Quantitative Macro-Labor
Empirical Regularities and Panel Data

Professor Griffy

UAlbany

Fall 2024

Announcements

- I still need to email about the campus cluster–moving was chaos.
- You can run Matlab/Python/R/Julia code on the campus cluster, and it will email you when done.
- Stata should work as well, let me know if you want to use it.

Panel Data

- What is panel data?
  1. Repeated surveys of the same individuals.
  2. Surveys contain repeated questions, thus comparable across time/age.
  3. Generally, introduction of *new* cohorts, thus allowing time and age effects to be disentangled.
- Why is it useful?
  1. Repeated individual observations help separate marginal effects of observables from innate ability.
  2. (With enough data), individual fixed effects control for time-invariant innate characteristics.
  3. Can control for geography-by-time trends, as well as the marginal effects of other "nuissance" covariates.
- Excellent discussion of use in macroeconomics: Browning, Heckman, and Hansen (1999).

"Fundamental Equation of (Reduced-Form) Labor Economics"

- Standard regression analysis:

$$y = X\beta + \epsilon \tag{1}$$

- Most of "reduced-form" labor economics comes down to arguing the following:

$$E[X\epsilon] = 0 \tag{2}$$

$$\text{or } E[\epsilon|X] = 0 \tag{3}$$

- i.e., that your covariates are uncorrelated with the error term,
- or alternatively that you aren't capturing variation with a covariate that is actually caused by an omitted variable.
- If you can successfully argue this, you have argued for *ex-post* identification.
- Note: "reduced-form" is not intended as a pejorative.

Two Basic Panel Models

- Fixed Effects:
  - "Fixed Effect Model" means that you have *individual* (or firm, etc.) fixed effects in your regression.
  - i.e., an intercept for every individual.
  - Don't confuse this with *using* fixed effects, i.e., state, year.
- Random Effects:
  - There is an individual unobserved heterogeneity, but it is *random*, i.e., uncorrelated with your observable characteristics.
  - I can't come up with a good example of this, and in almost every case people use fixed effects models.
- We'll focus briefly on the fixed effects model.

Fixed Effects Model

- Generic linear regression:

$$y_{it} = x_{it}\beta + c_i + \epsilon_{it} \tag{4}$$

- c_i is the individual heterogeneity/effect.
- Typically, we would run

$$y_{it} = x_{it}\beta + \epsilon_{it} \tag{5}$$

- But, this would be wrong if $E[X_{it}c_i] \not= 0$.
- "The point of using panel data is to allow c_i to be arbitrarily correlated with the x_{it}" (Wooldridge, 2002).
- Some good references:
  - "Econometric Analysis of Cross Section and Panel Data": Wooldridge (2002)
  - "Mostly Harmless Econometrics": Angrist and Pischke (2009)

Fixed Effects Model

- Generic linear regression:

$$y_{it} = x_{it}\beta + c_i + \epsilon_{it} \tag{6}$$

- How do we solve this problem?
  - "within transformation" FE estimator
  - "first-difference" estimator
- Within-transformation: difference out the mean over time of each observation

$$y_{it} - \bar{y}_i = x_{it}\beta + c_i + \epsilon_{it} - \bar{y}_i \tag{7}$$
$$y_{it} - \bar{y}_i = x_{it}\beta + c_i + \epsilon_{it} - \bar{x}_i\beta - \bar{c}_i - \bar{\epsilon}_i \tag{8}$$
$$y_{it} - \bar{y}_i = (x_{it} - \bar{x}_i)\beta + \cancel{(c_i - \bar{c}_i)} + (\epsilon_{it} - \bar{\epsilon}_i) \tag{9}$$
$$y_{it} - \bar{y}_i = (x_{it} - \bar{x}_i)\beta + (\epsilon_{it} - \bar{\epsilon}_i) \tag{10}$$
$$\tag{11}$$

- Identical to having an indicator variable for each individual.

First-Difference Approach

- Generic linear regression:

$$y_{it} = x_{it}\beta + c_i + \epsilon_{it} \tag{12}$$

- First-difference estimator:

$$y_{it} - y_{it-1} = x_{it}\beta + c_i + \epsilon_{it} - y_{it-1} \tag{13}$$
$$y_{it} - y_{it-1} = x_{it}\beta + c_i + \epsilon_{it} - x_{it-1}\beta - c_i - \epsilon_{it-1} \tag{14}$$
$$y_{it} - y_{it-1} = (x_{it} - x_{it-1})\beta + \cancel{(c_i - c_i)} + (\epsilon_{it} - \epsilon_{it-1}) \tag{15}$$
$$y_{it} - y_{it-1} = (x_{it} - x_{it-1})\beta + (\epsilon_{it} - \epsilon_{it-1}) \tag{16}$$

- Identical to fixed-effects estimator if errors not serially correlated.

Guvenen (2009)

- What Guvenen ultimately ends up estimating:

$$y^i_{h,t} = g(\theta^0_t, X^i_{h,t}) + c^i + d^i \times t + z^i_{h,t} + \phi_t \epsilon^i_{h,t} \tag{17}$$

$$y^i_{h,t} = g(\theta^0_t, X^i_{h,t}) + c^i + d^i \times t + \hat{\epsilon}^i_{h,t} \tag{18}$$

$$c^i : \textit{Ind. FE that affects intercept} \tag{19}$$

$$d^i : \textit{Ind. FE that affects slope} \tag{20}$$

- where $\hat{\epsilon}^i_{h,t}$ includes all unobserved components (persistent and transitory shocks).
- Estimate this jointly with transition equation for z_{it} to recover ρ, σ_ϵ and σ_ν.

