Regressing individual outcomes on group characteristics

- y scalar outcome (student's grade), $x \in R^J$ attributes of an individual's reference group, $z \in R^k$ observable individual attributes that affect directly y (e.g parent's income), scalar non-observable individual attribute $u = x'\delta + \epsilon$, with ϵ i.i.d. and $E[\epsilon] = 0$. (ability)

-

$$y = \alpha + \beta E[y|x] + E[z|x]'\gamma + z'\eta + x'\delta + \epsilon \quad (1)$$

  - β endogenous effect.
  - γ exogenous effect.
  - δ correlated effect.

- $$E[y|x,z] = \alpha + \beta E[y|x] + E[z|x]'\gamma + z'\eta + x'\delta \quad (2)$$

- Equilibrium condition:

$$E[y|x] = \alpha + \beta E[y|x] + E[z|x]'\gamma + E[z|x]'\eta + x'\delta \quad (3)$$

- If $\beta \neq 1$ the solution is unique and is:

$$E[y|x] = \frac{\alpha}{1-\beta} + E[z|x]'\frac{\gamma+\eta}{1-\beta} + \frac{x'\delta}{1-\beta} \quad (4)$$

- $E[y|x]$ is a linear function of $[1, E(z|x), x]$.

- Cannot distinguish endogenous from exogenous or correlated effects.

- Using (3) in (2) we obtain:

$$E[y|x,z] = \frac{\alpha}{1-\beta} + E[z|x]'\frac{\gamma+\beta\eta}{1-\beta} + z'\eta + x'\frac{\delta}{1-\beta} \quad (5)$$

- Identify $\frac{\alpha}{1-\beta}, \frac{\gamma+\beta\eta}{1-\beta}, \eta$ and $\frac{\delta}{1-\beta}$ if $[1, E(z|x), x, z]$ are linearly independent.

- If $\gamma = \delta = 0$,

$$E[y|x,z] = \frac{\alpha}{1-\beta} + E[z|x]'\frac{\beta\eta}{1-\beta} + z'\eta \quad (6)$$

- Can identify, $\frac{\alpha}{1-\beta}, \frac{\beta\eta}{1-\beta}$ and η.

- $1 + \frac{\beta}{1-\beta}$ is the social multiplier.

- Randomized experiments.

An example with random groups

- Peer effects with random assignment: Results for Dartmouth roommates (Sacerdote)

- Freshmen students at Dartmouth are assigned at random except for sex and response to four questions: smoking, study with music, neat or messy, keep late hours?

- Have data on academic achievement pre College (including, private vs public school, grades,

Table I (from Sacerdote)
Summary Statistics for Sample of Dartmouth Roommates
Graduating Classes of 1997 and 1998

Variable	*Obs*	*Mean*	*Std. Dev.*	*Min*	*Max*
freshman year GPA	1589	3.20	0.43	0.67	4.00
sophomore year GPA	1552	3.28	0.44	0.30	4.00
junior year GPA	1529	3.35	0.45	0.60	4.00
senior year GPA	1508	3.41	0.45	0.50	4.00
roommate freshman year GPA	1589	3.19	0.39	1.15	4.00
fraternity/sorority/co-ed house	1589	0.49	0.50	0.00	1.00
graduate late	1589	0.03	0.18	0.00	1.00
economics major	1589	0.10	0.31	0.00	1.00
social science major	1589	0.33	0.47	0.00	1.00
science major	1589	0.29	0.45	0.00	1.00
humanities major	1589	0.35	0.48	0.00	1.00
black	1589	0.05	0.22	0.00	1.00
SAT Math	1589	691.26	67.08	420.00	800.00
SAT Verbal	1589	632.86	70.07	360.00	800.00
academic score (incoming)	1589	204.20	12.88	151.00	231.00
high school class rank (incoming)	993	9.14	12.27	1.00	75.00
high school class rank missing	1589	0.38	0.48	0.00	1.00
private high school	1589	0.11	0.32	0.00	1.00
smokes (housing form)	1589	0.01	0.12	0.00	1.00
more neat than messy (housing form)	1589	0.69	0.46	0.00	1.00
stays up late (housing form)	1589	0.60	0.49	0.00	1.00
listens to music (housing form)	1589	0.47	0.50	0.00	1.00
same roommate sophomore year	1589	0.14	0.35	0.00	1.00
HS GPA	1328	3.56	0.51	2.00	4.00
Pre-Dart: drank beer in past year	1337	0.59	0.49	0.00	1.00

Notes: Use of beer in past year is coded 0-1 as follows: 0=not at all, occasionally or frequently=1. Use of beer and high school GPA come the UCLA Higher Education Research Institute's Survey of Incoming Freshman. Housing form variables come from Dartmouth's Office of Residential Life. All other data are from Dartmouth's Computing Services Group.

