:
  1. Balanced panel: each cross-sectional unit is observed for the same time periods.
  2. Unbalanced panel: cross-sectional units are observed for different amounts of time.

The Omitted Variables Problem

- Panel data is useful to solve common omitted variables problems.
- Suppose you are interested in understanding the linear relationship between x and y using the following linear model:

$$E(y \mid \mathbf{x}, c) = \beta_0 + \mathbf{x}\boldsymbol{\beta} + c$$

- where interest lies in the $K \times 1$ vector $\boldsymbol{\beta}$.
- If $Cov(x_j, c) = 0$ for all j, there is no issue and we can estimate the model using OLS.
- However, if $Cov(x_j, c) \neq 0$ for some j, not considering c will lead to the standard endogeneity problem and to biased OLS estimates.

The Omitted Variables Problem - Cross-Sectional Data

- What could you do about the omitted variable problem in cross-sectional data?
  - find a proxy
  - find a valid IV that is correlated with the elements of $\mathbf{x}$ that are correlated with c.
- Panel data allows gives additional possibilities to deal with the omitted variables problem.
  - Under relatively strong assumptions we can use RE models
  - Eliminating c using fixed effects methods.

Strict Exogeneity Assumption

- To estimate the most basic panel data models (random effects estimator and fixed effects estimator) we assume *strict exogoneity*:

$$E(y_{it} \mid \mathbf{x}_{i1}, \mathbf{x}_{i2}, ..., \mathbf{x}_{iT}, c_i) = E(y_{it}|\mathbf{x}_{it}, c_i) = \mathbf{x}_{it}\beta + c_i$$

- In words: once $\mathbf{x}_{it}$ and c_i are controlled for, $\mathbf{x}_{is}$ has no partial effect on y_{it} for $s \neq t$.
- In the regression model $y_{it} = \mathbf{x}_{it}\boldsymbol{\beta} + c_i + u_{it}$ the strict exogeneity assumption can be stated in terms of idiosyncratic errors as:

$$E(u_{it} \mid \mathbf{x}_{i1}, \mathbf{x}_{i2}, ..., \mathbf{x}_{iT}, c_i) = 0, \qquad t = 1, 2, ..., T$$

- This assumption implies that explanatory variables in each time period are uncorrelated with the idiosycratic error u_{it} in each time period:

$$E(\mathbf{x}_{is}' u_{it}) = 0 \qquad s, t = 1, ..., T$$

- This assumption is much stronger than assuming no contemporaneous correlation $E(\mathbf{x}_{it}' u_{it}) = 0, t = 1, ..., T$.

Random Effects Methods

- Random effects models effectively put c_i in the error term under the assumption that c_i is orthogonal to $\mathbf{x}_{it}$ and then accounts for the serial correlation in the composite error.
- Random effects models therefore impose strict exogeneity *plus* orthogonality between c_i and $\mathbf{x}_{it}$:
  1. - $E(u_{it} \mid \mathbf{x}_i, c_i) = 0, \qquad t = 1, 2, ..., T$
     - $E(c_i \mid \mathbf{x}_i) = E(c_i) = 0$
- where $\mathbf{x}_i = (\mathbf{x}_{i1}, \mathbf{x}_{i2}, ..., \mathbf{x}_{iT})$
- The important part of assumption 2 is $E(c_i \mid \mathbf{x}_i) = E(c_i)$ the assumption $E(c_i) = 0$ is without loss of generality as long as an intercept is included in $\mathbf{x}_{it}$.
- With the second part of this assumption even OLS would be consistent but not efficient $\Rightarrow$ use GLS.

- The random effects approach accounts for the serial correlation in the composite error $\nu_{it} = c_i + u_{it}$.
- Rewriting our regression model including the composite error:

$$y_{it} = \mathbf{x}_{it}\boldsymbol{\beta} + \nu_{it}$$

- The random effects assumptions imply:

$$E(\nu_{it} \mid \mathbf{x}_i) = 0 \qquad t = 1, 2, ..., T$$

- We can therefore apply GLS methods that account for the particular error structure in $\nu_{it} = c_i + u_{it}$.