Fixed Effects Model

- What does this mean?
  - If unobserved heterogeneity is **not** time-varying,
  - and we have correctly specified our model,
  - we can identify the marginal effect, β, of each covariate in x_{it}.
- We are using "within individual" variation to identify the effects.
- Potential problems:
  - The covariates of interest may also be time-invariant.
  - May have relatively few individual-level observations.
  - Then, we would use between-individual variation and try to argue that our inference can be interpreted causally.
  - Or use a structural model to try and interpret our results.

Some Valuable Micro-Data Sources

- In class, we will typically discuss "micro-data":
  1. Panel Study of Income Dynamics (PSID): a panel of households from 1968-present, annually.
  2. National Longitudinal Survey of Youth (NLSY79, NLSY97): Two separate cohorts interviewed repeatedly at an annual frequency 1979-present and 1997-present.
  3. Survey of Income and Program Participation (SIPP): Series of panels that last 3-4 years. Each panel contains new participants.
  4. Current Population Survey (CPS): The standard for labor market information. A monthly survey that is representative. Some panel dimensions, but note that this lacks important panel components that the others retain.
- I will upload some code to the lab storage.
- Link to a good description on website.

Panel Study of Income Dynamics

- Longitudinal study of a representative sample of US individuals and their families from 1968-present.
- New individuals enter and exit, meaning many cohorts.
- Excellent panel for life-cycle analysis (almost the exclusive source of data).
- Good labor market information: employment spells, income, wages, some employer-to-employer and job-to-job mobility.
- The bad:
  - Annual frequency.
  - (potentially) substantial measurement error.
  - Can be hard to work with: variables renamed each year.
- Can be used for intergenerational analysis as well (only dataset that can).

Survey of Income and Program Participation

- The SIPP is a series of short panels, rarely more than 3 years in length.
- Conducted annually 1984-1993, then in 1996, 2001, 2004, 2008.
- Households are assigned a "rotation group," and interviewed every four months about the previous four months.
- Great for labor market information: weekly labor force status, income, hours, wages, UI, mobility, etc.
- The bad:
  - It's a very short panel: no life-cycle components
  - Might only observe a single unemployment spell by individuals
  - Problems with censoring because of survey length
  - Survey design is a little tricky
- Probably best publicly available panel data for labor market.

National Longitudinal Survey of Youth (1979)

- The NLSY is (sort-of) a medium between the PSID and the SIPP.
- A cohort of 14-22 year olds are surveyed identical questions each year from 1979 to present.
- Has very detailed labor market information, and can be at a monthly frequency.
- Best (IMHO): has Armed Forces Qualifying Test (AFQT) scores, which are a rough measure of individual ability.
- Also has relatively consistent wealth observations.
- The bad:
  1. Annual frequency;
  2. Single cohort;
  3. Geographic information only available in restricted version.
- Easiest of the 3 to work with.

Micro-Data

- Which should you work with? Depends on the question.
- Broadly,
  1. If you aren't interested in life-cycle effects, choose the SIPP.
  2. If you are, need geographic location, or need to separate time and age effects, choose the PSID.
  3. If you want a measure of individual ability, the AFQT, choose the NLSY.
- If you aren't interested in the panel dimension, choose the CPS.

Labor Market Empirical Regularities

- What are some topics that are worth exploring in the labor market?
- Davis and Haltiwanger (1999) talk about six:
  1. Employer lifecycle dynamics;
  2. Worker reallocation and productivity growth;
  3. Worker reallocation over business cycles;
  4. Lumpiness, heterogeneity, and aggregation;
  5. Reasons for worker mobility;
  6. Worker sorting and job assignment.
- They argue that each of these topics (at least at the time) had unanswered questions.
- Davis and Haltiwanger papers (there are a lot) are a good source of "empirical regularities"
- FYI: they also use the term "empirical regularities"

Davis and Haltiwanger (1999)

- Lots of interest in worker flows.
- They are purely interested in measurement.
- Challenges:
  1. Few matched employer-employee data sets.
  2. Aggregation issues: flows between plants within same firm, etc.
  3. Few matched employer-employee panels, i.e., can't separate worker and firm fixed effects.

Davis and Haltiwanger (1999) - Key Definitions

- Primary Definitions:
  - (Gross) job creation at time t equals employment gains summed over all business units that expand or start up between $t-1$ and t.
  - (Gross) job destruction at time t equals employment losses summed over all business units that contract or shut down between $t-1$ and t.
- Secondary Definitions:
  - (Gross) job reallocation: job creation + job destruction
  - Excess job reallocation: job reallocation − net employment change
  - (Gross) **worker** reallocation: movement across place of employment.
- Excess reallocation: the amount of job reallocations *over and above* the amount required to accomodate net employment changes.

Measurement and Job/Worker Flows

- Tricky Proposition:
  - Flow is necessarily a continuous variable;
  - Surveys yield snapshots at various points in time.
  - Workers may transition jobs before being observed, i.e., $E_t, U_{t+0.5\times\Delta}, E_{t+\Delta}$.
- Definitions change over time.
- Workers flow E-O and O-E, so even out of labor force are searching for jobs.
- Rob Shimer has a lot of good work on this as well.