Sample consists of all members of the classes of 1997 and 1998 minus the following four groups: students who were assigned to singles (222), students for whom I could not find housing forms (209), students assigned to the substance free dorm (26), and students who were able to request a specific roommate (135).

Table II (from Sacerdote)
Own Pre-treatment Characteristics Regressed On Roommate Pre-treatment Characteristics
Evidence of the Random Assignment of Roommates

	(1) *SAT Math (self)*	*(2)* *SAT Verbal (self)*	*(3)* *HS Academic Index*	*(4)* *HS Class Rank*	*(5)* *HS Academic Index*
roommates' math SAT scores	-0.025 (0.028)				-0.005 (0.008)
roommates' verbal SAT scores		-0.009 (0.029)			-0.005 (0.007)
roommates' HS academic scores			0.010 (0.028)		0.055 (0.056)
roommates' HS class ranks				-0.032 (0.028)	0.031 (0.042)
roommates' HS class rank missing					-0.512 (0.838)
Dummies for housing questions	yes	yes	yes	yes	yes
F test: All roommate background coeff=0					F(5,1543) =0.50 P>F = .78
R^2	.09	.03	.04	.03	.04
N	1589	1589	1589	993	1589

Standard errors are in parentheses. In cases with more than one roommate, roommate variables are averaged.

Columns (1)-(5) are OLS. All regressions include 41 dummies representing non-empty blocks based upon responses to the housing questions.

The lack of statistical significance on the coefficients is intended to demonstrate that the assignment process resembles a randomized experiment. In earlier non-randomly assigned classes (such as the classes of 1995-1996), own and roommate background are highly correlated.

- $$\begin{aligned} y_i &= \delta + \alpha(z_i + m_i) + \gamma(z_j + m_j) \\ &+ \beta y_j + \epsilon_i \end{aligned} \tag{7}$$

- y_i GPA at freshman year, z_i summary of academic background, m_i is measurement error in academic background

- β measures endogenous effect, γ endogenous effect

- $$\begin{aligned} & (1-\beta^2)y_i = (1+\beta)\delta + (\alpha + \gamma\beta)z_i \\ + \; & (\gamma + \beta\alpha)z_j + (\alpha + \gamma\beta)m_i + \\ & (\gamma + \beta\alpha)m_j + \beta\epsilon_j + \epsilon_i \end{aligned} \tag{8}$$

- $$y_i = \pi_0 + \pi_1 z_i + \pi_2 z_j + \eta \tag{9}$$

Table III (from Sacerdote)
Peer Effects in Academic Outcomes

	(1) Fresh year GPA	*(2) Fresh year GPA w/ dorm f.e.*	*(3) Senior year GPA*	*(4) Fresh year GPA*	*(5) Fresh year GPA*	*(6) Fresh year GPA*	*(7) Grad -uate late*	*(8) Econ Major*
Roommates' GPA	0.120** (0.039)	0.068** (0.029)	0.008 (0.026)					
HS academic score (self)	0.014** (0.0008)	0.015** (0.0007)	0.013** (0.0009)				-0.0001 (0.0003)	0.003** (0.0006)
HS academic score (roommates')	-0.001 (0.001)	-0.0003 (0.0009)	0.0009 (0.001)				0.0003 (0.0003)	-0.0001 (0.0006)
roommates' academic score bottom 25%				0.016 (0.028)	0.014 (0.025)	0.017 (0.025)		
roommates' academic score top 25%				0.060** (0.028)	0.047* (0.026)	0.043* (0.026)		
roommates' intention to graduate w/ honors (1-4)						0.082** (0.037)		
own academic score bottom 25%					-0.284** (0.025)	-0.282** (0.025)		
own academic score top 25%					0.174** (0.025)	0.175** (0.025)		
Roommate graduate late							0.008 (0.029)	
Roommate econ major								-0.018 (0.026)
Dummies for housing questions	yes	yes	yes	yes	yes	yes	yes	yes
F test of Roommate background coefficients=0				F=2.31 P=0.10	F=1.63 P=0.20	F=2.74 P=0.04		
R^2	.24	.38	.18	.05	.19	.19	.06	.07
N	1589	1589	1441	1589	1589	1589	1589	1589

Standard errors are in parentheses and are corrected for clustering at the room level. In cases with more than one roommate, roommate variables are averaged. **=p-value<.05. *=p-value<.10.