Random Effects = Feasible GLS

- The model for all time periods can be written as:

$$\mathbf{y}_i = \mathbf{X}_i \boldsymbol{\beta} + \mathbf{v}_i$$

- Define the (unconditional) variance matrix of $\mathbf{v}_i$ as:

$$\Omega \equiv E(\mathbf{v}_i \mathbf{v}_i')$$

- A $T \times T$ matrix. This matrix is the same for all i because of the random sampling assumption in the cross section.
- For consistency of GLS we need the usual rank condition for GLS:
  2. *rank* $E(\mathbf{X}_i' \Omega^{-1} \mathbf{X}_i) = K$

- A standard random effects analysis adds assumptions on the idiosyncratic errors that give Ω a special form.
  3. - $E(u_i u_i' \mid x_i, c_i) = \sigma_u^2 I_T$
     - $E(c_i^2 \mid x_i) = \sigma_c^2$
- Under this assumption Ω takes the following form:

$$\Omega = \begin{bmatrix} \sigma_c^2 + \sigma_u^2 & \sigma_c^2 & \cdots & \cdots & \sigma_c^2 \\ \sigma_c^2 & \sigma_c^2 + \sigma_u^2 & & & \vdots \\ \vdots & & \ddots & & \vdots \\ \vdots & & & \ddots & \sigma_c^2 \\ \sigma_c^2 & \cdots & \cdots & \sigma_c^2 & \sigma_c^2 + \sigma_u^2 \end{bmatrix}$$

Random Effects = Feasible GLS

- If we assume that we have consistent estimators of σ_u^2 and σ_c^2 (see Wooldridge, pp. 260-261 how to get consistent estimates of σ_u^2 and σ_c^2) we can obtain an estimate of Ω as.

$$\widehat{\Omega} \equiv \widehat{\sigma}_u^2 \mathbf{I}_T + \widehat{\sigma}_c^2 \mathbf{j}_T \mathbf{j}_T'$$

- where $\mathbf{j}_T \mathbf{j}_T'$ is the $T \times T$ matrix with unity in every element.
- These gives the standard random effects estimator as:

$$\widehat{\beta}_{RE} = \left(\sum_{i=1}^{N} \mathbf{X}_i' \widehat{\Omega}^{-1} \mathbf{X}_i\right)^{-1} \left(\sum_{i=1}^{N} \mathbf{X}_i' \widehat{\Omega}^{-1} \mathbf{y}_i\right)$$

- RE is one particular way of estimating a feasible GLS model (with only two estimated parameters in the variance-covariance matrix).
- If the RE assumptions are satisfied it is consistent and efficient.

Why Not Always Estimate A More Flexible FGLS?

- RE is one particular way of estimating a feasible GLS model (with only two estimated parameters in the variance-covariance matrix).
- One could also estimate a more flexible FGLS model that allows for heteroscedasticity and autocorrelation.
- If the RE assumption 3) was not satisfied this alternative model would be preferable.
- And even if assumption 3) is satisfied this alternative FGLS model would be just as efficient as RE if N is large.
- Why would we ever use RE then?
  $\Rightarrow$ If N is not several times larger than T an unrestricted FGLS analysis can have poor finite sample properties because $\widehat{\Omega}$ has many elements $(T(T+1)/2)$ that would have to be estimated.

Fixed Effects

- RE assumes that c_i is orthogonal to $\mathbf{x}_{it}$ which is a very strong assumption.
- In many applications the whole point of using panel data is to allow for arbitrary correlations of c_i with $\mathbf{x}_{it}$.
- Fixed effects explicitly deals with the fact that c_i may be correlated with $\mathbf{x}_{it}$.
- For fixed effects models we assume strict exogeneity.
  1. $E(u_{it} \mid \mathbf{x}_i, c_i) = 0, \qquad t = 1, 2, ..., T$
- Unlike the stricter RE assumption we do not assume $E(c_i \mid \mathbf{x}_i) = E(c_i)$.
- In other words $E(c_i \mid \mathbf{x}_i)$ is allowed to be any function of $\mathbf{x}_i$.
- We thus need a much weaker assumption than for RE.
- Cost: we cannot include time-constant variables in $\mathbf{x}_{it}$.

Fixed Effects - 3 Ways to Eliminate c

- In FE models there are 3 ways to eliminate c_i that causes the error term to be correlated with the regressors:
  1. Within-transformation (FE transformation).
  2. Estimating c_i with dummies.
  3. First differencing.