Previous work across countries

International comparison of annual gross job flow rates (annual averages as percentages of employment)

Country	Period	Coverage	Employer unit	Job creation	Job destruction	Net growth	Job reallocation
Australia	1984–1985	Manufacturing	Establishments	16.1	13.2	3.9	29.3
Canada	1974–1992	Manufacturing	Establishments	10.9	11.1	−0.2	21.9
Canada	1983–1991	All employees	Firms	14.5	11.9	2.6	26.3
Chile	1976–1986	Manufacturing	Establishments	13.0	13.9	−1.0	26.8
Colombia	1977–1991	Manufacturing	Establishments	12.5	12.2	0.3	24.6
Denmark	1983–1989	Private sector	Establishments	16.0	13.8	2.2	29.8
Denmark	1981–1991	Manufacturing	Establishments	12.0	11.5	0.5	23.5
Estonia	1992–1994	All employees	Firms	9.7	12.9	−2.2	22.6
Finland	1986–1991	All employees	Establishments	10.4	12.0	−1.6	22.4
France	1984–1992	Private sector	Establishments	13.9	13.2	0.6	27.1
France	1985–1991	Manufacturing	Firms	10.2	11.0	−0.8	21.2
France[a]	1985–1991	Non-manufacturing	Firms	14.3	11.8	2.4	26.1
Germany	1983–1990	All employees	Establishments	9.0	7.5	1.5	16.5
Germany (Lower Saxony)	1979–1993	Manufacturing	Establishments	4.5	5.2	−0.7	9.7
Italy[b]	1984–1993	Private sector	Firms	11.9	11.1	0.8	23.0
Israel	1971–1972	Manufacturing	Establishments	9.7	8.2	1.5	17.9
Morocco	1984–1989	Manufacturing	Firms	18.6	12.1	6.5	30.7
Netherlands	1979–1993	Manufacturing	Firms	7.3	8.3	−1.0	15.6
New Zealand	1987–1992	Private sector	Establishments	15.7	19.8	−4.1	35.5
Norway	1976–1986	Manufacturing	Establishments	7.1	8.4	−1.2	15.5
Sweden	1985–1992	All employees	Establishments	14.5	14.6	−0.1	29.1
USA	1973–1993	Manufacturing	Establishments	8.8	10.2	−1.3	19.0
USA[c]	1979–1983	Private sector	Establishments	11.4	9.9	1.4	21.3
USA[c]	1979–1983	Manufacturing	Establishments	10.2	11.5	−1.3	21.6
United Kingdom	1985–1991	All employees	Firms	8.7	6.6	2.1	15.3

Between vs. within employers

Job reallocation rates between and within employers[a]

Study	Coverage	Employer unit	Between reallocation rate	Within reallocation rate
Hamermesh et al. (1996, Table 2)	Netherlands, all sectors, 1988–1990	Firms	6.2	0.8
Lagarde et al. (1994, Table 1)	France, all sectors, 1984–1991	Establishments	7.9	6.7
Dunne et al. (1997, Table 5)	USA, manufacturing 1972–1988	Establishments	19.2	2.7

By industry

Average annual job reallocation rates by country and industry[a]

	USA 1974–1992	Canada 1974–1992	Netherlands 1979–1993	Norway 1976–1986
Food	17.9	19.5	18.4	15.3
Tobacco	12.7	12.3		
Textiles	16.9	21.3	19.1	18.3
Apparel	25.2	27.8	23.4	
Lumber	25.8	26.2	20.8	15.7
Furniture	20.7	27.7		
Paper	12.5	11.1	14.6	12.6
Printing	17.1	22.0	16.3	
Chemicals	14.0	18.7	12.1	12.7
Petroleum	14.2	15.6	10.1	13.2
Rubber	20.3	21.5	12.1	
Leather	22.4	24.2	17.5	
Stone, clay, glass	20.4	23.0	15.6	
Primary metals	16.0	13.3	5.2	6.3
Fabricated metals	20.0	27.7	18.8	18.7
Non-electric machinery	20.5	27.8	16.4	
Electric machinery	19.5	24.6	11.3	
Transportation	18.4	20.6	14.6	
Instruments	10.5	28.1	19.7	
Miscellaneous	14.4		28.5	18.3
Total manufacturing	19.0	21.9	15.6	15.5

Excess Reallocation

Fraction of excess job reallocation accounted for by employment shifts between sectors

Country	Period	Classification scheme	Unit of analysis	Number of sectors	Average number of workers per sector (in 000's)	Fraction resulting from shifts between sectors
USA	1972–1988	4-Digit SIC manufacturing	Plant	448/456	39.1[a]	0.13
USA	1972–1988	2-Digit SIC manufacturing by state	Plant	980	17.9	0.14
Denmark	1983–1989	1-Digit ISIC private sector	Plant	8	196.1	0.00
Finland	1986–1991	2-Digit ISIC	Plant	27	48.9	0.06
Germany	1983–1990	2-Digit ISIC	Plant	24	1171.2	0.03
Italy	1986–1991	2-Digit ISIC private sector	Firm	28	321.5	0.02
Netherlands	1979–1993	2-Digit SIC	Firm	18	10.0	0.20
Sweden	1985–1991	2-Digit ISIC	Plant	28	112.4	0.03
Norway	1976–1986	5-Digit ISIC manufacturing	Plant	142	2.4	0.06
France	1984–1988	NAP private sector	Plant	15	883.3	0.06
France	1985–1991	Detailed industry	Firm	600	36.6	0.17
France	1984–1991	NAP	Plant	100		0.12
New Zealand	1987–1992	2-Digit ISIC	Plant	28	27.5	0.01
Chile	1979–1986	4-Digit manufacturing	Plant	69	3.7	0.12
Colombia	1977–1991	4-Digit manufacturing	Plant	73	6.31	0.13
Morocco	1984–1989	4-Digit manufacturing	Plant	61	4.0	0.17