Regression (1) is OLS of own GPA on roommate GPA and controls. If own and roommate academic indices are excluded, the coefficient on roommate GPA falls to .111 and the standard error falls to 0.037. If roommate GPA is excluded, the coefficient on roommate academic index remains small and insignificant.

Regression (2) adds dorm fixed effects. The coefficient on roommate GPA falls, but remains significant. Regression (3) is OLS of own senior year GPA on freshman year roommates' senior year GPA. Senior year GPA includes all grades in final year and excludes grades from earlier years.

Regressions (4)-(6) are OLS of own GPA on own and roommate background. These regressions use dummies for own and roommate academic index are in the bottom 25%, middle 50% (excluded category), or top 25% of their respective distributions. Regression (4) shows that "roommate top 25%" is significant in predicting own GPA. The level of significance on "roommate top 25%" falls to .10 when two dummies for own academic index are added. (This is regression (5).) Regression (6) shows that roommate intention to graduate with honors also predicts own GPA. This variable is a self-assessed probability of graduating with honors and is coded as a 1,2,3 or 4 for the responses of no chance, very little chance, some chance, or a very good chance. Regression (6) also includes a dummy for "roommate intend to graduate with honors" missing. See text for more discussion of this variable.

Regressions (7) and (8) are probits of own "graduate late" and own "major choice=econ" on roommate graduate late and roommate major choice=econ. $\partial y/\partial x$ is shown.

Social interactions and unemployment

- Topa, *Review of Economic Studies,* 2002.

- Motivation: distribution of unemployment rates in Chicago display a lot of spacial correlation. Try to distinguish between sorting of types and social interactions.

- Employed agents transmit information about job opening to unemployed "contacts."

- In US 50% of all new jobs are found through friends, relatives, neighbors, colleagues.

  - especially true for low-skill jobs, less educated and blacks.

  - Studies show contacts live nearby.

- Contact process (Harris)

Figure 1: Map of Unemployment rate, 1980.