1. Within Estimator

- Estimating equation:

$$y_{it} = \mathbf{x}_{it}\boldsymbol{\beta} + c_i + u_{it} \qquad (1)$$

- *Step 1:* Average estimating equation over $t = 1, ..., T$:

$$\overline{y}_i = \overline{\mathbf{x}}_i\boldsymbol{\beta} + c_i + \overline{u}_i \qquad (2)$$

- Where: $\overline{y}_i = \frac{1}{T}\sum_{t=1}^{T} y_{it}, \qquad \overline{\mathbf{x}}_i = \frac{1}{T}\sum_{t=1}^{T} x_{it}, \qquad \overline{u}_i = \frac{1}{T}\sum_{t=1}^{T} u_{it}$
- *Step 2:* Substract equation (2) from equation (1) to get:

$$y_{it} - \overline{y}_i = (\mathbf{x}_{it} - \overline{\mathbf{x}}_i)\boldsymbol{\beta} + u_{it} - \overline{u}_i$$
$$\widetilde{y}_{it} = \widetilde{\mathbf{x}}_{it}\boldsymbol{\beta} + \widetilde{u}_{it}$$

- Where: $\widetilde{y}_{it} = y_{it} - \overline{y}_i, \qquad \widetilde{\mathbf{x}}_{it} = (x_{it} - \overline{\mathbf{x}}_i), \qquad \widetilde{u}_{it} = u_{it} - \overline{u}_i$
- *Step 3:* Run a regression of $\widetilde{y}_{it}$ on $\widetilde{\mathbf{x}}_{it}$ using pooled OLS.

- To ensure that the FE estimator is well behaved asymptotically we need the standard rank condition:
    2. $rank(\sum_{t=1}^{T} E(\widetilde{\mathbf{x}}_{it}'\widetilde{\mathbf{x}}_{it}) = K$
- If $\mathbf{x}_{it}$ contains an element that does not vary over time for any i then the corresponding element in $\widetilde{\mathbf{x}}_{it}$ would be 0 and the rank condition would fail.
  $\Rightarrow$ we cannot include time-invariant variables in fixed effects models.
- Without further assumptions the FE estimator is not necessarily the most efficient estimator. The next assumption ensures that it is efficient (and that we get the proper variance matrix estimator):
    3. $E(\mathbf{u}_i\mathbf{u}_i' \mid \mathbf{x}_i, c_i) = \sigma_u^2 I_T$

1. Within Estimator Standard Errors

- The standard errors from a standard OLS regression estimated under Step 3 above would not be correct. Why?
- Demeaning introduces serial correlation of the error terms:
- Variance of $\widetilde{u}_{it}$:

$$E(\widetilde{u}_{it}) = E[(u_{it} - \overline{u}_i)^2] = E(u_{it}^2) + E(\overline{u}_i^2) - 2E(u_{it}\overline{u}_i)$$
$$= \sigma_u^2 + \sigma_u^2/T - 2\sigma_u^2/T = \sigma_u^2(1 - 1/T)$$

- Covariance between $\widetilde{u}_{it}$ and $\widetilde{u}_{is}$ for $t \neq s$:

$$E(\widetilde{u}_{it}\widetilde{u}_{is}) = E[(u_{it} - \overline{u}_i)(u_{is} - \overline{u}_i)]$$
$$= E(u_{it}u_{is}) - E(u_{it}\overline{u}_i) - E(u_{is}\overline{u}_i) + E(\overline{u}_i^2)$$
$$= 0 - \sigma_u^2/T - \sigma_u^2/T + \sigma_u^2/T = -\sigma_u^2/T$$

- As a result the correlation between $\widetilde{u}_{it}$ and $\widetilde{u}_{is}$ is:

$$Corr(\widetilde{u}_{it}, \widetilde{u}_{is}) = \frac{Cov(\widetilde{u}_{it}, \widetilde{u}_{is})}{\sqrt{Var(\widetilde{u}_{it})Var(\widetilde{u}_{is})}} = \frac{-\sigma_u^2/T}{\sigma_u^2(1-1/T)} = \frac{-1}{(T-1)}$$

1. Within Estimator Standard Errors

- Assumption 3 allows us to derive an estimand of the asymptotic variance:

$$Av\hat{a}r(\widehat{\beta}_{FE}) = \widehat{\sigma}_u^2 (\sum_{i=1}^{N} \sum_{i=1}^{T} \widetilde{\mathbf{x}}_{it}' \widetilde{\mathbf{x}}_{it})^{-1}$$