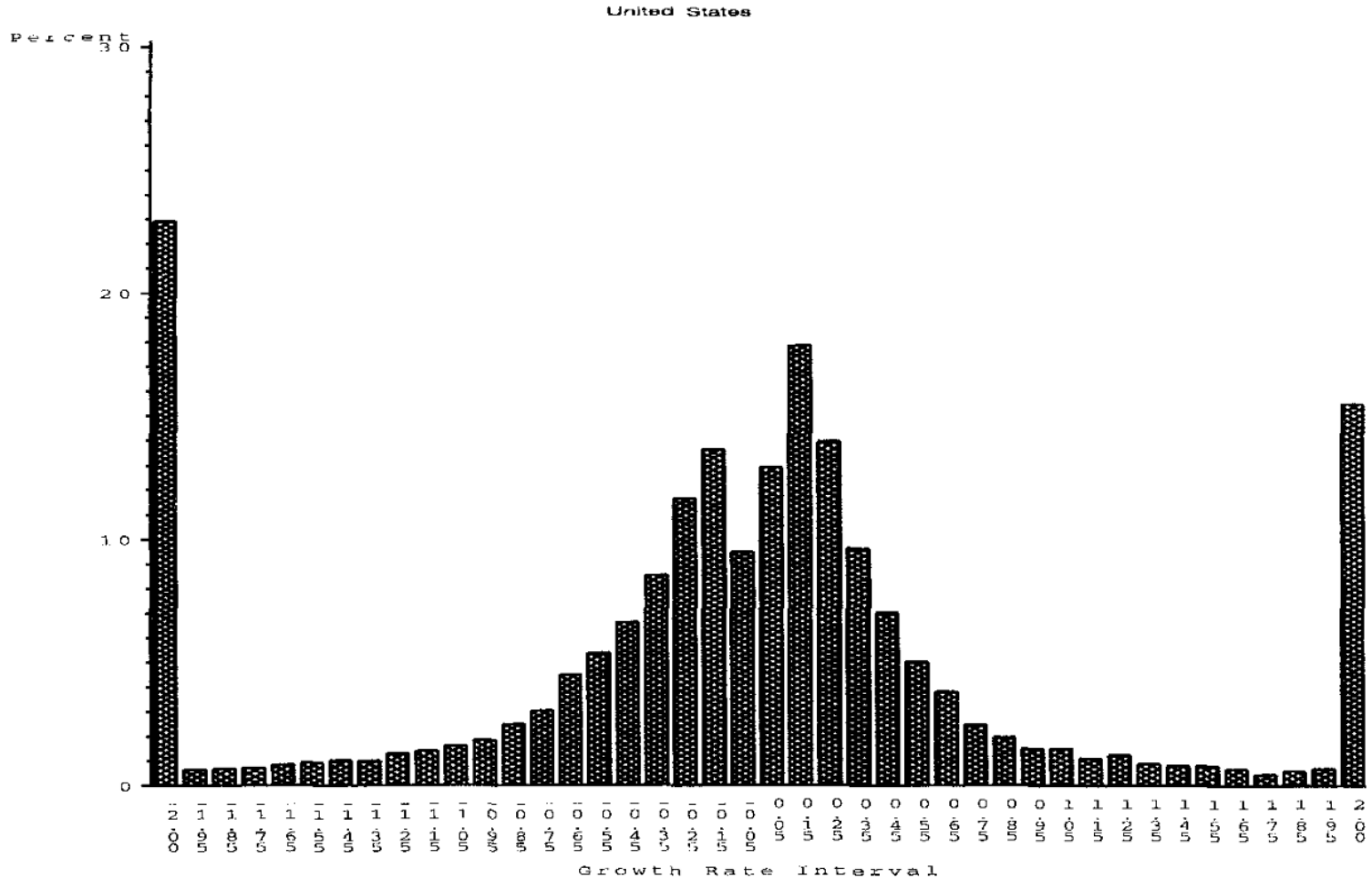
Persistence

- Persistence of job creation: % of jobs at time t that remain filled at $t + n$
- Persistence of job destruction: % of jobs at time t that do not reappear by $t + n$

Average persistence rates for annual job flows

	USA 1973–1988		Denmark 1980–1991		Netherlands 1979–1993		Norway 1977–1986		France 1985–1990	
	1 year	2 years	1 year	2 years	1 year	2 years	1 year	2 years	1 year	2 years
Job creation	70.2	54.4	71.0	58.0	77.9	58.8	72.7	65.1	73.4	51.5
Job destruction	82.3	73.6	71.0	58.0	92.5	87.3	84.2	79.8	82.1	68.2
Sources	Davis et al. (1996, Table 2.3)		Albaek and Sorensen (1996, Table 3)		Gautier (1997, Tables 3.16 and 3.19)		Klette and Mathiassen (1996, Table 5 and 4)		Nocke (1994, Table 4)	