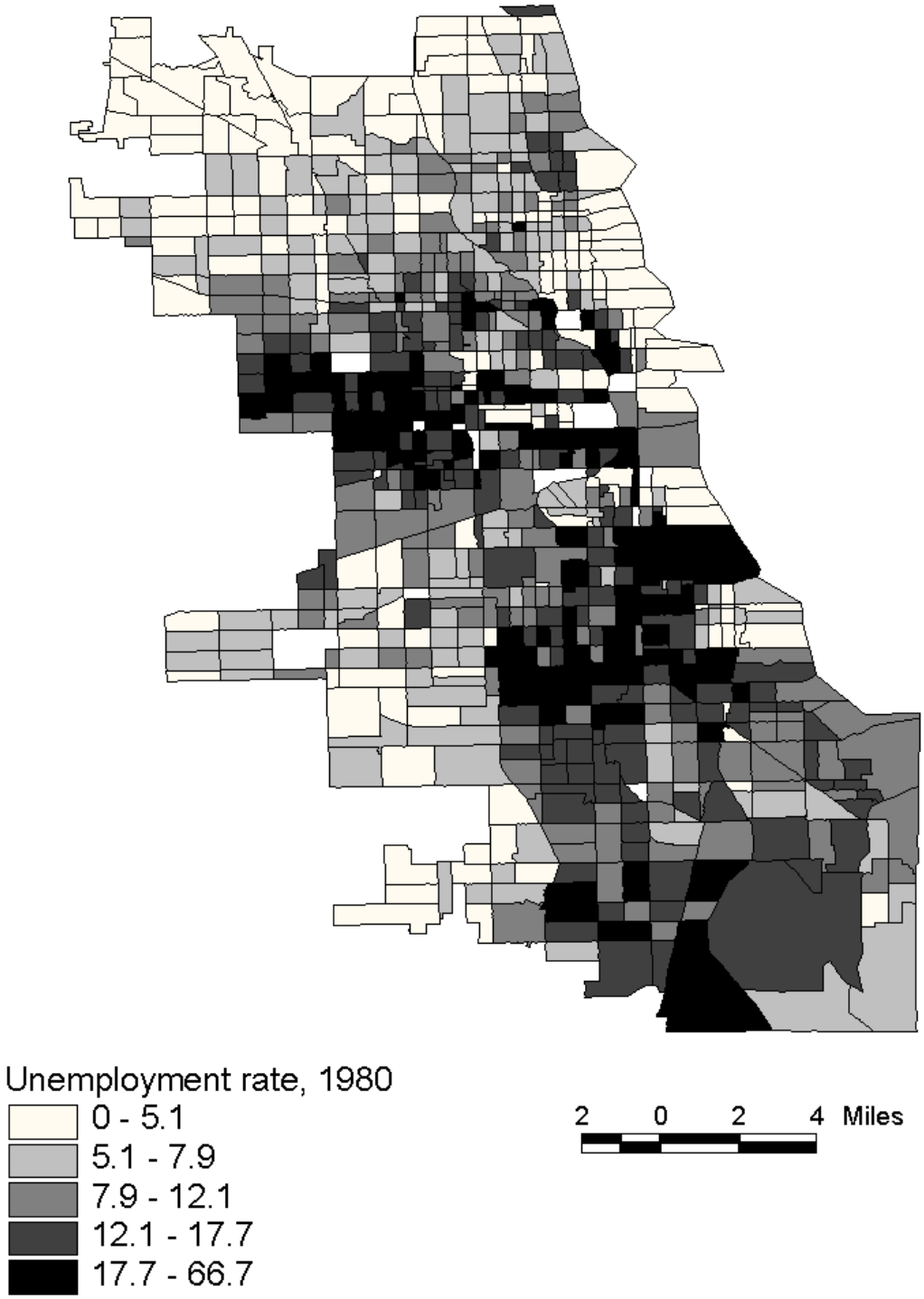
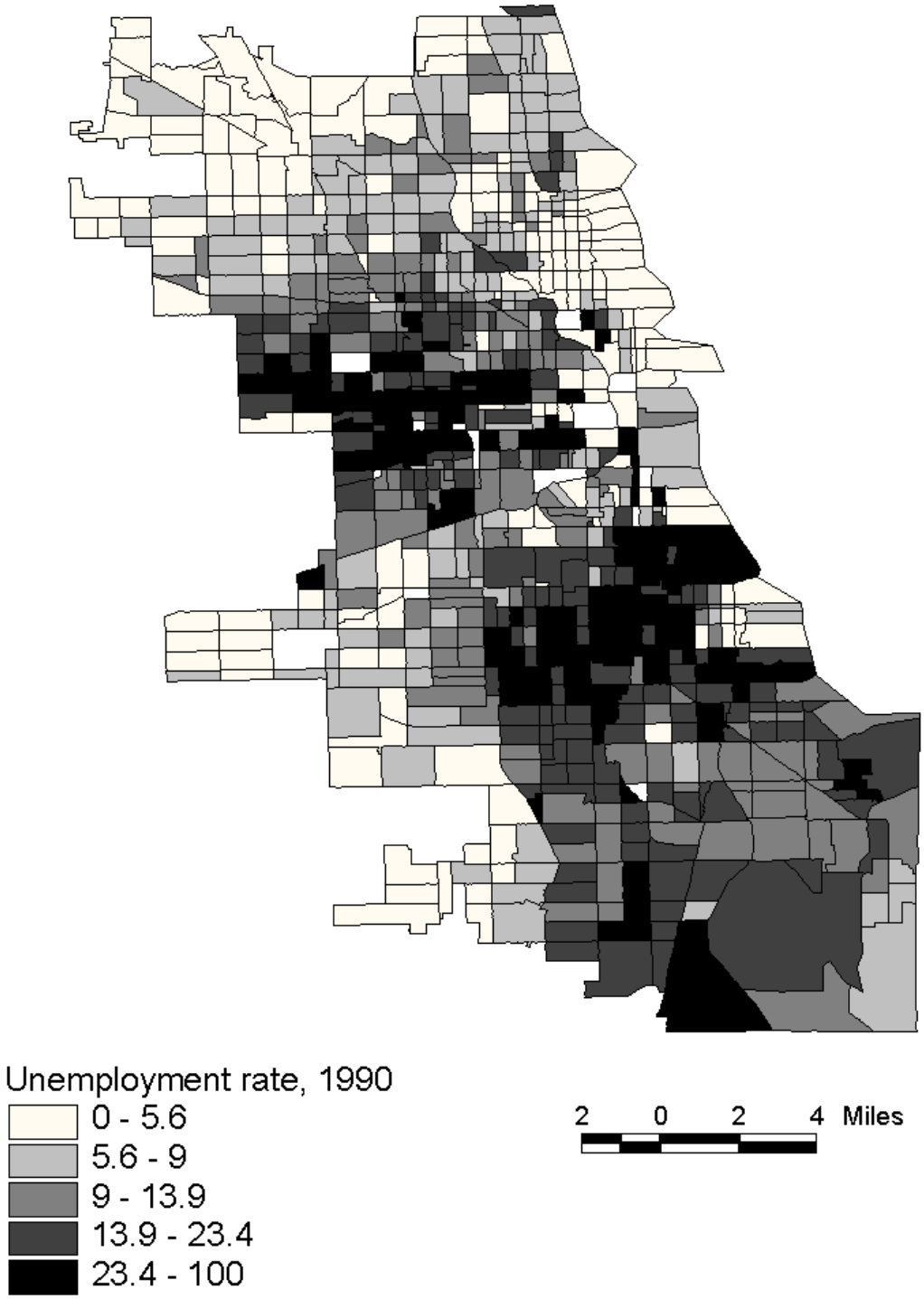