- Note $\widehat{\sigma}_u^2$ is a consistent estimate for the variance of u_{it} not $\widetilde{u}_{it}$.
- As a result, we cannot get $\widehat{\sigma}_u^2$ straight from our OLS regression in step 3 above.
- The standard variance estimate from the regression in step 3 would be:

$$SSR/(NT - K)$$

- This would however be the wrong variance estimate as we need the variance of $\widehat{\sigma}_u^2$ not $\widehat{\sigma}_{\widetilde{u}}^2$.
- Note the substraction of K in the denominator does not matter asymptotically but it is standard to make such a correction.

1. Within Estimator Standard Errors

- It turns out that the variance of $\widetilde{u}_{it}$ is $\sigma_u^2(1 - 1/T)$ (see above). If we sum this across t we get:

$$\sum_{t=1}^{T} E(\widetilde{u}_{it}) = \sigma_u^2(T - 1)$$

- Further summing across N we get:

$$\sum_{i=1}^{N}\sum_{t=1}^{T} E(\widetilde{u}_{it}) = \sigma_u^2(T - 1)N \Rightarrow \sigma_u^2 = \sum_{i=1}^{N}\sum_{t=1}^{T} E(\widetilde{u}_{it})/[(T - 1)N]$$

- Thus we can get a consistent estimate for σ_u^2 by estimating the equation in step 3 and getting an estimate for σ_u^2 from:

$$SSR/(N(T - 1) - K)$$

- The difference between $SSR/(NT - K)$ and $SSR/(N(T - 1) - K)$ will be substantial when T is small.

1. Within Estimator Standard Errors

- Standard regression packages (such as STATA) will do the adjustment of standard errors automatically if you specify a fixed effects model.
- But if you wanted to estimate the fixed effects model step by step you could do the three steps outlined above and then adjust the standard errors of the regression that you obtained in step 3 above by the factor:

$$\{(NT - K)/[(N(T-1) - K)]\}^{1/2}$$

2. Dummy Variables Estimator

- An alternative way of estimating fixed effects models (especially if you have small N or if you are actually interested in the FE) would be to estimate the c_i using a set of dummies for all i in the sample.
- We would include N dummies (one for each i) in the regression and estimate:

$$y_{it} = \mathbf{x}_{it}\boldsymbol{\beta} + c_i + u_{it}$$

- using standard OLS. This is sometimes referred to as the *dummy variables estimator.*
- One benefit of this regression is that it produces the correct standard errors because it uses $NT - N - K = N(T-1) - K$ degrees of freedom.
- The cost is that it is computing power intensive if N is large.

3. First Differencing

- Yet another alternative to estimate fixed effects models would be to use first differences.
- Again we assume strict exogeneity conditional on c_i.
  1. $E(u_{it} \mid \mathbf{x}_i, c_i) = 0, \qquad t = 1, 2, ..., T$
- Lagging the model $y_{it} = \mathbf{x}_{it}\boldsymbol{\beta} + c_i + u_{it}$ by one period and substracting gives:

$$y_{it} - y_{it-1} = \mathbf{x}_{it}\boldsymbol{\beta} - \mathbf{x}_{it-1}\boldsymbol{\beta} + c_i - c_i + u_{it} - u_{it-1}$$
$$\Delta y_{it} = \Delta \mathbf{x}_{it}\boldsymbol{\beta} + \Delta u_{it}$$

- First differencing eliminates c_i.
- In differencing we lose the first time period for each cross section: we now have $T - 1$ time periods for each i instead of T.

- The first-difference (FD) estimator $\widehat{\beta}_{FD}$ is the pooled OLS estimator from the regression of:

$$\Delta y_{it} \text{ on } \Delta \mathbf{x}_{it}$$

- Under assumption FD 1 pooled OLS estimation of the first-differenced equations is consistent and unbiased.
- As above we have the rank condition for the FD estimator:
  2. $rank(\sum_{t=2}^{T} E(\Delta \mathbf{x}_{it}' \Delta \mathbf{x}_{it}) = K$
- Which again rules out time-constant explanatory variables and perfect collinearity among the time-varying variables.

