Women and the Econometrics of Family Trees

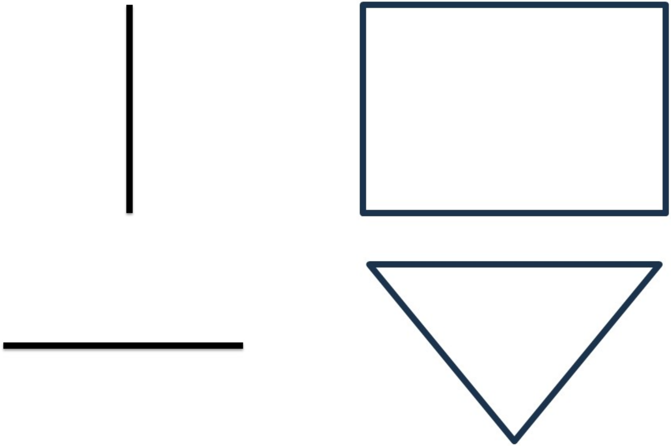
José-Antonio Espín-Sánchez[1] Joseph Ferrie[2] Chris Vickers[3]

[1]Department of Economics
Yale University

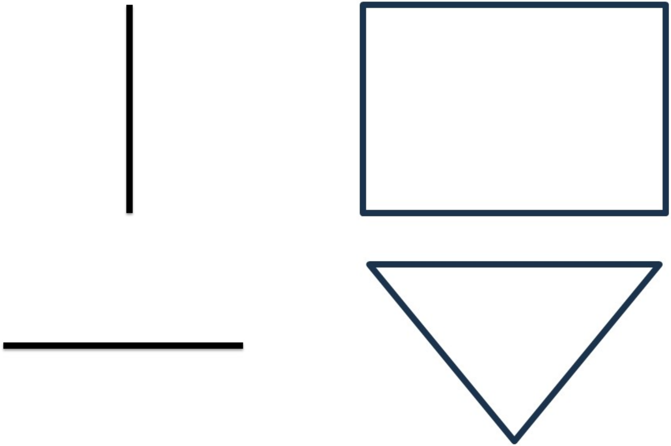
[2]Department of Economics
Northwestern University and NBER

[3]Department of Economics
Auburn University

Which one is different?



Which one is different?



- *Hint*: Econometric Structures

Women and Social Mobility

- Econometrics
  - No Econometrics of Family Trees
- Why the lack of studies on Women and Social Mobility?
  - Cholli and Durlauf (2021) (NBER): 0 times
  - Deutscher and Mazumber (2023) (JEL): 1 time
    - ★ Chadwick and Solon (2002), Olivetti and Paserman (2015), Jácome et al (2021), Craig et al. (2023): daughters not mothers
- Data
  - Most sources only have status information for males
  - Hard to link matrilineally
    - ★ Women change names upon marriage

Solution

- Use women's birth names and create full Family Trees
- Family Trees: income from males in the matrilineal side
  - Maternal Grandfathers
  - Maternal Uncles
- Fully specify an econometric model of Family Trees
  - Structural parameters: Mobility and Assortment
  - Nuisance parameters: Correlation among all Grandparents
  - Use Grandfathers and Uncles as instruments

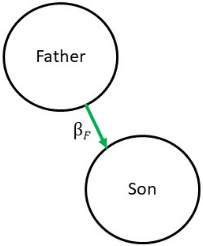
OLS and IV as GMM

- OLS as GMM
  - $Y_i = \beta X_i + \varepsilon_i$
  - $\mathbb{E}[X_i Y_i] = \beta \mathbb{E}[X_i X_i] + \mathbb{E}[X_i \varepsilon_i]$
  - If $\mathbb{E}[X_i \varepsilon_i] = 0$, then $\beta_{OLS} = \frac{\mathbb{E}[X_i Y_i]}{\mathbb{E}[X_i X_i]}$

- IV as GMM
  - $Y_i = \beta X_i + \nu_i$
  - $\mathbb{E}[Z_i Y_i] = \beta \mathbb{E}[Z_i X_i] + \mathbb{E}[Z_i \nu_i]$
  - If $\mathbb{E}[X_i \nu_i] \neq 0$, but $\mathbb{E}[Z_i \nu_i] = 0$, then $\beta_{IV} = \frac{\mathbb{E}[Z_i Y_i]}{\mathbb{E}[Z_i X_i]}$