Firm Distribution of Growth Rates

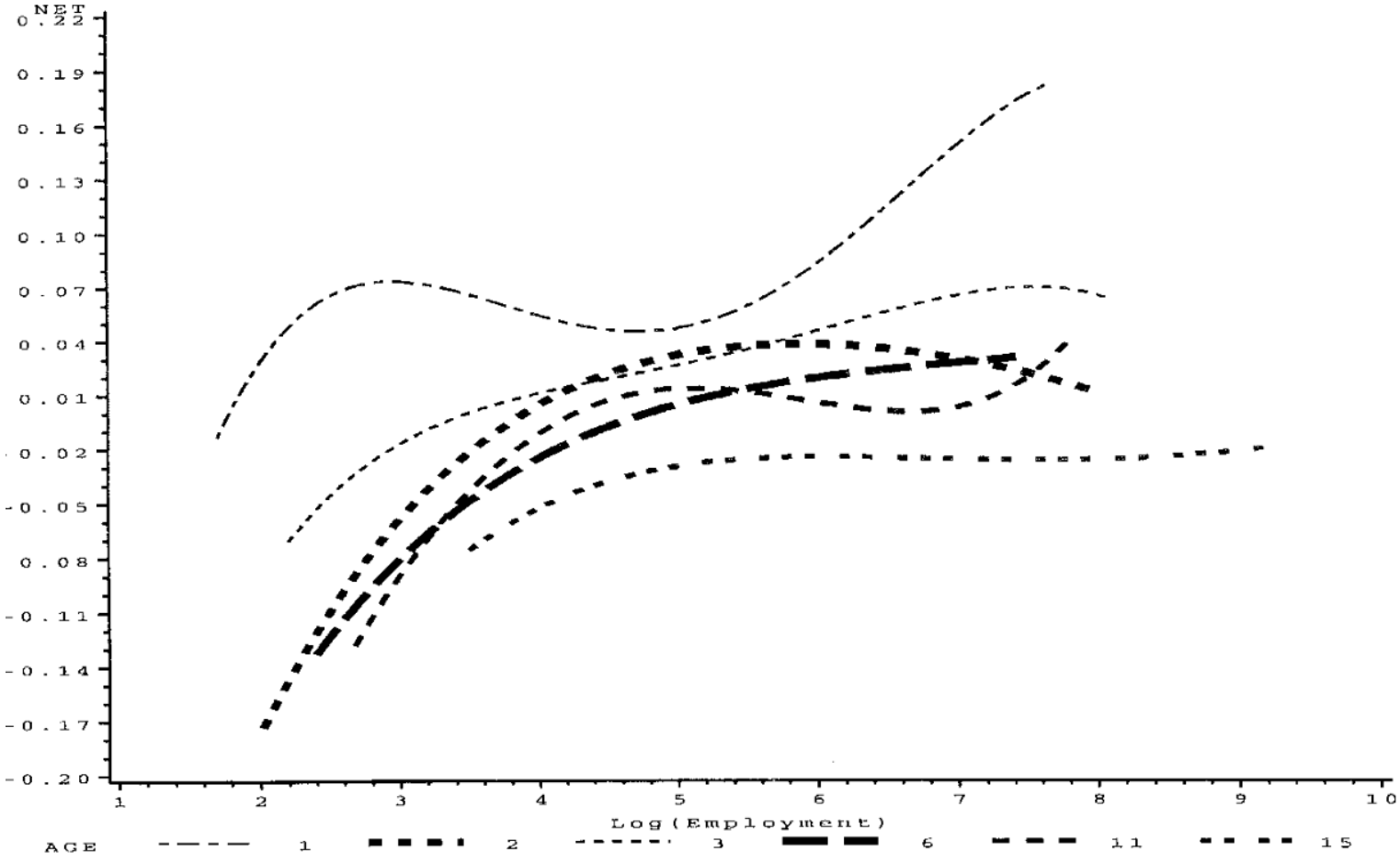


Firm Distribution of Growth Rates

The concentration of job creation and job destruction[a]

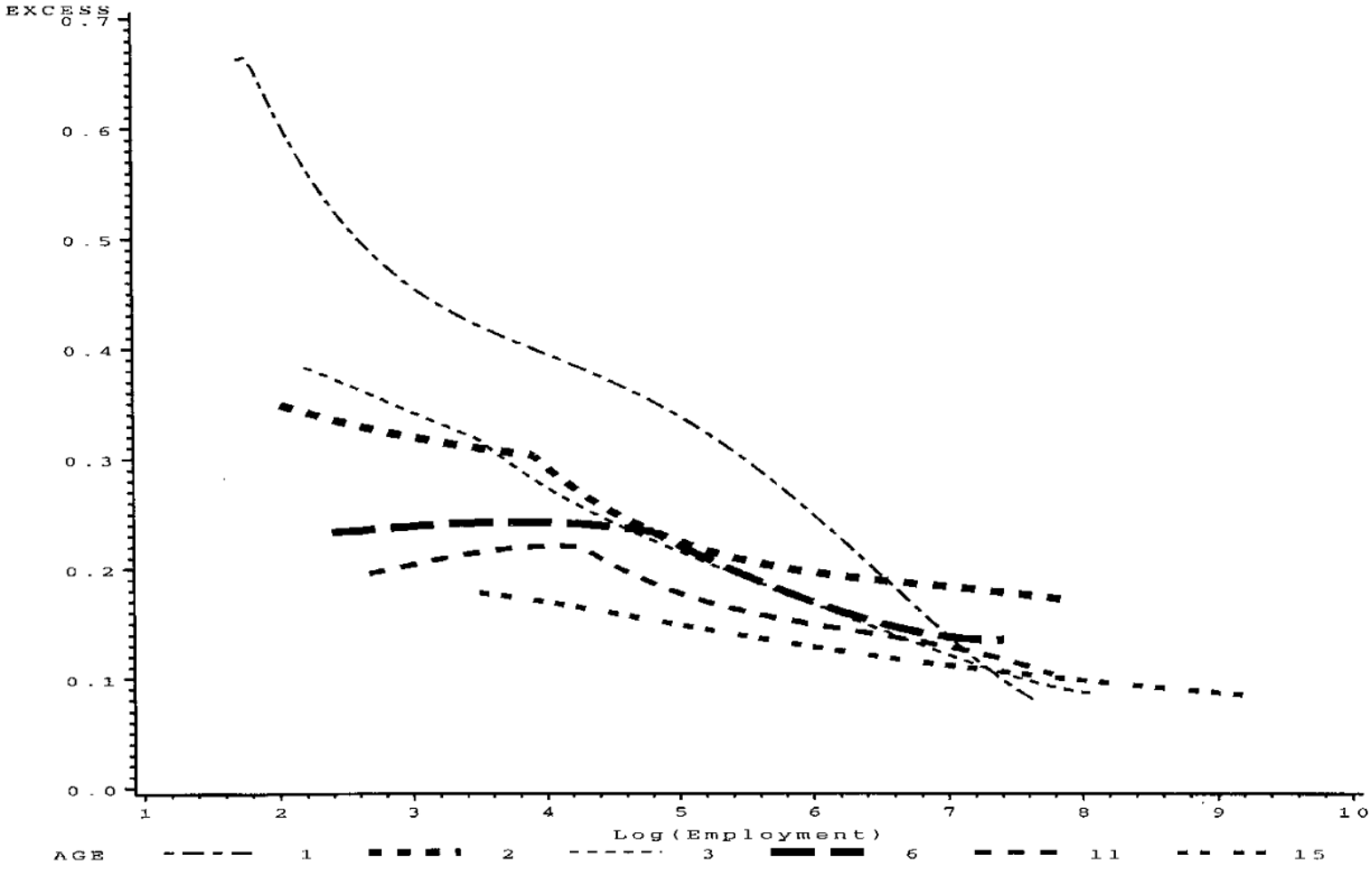
Country	Sector	Percent of job creation or destruction accounted for by plants with growth rates in the indicated interval					
		[−2,−1)	[−1,−0.2)	[−0.2,0)	(0,0.2]	(0.2,1]	(1,2]
United States	Manufacturing	32.9	44.0	23.1	30.7	45.1	24.2
Canada	Manufacturing	77.7		22.3	24.8	75.2	
Denmark	Manufacturing	45.9	33.7	20.4	23.4	37.4	39.1
Israel	Manufacturing	84.7		15.3	21.8	78.2	