3. First Differencing - Standard Errors

- Under assumptions FE1-FE3 the fixed effects (within-estimator) is asymptotically efficient.
- But FE3 assumes that the u_{it} are serially uncorrelated.
- Under the alternative assumption that Δu_{it} are serially uncorrelated the FD estimator would be efficient.
- It can be shown that with only two time periods the FD is identical to the FE estimator (try this at home: plug in the values of x and y for two time periods into the respective equations for each estimator).

Practical Tips: RE versus FE, FD, or Dummy Variables

- The RE assumption that $E(c_i \mid \mathbf{x}_i) = E(c_i) = 0$ is very strong and unlikely to hold in many cases.
- So we usually prefer fixed effect estimators.
- Which fixed effect estimator should we choose?
- With only two time periods the FE, FD and Dummy variables estimators are identical.
- if $T > 2$ the preferred estimator depends on the assumptions about the errors u_{it}.
- In practice it is more common to assume FE 3) and use FE.
- If you are interested in the FE (which we often are, see below) we would want to use the dummy variables estimator.

Introductory Example on the Use of Fixed Effects - The Effect of Unionization on Wages

- Suppose you are interested in the question whether union workers earn higher wages.
- Problem: unionized workers may be different (e.g. higher skilled, more experienced) from non-unionized workers.
- Many of these factors will not be observable to the econometrician (standard omitted variable bias problem).
- Therefore the error term and union status will be correlated and OLS will be biased.

Estimation of Fixed Effects Models

- A natural model of the effect of unionization on wages would be:

$$Y_{it} = c_i + \lambda_t + \rho D_{it} + X_{it}'\beta + u_{it} \qquad (1)$$

- Suppose you simply estimate this model with OLS (without including individual fixed effects).
- You therefore estimate:

$$\text{Y}_{it} = constant + \lambda_t + \rho D_{it} + X_{it}'\beta + \underbrace{c_i + u_{it}}_{\varepsilon_{it}}$$

- As c_i is correlated with union status D_{it} there is a correlation of D_{it} with the error term. This will lead to biased OLS estimates.
- Solution: estimate the model including individual FEs.

The Effect of Unionization on Wages - OLS vs. FE

- Freeman (1984) analyzed unionization comparing OLS and FE models for a number of datasets:

Survey	OLS	Fixed Effects
CPS 74-75	0.19	0.09
NLSY 70-78	0.28	0.19
PSID 70-79	0.23	0.14
QES 73-77	0.14	0.16

- These results suggest that union workers are positively selected.

Measurement Error and Fixed Effects Models

- OLS results were larger than FE $\rightarrow$ selection may be important.
- Another plausible explanation is measurement error.
- Measurement error introduces attenuation bias.
- As the signal to noise ratio is smaller with fixed effects (as we just use the deviations from the mean as signal) measurement error is typically a more important problem in fixed effect models.
- In this case union status may be misreported for some individuals in each year. Observed year to year changes in union status for one individual may thus be mostly noise.

Example Measurement Error

- Suppose the we have data on two individuals

	Union Status [Actual (Measured)]			
Individual	2010	2011	2012	2013
1	1	1	1 (0)	1
2	0	0	0	0

- If we ran a pooled OLS regression of union status on wages 1/8th of the observations would be mismeasured.
- Suppose you run a fixed effects model $\Rightarrow$ identification comes from changes in union status within individuals.
  - Individual 2 does not contribute to the FE estimation
  - The variation in individual 1's union status is only measurement error $\Rightarrow$ 100% of the variation is noise.

Estimating and Analyzing Fixed Effects - Bertrand and Schoar (2003)

- Sometimes explicitly estimating fixed effects can be useful because the fixed effects can inform us about parameters of interest.
- Bertrand and Schoar (2003) explicitly estimate CEO fixed effects to:
  - document that CEOs matter.
  - analyze how different FEs are correlated with performance.
- Data come from two sources:
  - Forbes 800 files (1969-1999)
  - Execucomp (1992-1999)
- Sample restriction: all firms for which at least one top executive can be observed in at least one other firm.