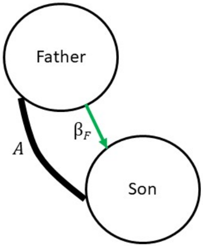
Standard Approach

$$S_i = \tilde{\beta}_F F_i + \nu_i$$



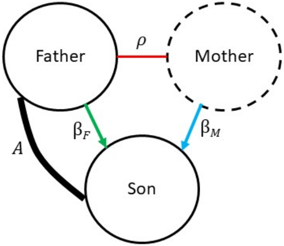
GMM

$$S_i = \tilde{\beta}_F F_i + \nu_i$$
$$\mathbb{E}[F_i S_i] = \tilde{\beta}_F \mathbb{E}[F_i F_i]$$
$$A \equiv \mathbb{E}[F_i S_i] = \tilde{\beta}_F$$



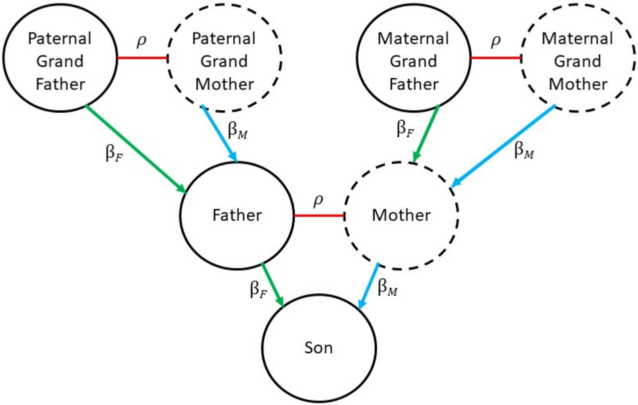
Women Matter

$$S_i = \beta_F F_i + \beta_M M_i + \varepsilon_i^S$$
$$\mathbb{E}[F_i S_i] = \beta_F \mathbb{E}[F_i F_i] + \mathbb{E}[F_i M_i]\beta_M + \varepsilon_i^S$$
$$\mathbb{E}[F_i S_i] = \beta_F + \rho\beta_M$$



Full Trees

$$S_i = \beta_F F_i + \beta_M M_i + \varepsilon_i^S$$
$$F_i = \beta_F PGF_i + \beta_M PGM_i + \varepsilon_i^F$$
$$M_i = \beta_F MGF_i + \beta_M MGM_i + \varepsilon_i^M$$



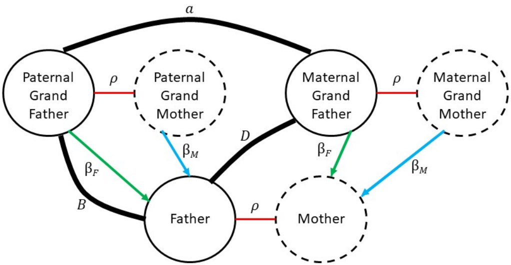
Two Generations Data

$$F_i = \beta_F PGF_i + \beta_M PGM_i + \varepsilon_i^F$$
$$M_i = \beta_F MGF_i + \beta_M MGM_i + \varepsilon_i^M$$

$$a \equiv \mathbb{E}[PGF_i MGF_i]$$

$$B \equiv \mathbb{E}[PGF_i F_i]$$

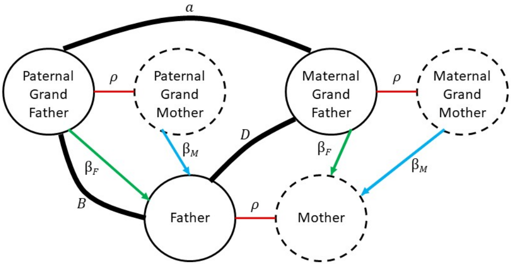
$$D \equiv \mathbb{E}[MGF_i F_i]$$



First Moment

$$F_i = \beta_F PGF_i + \beta_M PGM_i + \varepsilon_i^F$$
$$\mathbb{E}[F_i PGF_i] = \beta_F \mathbb{E}[PGF_i PGF_i] + \beta_M \mathbb{E}[PGM_i PGF_i]$$

$$B = \beta_F + \rho \beta_M$$

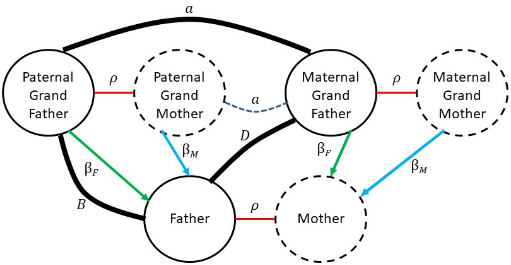


Second Moment

$$F_i = \beta_F PGF_i + \beta_M PGM_i + \varepsilon_i^F$$
$$\mathbb{E}[F_i MGF_i] = \beta_F \mathbb{E}[PGF_i MGF_i] + \beta_M \mathbb{E}[PGM_i MGF_i]$$