Figure 2: Map of Unemployment rate, 1990.



- Unemployment in model evolves according to a Markov process over locations.
  - employed individuals become unemployed with exogenous probabilities.
  - unemployed individuals become employed with a probability that depends on the number of employed contacts.
  - transitions also depend on neighborhood characteristics.

- Model estimated using census tracts, assume that agents in a tract have contacts in a nearby tract.

- Model estimated using indirect inference (Gourieroux, Monfort and Renault.)

The formal model

- Locations (tract) indexed by $i \in S$ a finite set.
  - Chicago's tracts have on average 2,500 people (16+).
  - aggregate from individual level. (Conley and Topa (2000).
- Location i has vector of characteristics X_i, % non-white, % that finished college, crime rate, ...
- The state of tract y_{it} is the employment rate. It assumes values in $E = \{0, 0.1, 0.2, \cdots, 1\}$.
- A tract is adjacent to another if they share an edge in the physical map.

- Parametrization
  - $p_d = \gamma(X_i)$, $p_u = \alpha(X_i) + \lambda(X_i)\bar{y}_t$.
  - functions are linear, $\bar{y}_t$ the average employment rate of neighbors, p_u, $p_d < 1$.
  - if state is interior y_{it} goes up (down) with prob. $\frac{1}{2}p_u$ ($\frac{1}{2}p_d$), if $y_{it} = 1$ (0) it goes down (up) with probability p_d (p_u).
  - shocks can be correlated
  - variation: divide neighboring tracts into those with similar ethnic composition and others. Lower impact of the others.

- Estimate $\theta \in \mathbb{R}^p$, the vector of parameters of the functions γ, α, λ.

- Look at stationary distribution of simulated model and compare with empirical cross section distribution.

Indirect inference

- Write auxiliary model parameterized by $\rho \in \mathbb{R}^q$, $q \geq p$.

- Estimate of auxiliary model using data is $\widehat{\rho}(\theta_0)$, where θ_0 is true value of parameter of structural model.

- Simulate data of structural model for different $\theta \in \Theta$ and estimate $\tilde{\rho}(\theta)$.

- Choose $\widehat{\theta}$ to minimize the distance between $\widehat{\rho}(\theta_0)$ and $\tilde{\rho}(\theta)$.

- If enough data this works (consistency), but need "good" auxiliary model for efficiency.

Auxiliary model

- Spatial auto regression.
- $y_i = \sum_{j>0}^{K} \phi_j y^{N_{ij}} + (\phi_0 \cdot x_i) y^{N_{i1}} + \beta \cdot x_i + \epsilon_i.$
- $y^{N_{ij}}$ is the average employment in tracts at distance j from i, where a tract has distance 1 if it shares an edge, distance 2 if it shares an edge with a tract that shares an edge etc...
- ϵ_i uncorrelated, with mean zero and variance σ_i^2.
- Also add some raw spatial moments of the unemployment variable to the auxiliary model.

Some empirical conclusions

- $\lambda = 0$ is rejected, λ goes down with education, up with non-white.

- When ethnicity of neighbors is considered, same ethnicity raises effect of neighbors employment level.

- Effect of one standard deviation on the average unemployment of neighbors is 1.2%, similar to effect of one standard deviation on educational level (0.8%).

- Consider two poor tract and two rich ones.

  - Giving the poorest (Grand Boulevard) the neighbors of the richest (Lake View) lowers unemployment by 4% points.