Bertrand and Schoar (2003) - Regression of Interest

- They estimate executive FE with the following regression:

$$y_{it} = \alpha_t + \gamma_i + \beta X_{it} + \lambda_{CEO} + \lambda_{CFO} + \lambda_{Others} + \epsilon_{it}$$

- Where y_{it} is a firm level corporate policy variable, α_t are year FE, γ_i are firm fixed effects, $\lambda's$ are executive FE where λ_{CEO} are fixed effects for the group of managers who are CEOs in the last firm where they can be observed, and so on.
- The executive FEs can only be separately identified from the γ_i if managers move across firms.
- Why do managers move across firms? If one wanted to identify the causal effect of managers on firm performance we would have to worry about these movers.

Manager Transitions Across Firms

TABLE II
EXECUTIVE TRANSITIONS BETWEEN POSITIONS AND INDUSTRIES

to:	CEO	CFO	Other
from:			
CEO	117	4	52
	63%	75%	69%
CFO	7	58	30
	71%	71%	57%
Other	106	0	145
	60%		42%

Bertrand and Schoar (2003) - How Do Manager FE Affect R-Square?

- The inclusion of manager FE increases R^2 for most outcomes. It increases from 0.91 to 0.96 for investment for example.

	Panel A: Investment policy *F-tests on fixed effects for*				
	CEOs	*CFOs*	*Other executives*	*N*	*Adjusted* R^2
Investment				6631	.91
Investment	16.74 (<.0001, 198)			6631	.94
Investment	19.39 (<.0001, 192)	53.48 (<.0001, 55)	8.45 (<.0001, 200)	6631	.96

Bertrand and Schoar (2003) - How Do Manager FE Affect R-Square?

- Managers also seem to affect the number of acquisitions, R&D, and advertising.

	Panel A: Organizational strategy *F-tests on fixed effects for*				
	CEOs	*CFOs*	*Other executives*	*N*	*Adjusted* R^2
N of diversifying acquis.				6593	.22
N of diversifying acquis.	2.06 (<.0001, 204)			6593	.25
N of diversifying acquis.	1.23 (.0163, 202)	1.74 (.0007, 53)	3.97 (<.0001, 202)	6593	.33
R&D				4283	.78
R&D	1.86 (<.0001, 145)			4283	.79
R&D	2.27 (<.0001, 143)	3.60 (<.0001, 45)	4.46 (<.0001, 143)	4283	.83
Advertising				2584	.79
Advertising	2.88 (<.0001, 95)			2584	.81
Advertising	4.03 (<.0001, 95)	0.84 (.6665, 21)	6.10 (<.0001, 80)	2584	.84
SG&A				2397	.46
SG&A	33.55 (<.0001, 123)			2397	.83
SG&A	13.80 (<.0001, 118)	0.82 (.7934, 42)	0.77 (.9777, 146)	2397	.83

Bertrand and Schoar (2003) - How Do Manager FE Affect R-Square?

- Managers also seem to affect firm performance.

	Panel B: Performance F-tests on fixed effects for				
	CEOs	*CFOs*	*Other executives*	N	Adjusted R^2
Return on assets				6593	.72
Return on assets	2.04 (<.0001, 217)			6593	.74
Return on assets	2.46 (<.0001, 201)	3.39 (<.0001, 54)	4.46 (<.0001, 202)	6593	.77
Operating return on assets				5135	.34
Operating return on assets	2.61 (<.0001, 217)			5135	.39
Operating return on assets	1.60 (<.0001, 216)	0.66 (.9788, 58)	1.01 (.4536, 217)	5135	.39

Bertrand and Schoar (2003) - Using Manager FEs to Understand "Managing Styles"

- In a second part of the paper Bertrand and Schoar investigate how the different manager fixed effects estimated (but not reported) above are correlated.
- They first generate a dataset that includes one row for each manager with his fixed effect from each of the regression above in columns.
- They then estimate regressions such as:

$$FE(y_j) = \alpha + \beta FE(z_j) + \epsilon_j$$

Bertrand and Schoar (2003) - Using Manager FEs to Understand "Managing Styles"

- Column (1) last row: there is negative correlation between investment and aquisitions. This suggests that some managers grow firms by investing while others grow firms by buying other firms.