Firm Size



- Declines with age.

Firm Size



- Excess reallocation: the amount of job reallocations *over and above* the amount required to accomodate net employment Δ.
- Declines with age.

Business Cycles

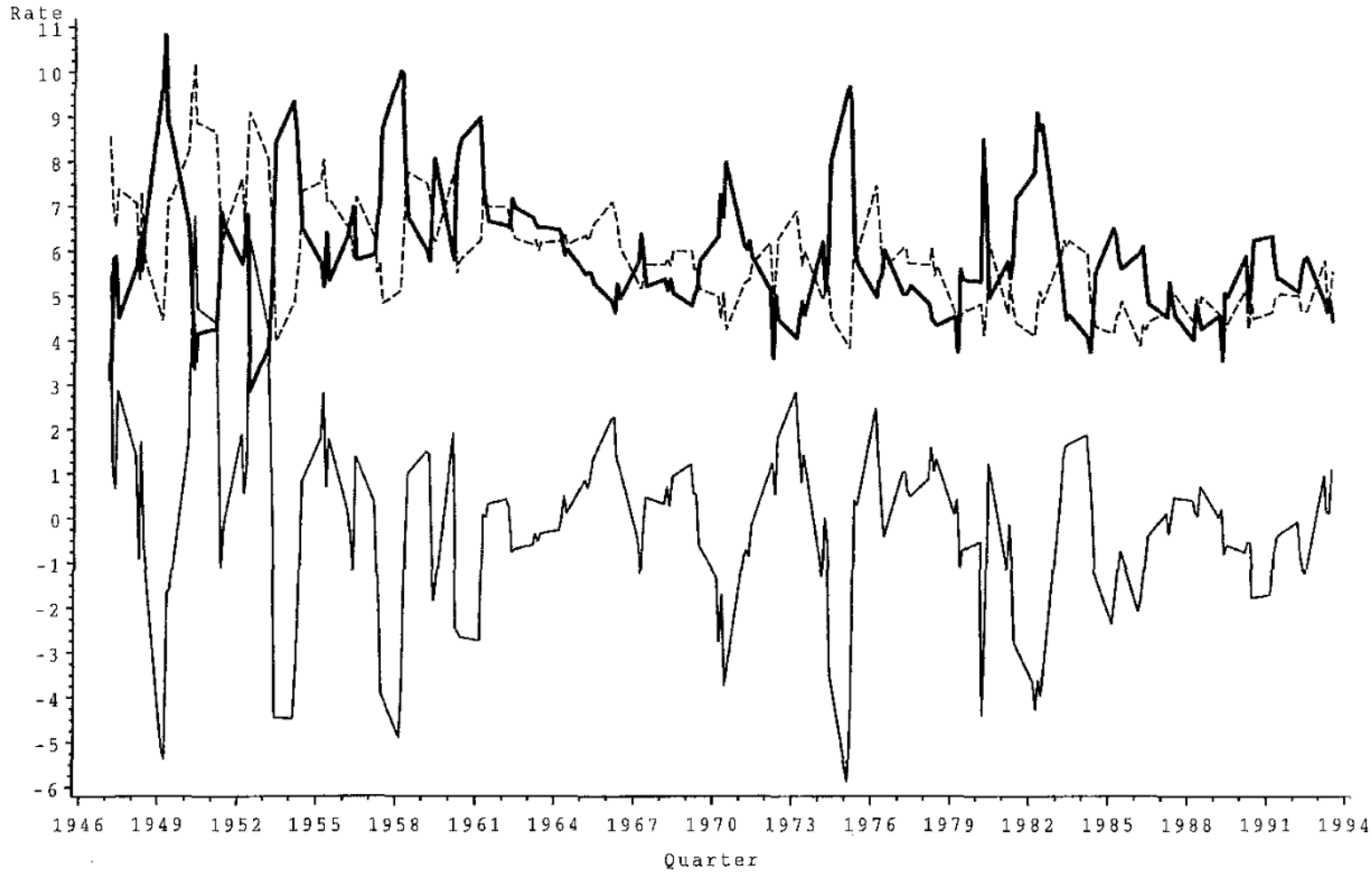


Fig. 4. Job creation and job destruction in US manufacturing. Dashed line, job creation; heavy solid line, job destruction; light solid line, net growth.