Figure 9: Map of *simulated* unempl. rate, 1990-80 (COR).

TABLE 1

Summary statistics of all variables

Variables	1980		1990		1990-80	
	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
Unemployment rate	11.71	8.16	14.93	12.06	3.22	8.95
Segregation index	87.20	13.03	91.82	11.52	4.62	14.32
Non-white persons (%)	50.55	39.63	57.91	37.14	7.36	14.59
Hispanic persons (%)	7.19	12.92	19.53	26.36	12.34	17.24
High-school grad. or more (%)	53.65	16.94	62.53	17.20	8.88	10.23
College grad. or more (%)	10.86	12.91	15.62	17.37	4.77	8.62
Crime Index	22.59	15.51	19.39	13.50	-3.20	8.04
Females (%)	52.91	6.63	51.81	5.29	-1.10	5.68
Persons 18-24 y.o. (%)	19.35	6.70	15.94	7.06	-3.41	6.19
Persons 0-24 y.o. (%)	41.74	12.03	38.01	11.97	-3.73	7.74
Persons per household (average)	2.77	0.62	2.82	0.67	0.05	0.40
Vacant housing units (%)	7.36	5.99	10.48	8.18	3.12	8.14
Median gross rent	227.42	55.22	252.07	73.22	24.65	51.77
Average housing value	10.05	13.52	26.18	22.30	16.13	19.60
Employed in prfs/mngr jobs (%)	9.59	8.91	12.77	11.55	3.18	7.12
Out of labor force - males (%)	30.29	13.22	30.48	14.03	0.19	12.08
Out of labor force - females (%)	51.01	12.56	45.82	13.73	-5.19	11.87
Same county workers (%)	83.89	12.02	94.60	5.46	10.71	12.41
Median travel time to work	27.26	6.38	27.99	6.52	0.73	5.96

See Appendix A for a definition of all the variables and units.

TABLE 3

Structural parameter estimates

Parameter	Baseline	Ethnicities	Comm. Areas	lambda = 0	Corr. shocks	lambda = 0
lambda_0	0.2495	0.2057	0.1966		0.3001	
	(0.0104)	(0.0103)	(0.0211)		(0.0254)	
lambda_EDN(80)		0.0024				
		(0.0017)				
lambda_EDN(90)		-0.2094				
		(0.0226)				
lambda_NCA			-0.0159			
			(0.0046)			
lambda_ed	-0.1049	-0.1558	-0.0922		-0.1691	
	(0.0109)	(0.0175)	(0.0182)		(0.0312)	
lambda_nw	0.0910	0.1101	0.2718		-0.1657	
	(0.0107)	(0.0090)	(0.0328)		(0.0157)	
lambda_cr	-0.2592	-0.3262	-0.3583		-0.3510	
	(0.0257)	(0.0175)	(0.0165)		(0.0470)	
alpha_0	0.3919	0.3724	0.3719	0.6505	0.7528	0.2313
	(0.0170)	(0.0139)	(0.0225)	(0.0158)	(0.0352)	(0.0064)
alpha_ed	0.0778	0.0814	0.0926	0.0737	0.1662	0.1319
	(0.0046)	(0.0068)	(0.0095)	(0.0054)	(0.0274)	(0.0174)
alpha_nw	-0.1248	-0.1017	-0.1234	-0.1109	-0.0600	-0.0614
	(0.0040)	(0.0052)	(0.0064)	(0.0049)	(0.0054)	(0.0031)
alpha_cr	-0.0401	-0.0289	-0.0392	-0.2473	-0.2009	-0.0866
	(0.0048)	(0.0016)	(0.0029)	(0.0096)	(0.0238)	(0.0063)
delta	0.6483	0.6687	0.6784	0.4169	0.5754	0.3128
	(0.0171)	(0.0113)	(0.0118)	(0.0066)	(0.0274)	(0.0045)
K-test	12.0666	6.0698	10.3123	20.0642	18.4197	18.7334
p-value	(0.3586)	(0.7329)	(0.4135)	(0.1695)	(0.0723)	(0.2261)
Kc-test						
(Ho: lambda = 0)	7.9976	13.9944	9.7519		0.3137	
p-value	(0.0917)	(0.0297)	(0.0826)		(0.9889)	
Kc-test on						
(Ho: lambda_EDN = 0)		5.9967				
p-value		(0.0499)				
Kc-test						
(Ho: lambda_NCA = 0)			1.7542			
p-value			(0.1853)			

Standard errors in parenthesis.