	Investment	Inv to Q	Inv to CF
Investment			
Inv to Q sensitivity	**6.8**		
	(0.92)		
Inv to CF sensitivity	0.02	**−0.23**	
	(0.6)	**(0.11)**	
Cash holdings	−1.10	−0.79	−0.46
	(1.62)	(1.71)	(1.72)
Leverage	−0.39	−0.28	−0.63
	(0.55)	(0.59)	(0.60)
R&D	**0.07**	**0.08**	**−0.03**
	(0.00)	**(0.02)**	**(0.01)**
Advertising	0.01	**0.02**	−0.01
	(0.01)	**(0.01)**	(0.01)
N of acquisitions	**−0.27**	0.08	**0.23**
	(0.11)	(0.10)	**(0.10)**

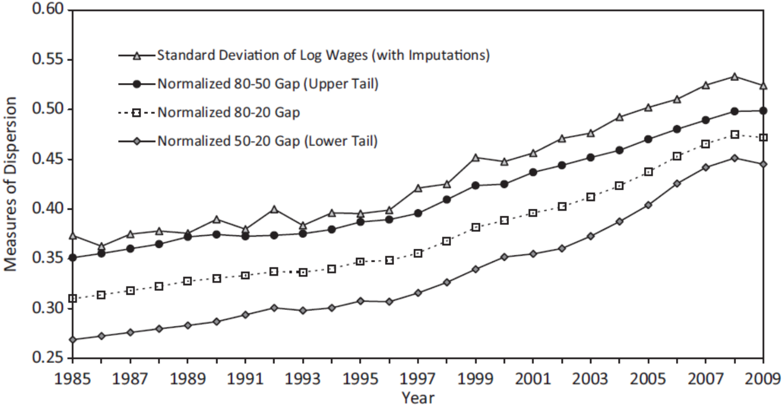
Bertrand and Schoar (2003) - Are Manager FEs Correlated with Compensation?

- They then investigate whether the different FEs are correlated with compensation.
- Row (1): managers who have positive FEs for returns on assets also receive higher compensation.

	Percent shares held by large block holders	*Residual compensation*	
		Total compensation	*Salary compensation*
Return on assets	0.012	0.72	2.86
	(0.006)	(0.24)	(0.57)
Investment	0.278	0.02	−0.08
	(0.252)	(0.01)	(0.06)
Inv to Q sensitivity	0.246	0.08	0.19
	(0.053)	(0.03)	(0.13)
Inv to CF sensitivity	−0.004	−0.06	−0.06
	(0.088)	(0.04)	(0.07)
Cash holdings	−0.001	−0.02	−0.26
	(0.007)	(0.15)	(0.29)
Leverage	−0.018	0.04	−0.01
	(0.021)	(0.26)	(0.18)

Card, Heining, and Kline (2013) - Use Firm and Worker FE to Decompose Changes in German Wage-Inequality

- Wage inequality has increased in Germany since the 1980s.



Card, Heining, and Kline (2013) - Empirical Specification

- They estimate the following wage equation with establishment ($\psi_{\mathbf{J}}$) and worker fixed effects (α_i):

$$y_{it} = \alpha_i + \psi_{\mathbf{J}} + x'_{it}\beta + r_{ir}$$

- Where the error term has a RE structure that allows for the correlation of errors within a worker-establishment pair.
- In matrix notation the model can be written as:

$$y = D\alpha + F\psi + X\beta + r = Z'\xi + r$$

- Where $Z \equiv [D, F, X]$ and $\xi \equiv [\alpha', \psi', \beta']$
- Because they have about 85-90 million person-year observations they do not estimate the model using standard fixed effects procedures but rely on methods that allows them to solve for ξ in:

$$Z'Z\xi = Z'y$$

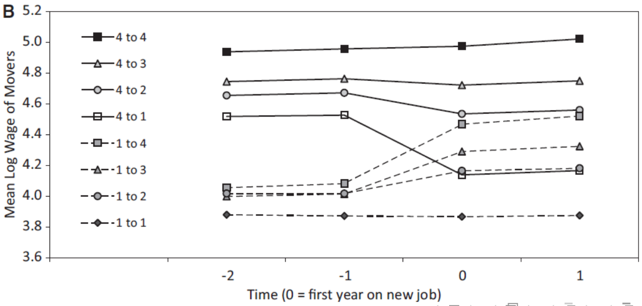
- without inverting $Z'Z$ (as in Abowd, Kramarz, and Margolis, 1999)