A Follow-Up: Davis, Faberman, Haltiwanger (2006)

- Again on measurement (they've done a lot of good work on it).
- Comparison of results across different datasets.
  - JOLTS: Job Openings and Labor Turnover Survey
  - BED: Business Employment Dynamics
  - LEHD: Longitudinal Employer Household Dynamics
- Updated with new papers and findings.

Firm Distribution of Growth Rates

Job and Worker Flow Rates by Sampling Frequency and Data Source

Sampling Frequency and Data Source	*Job creation*	*Job destruction*	*Hires*	*Separations*
Monthly				
JOLTS, continuous monthly units from microdata, Dec. 2000 to Jan. 2005	1.5	1.5	3.2	3.1
Quarterly				
JOLTS, continuous quarterly units from microdata, Dec. 2000 to Jan. 2005	3.4	3.1	9.5	9.2
BED, all private establishments, 1990:2–2005:1	7.9	7.6	—	—
LEHD, all transitions, ten selected states, 1993:2–2003:3	7.0	6.0	25.0	24.0
LEHD, "full-quarter" transitions, ten selected states, 1993:2–2003:3	7.6	5.2	13.1	10.7
Annual				
BED, from Pinkston and Spletzer (2004), private establishments, 1998–2002	14.6	13.7	—	—

Firm Distribution of Growth Rates

Job and Worker Flows by Selected Industries

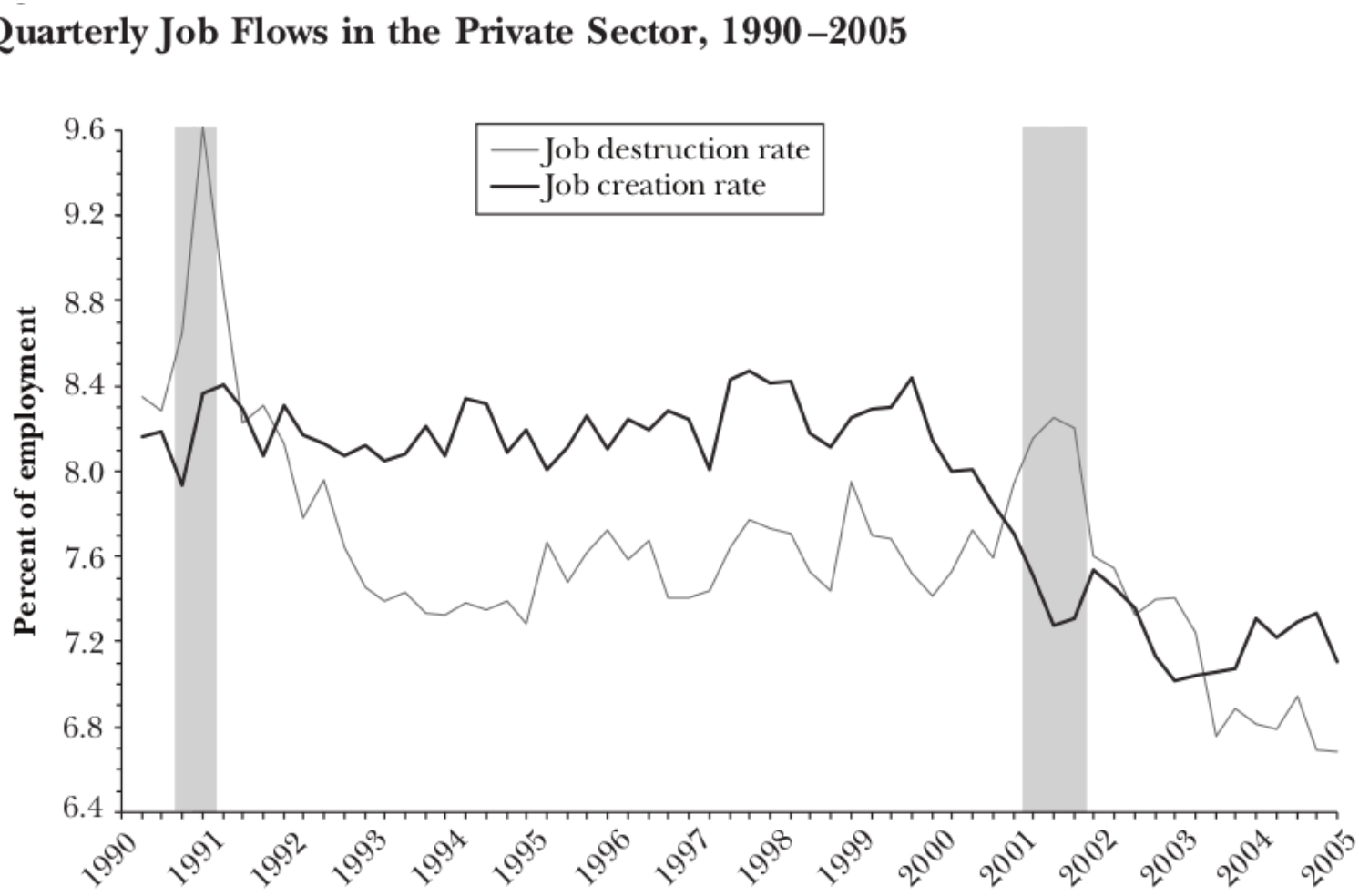
A. Average Quarterly Job Flow Rates in the BED, 1990:2–2005:1