Card, Heining, and Kline (2013) - Assumptions

- Establishment and worker FE can be separately estimated because workers move across establishments.
- Card, Heining, and Kline are very clear about the assumptions that need to hold for their approach to estimate consistent estimates.
- As with all fixed effects models they need strict exogeneity.
- They discuss scenarios under which this assumption would not be satisfied. "Endogenous mobility" that violate strict exogeneity would occur if:
  1. There was sorting based on the idiosyncratic match component of wages (i.e. the component of compensation that comes from a good match of workers and firms).
  2. If workers with rising wages (e.g. due to a drift component in the error term) are more likely to move to high wage jobs.
  3. If workers with a positive wage shock are more likely to move (if workers cycle to high-wage but less stable jobs when the particular industry is doing well and they move to low-wage but stable jobs if the industry is doing badly).

Card, Heining, and Kline (2013) - Wages of Workers Who Move Jobs

- To rule out the concerns above they show that wages of workers who move are relatively stable before a move (evidence against concerns 2) and 3)) and that wage increases of workers who move from low-wage to high-wage firms are almost symmetric to wage decreases of workers who move from high-wage to low-wage firm (evidence against concern 1) otherwise workers would always gain from moving)



- The variance of observed wages for workers in a given sample interval can be decomposed as:

$$Var(y_{it}) = Var(\alpha_i) + Var(\psi_{\mathbf{J}}) + Var(x'_{it}\beta)$$
$$+ 2Cov(\alpha_i, \psi_{\mathbf{J}}) + 2Cov(\psi_{\mathbf{J}}, x'_{it}\beta) + 2Cov(\alpha_i, x'_{it}\beta) + Var(r_{it})$$

- They use a feasible version of this decomposition by replacing each term with the corresponding sample analogue.

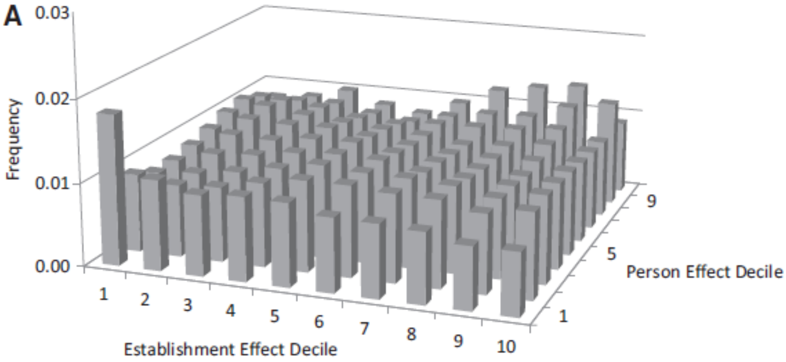
Card, Heining, and Kline (2013) - Variance Decompositions - Results

	(1) Interval 1 1985–1991	(2) Interval 2 1990–1996	(3) Interval 3 1996–2002	(4) Interval 4 2002–2009
Person and establishment parameters				
Number person effects	16,295,106	17,223,290	16,384,815	15,834,602
Number establishment effects	1,221,098	1,357,824	1,476,705	1,504,095
Summary of parameter estimates				
Std. dev. of person effects (across person-year obs.)	0.289	0.304	0.327	0.357
Std. dev. of establ. Effects (across person-year obs.)	0.159	0.172	0.194	0.230
Std. dev. of Xb (across person-year obs.)	0.121	0.088	0.093	0.084
Correlation of person/establ. Effects (across person-year obs.)	0.034	0.097	0.169	0.249
Correlation of person effects/Xb (across person-year obs.)	−0.051	−0.102	−0.063	0.029
Correlation of establ. effects/Xb (across person-year obs.)	0.057	0.039	0.050	0.112
RMSE of AKM residual	0.119	0.121	0.130	0.135
Adjusted R-squared	0.896	0.901	0.909	0.927

Card, Heining, and Kline (2013) - Variance Decompositions

- The person effects and establishment FE become more variable over time.
- The variation in the covariate index $x'_{it}\beta$ falls over time.
- The residual standard deviation of wages (measure by RMSE) is relatively small and relatively stable over time.
- The correlation of person effects and establishment effects increased massively over time (-$>$ sorting of high-wage workers into high-wage firms)

Card, Heining, and Kline (2013) - Graphical Evidence for Increase Assortative Matching - First Period



Card, Heining, and Kline (2013) - Graphical Evidence for Increase Assortative Matching - Last Period