	Job creation	*Job destruction*	*Net growth*
Total private	7.9	7.6	0.3
Construction	14.3	13.9	0.4
Manufacturing	4.9	5.3	−0.4
Retail trade	8.1	7.9	0.2
Professional & business services	9.9	9.1	0.8
Leisure & hospitality	10.7	10.2	0.5

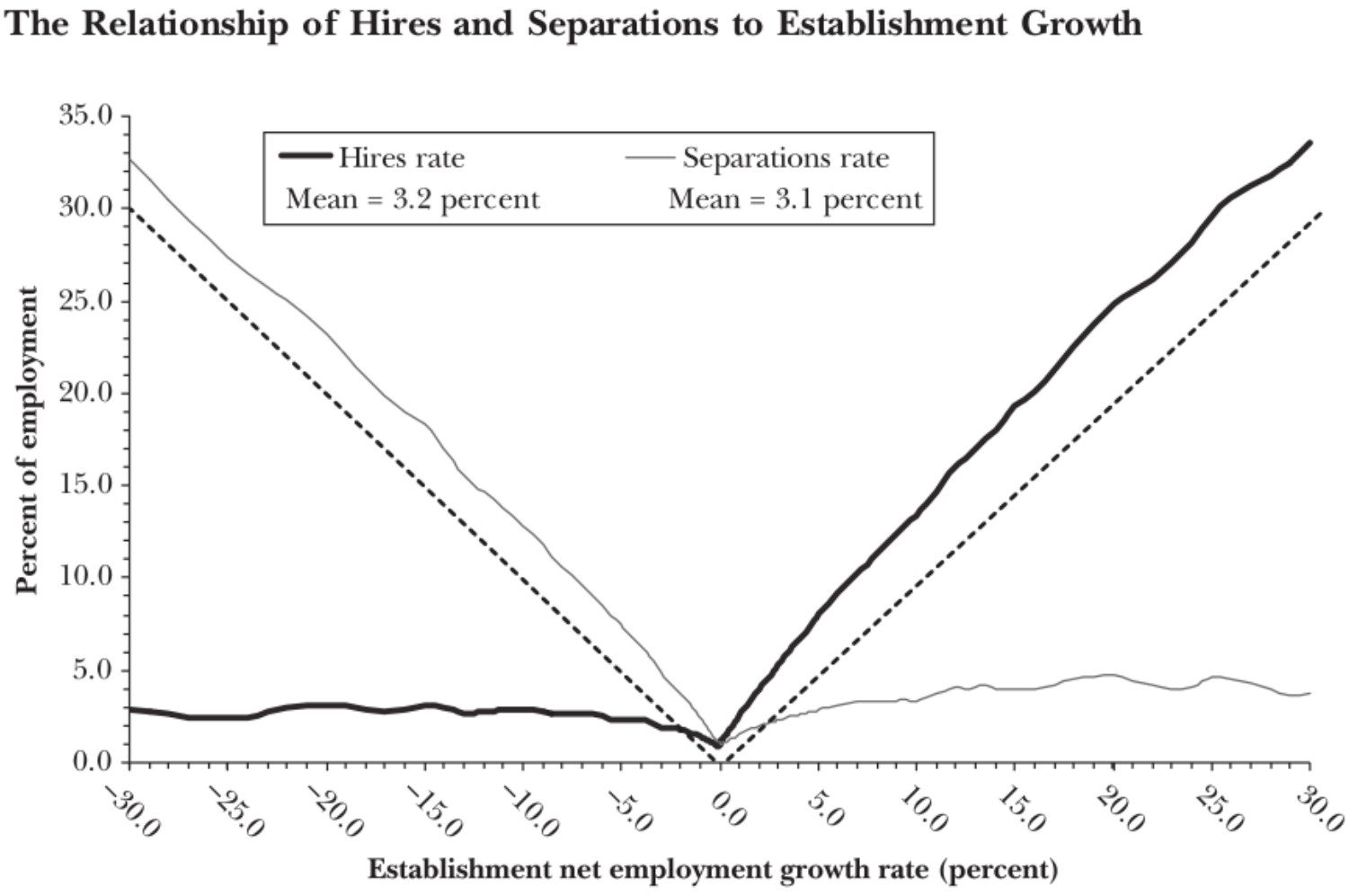
B. Average Monthly Worker Flow Rates in JOLTS, December 2000 to January 2005

					Layoffs per	
	Hires	*Separations*	*Quits*	*Layoffs*	*Quit*	*Destroyed job*
Total nonfarm	3.2	3.1	1.7	1.1	0.7	0.8
Construction	5.3	5.5	2.1	3.2	1.5	1.1
Manufacturing	2.2	2.7	1.2	1.2	1.1	0.8
Retail trade	4.3	4.2	2.6	1.3	0.5	0.7
Professional & business services	4.2	3.9	2.0	1.6	0.8	1.0
Leisure & hospitality	6.1	5.9	3.9	1.8	0.5	0.7

Firm Distribution of Growth Rates



Firm Distribution of Growth Rates



Next Time

- Job search: how can we explain wage dispersion?
- The McCall Model.
- Read Rogerson, Shimer, Wright (2005).
- Make sure you've installed some programming languages.