$$B = \beta_F + \rho \beta_M$$

$$D = a\beta_F + a\beta_M$$



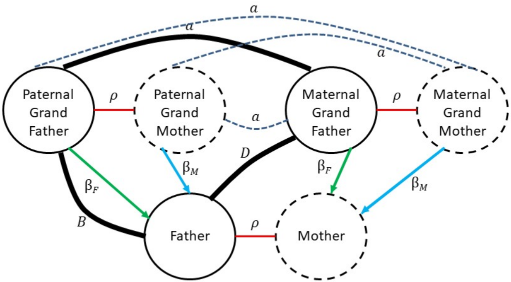
Third Moment

$$F_i = \beta_F PGF_i + \beta_M PGM_i + \varepsilon_i^F$$
$$M_i = \beta_F MGF_i + \beta_M MGM_i + \varepsilon_i^M$$
$$\mathbb{E}[F_i M_i] = \mathbb{E}[(\beta_F PGF_i + \beta_M PGM_i)(\beta_F MGF_i + \beta_M MGM_i)]$$

$$B = \beta_F + \rho\beta_M$$

$$D = (\beta_F + \beta_M)\, a$$

$$\rho = (\beta_F + \beta_M)^2\, a$$



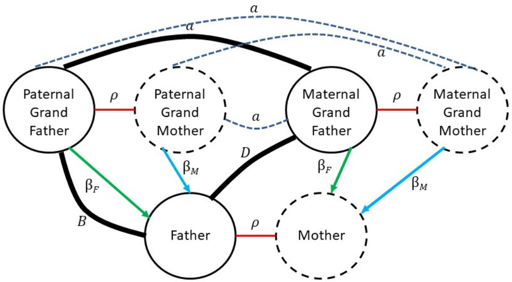
Solving the system

$$B = \beta_F + \rho\beta_M$$
$$D = (\beta_F + \beta_M)\, a$$
$$\rho = (\beta_F + \beta_M)^2\, a$$

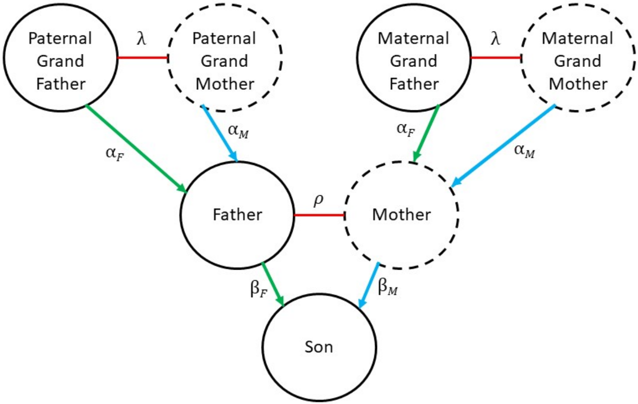
$$\rho = \frac{D^2}{a}$$

$$\beta_F = \frac{D}{a} + \frac{aB - D}{a - D^2}$$

$$\beta_M = -\frac{aB - D}{a - D^2}$$

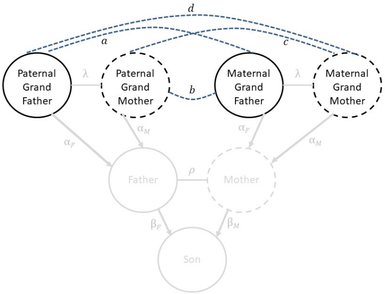


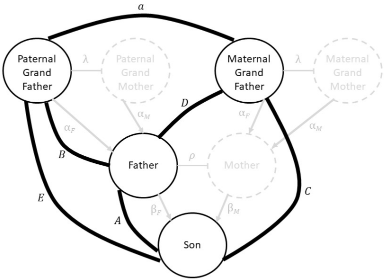
General Tree



Family Trees

Nuisance Parameters and Empirical Relations





Main Results

Table 1: Summary of Main Identification results.

Prop.	Data	Nuisance Assumptions	Structural Assumptions	Point Identified Parameters
Identification using two generations				
Prop. 1	(F, PGF, MGF)	$a = b = c = d$	$\lambda = \rho$	(β_F, β_M, ρ)
Prop. 2	(F, PGF, MGF)		$\beta_M = 0$	(β_F, ρ)
Prop. 3	(F, PGF, MGF)	$a = b = c = d$	$\beta_F = \beta_M$	(β_F, λ, ρ)
Identification using three generations				
Prop. 4	(S, F, PGF, MGF)	$b = d$		$(\beta_F, \beta_M, \lambda, \rho)$
Identification from maternal uncles				
Prop. 5	(S, F, MGF)	$a = b = c = d$	$\lambda = \rho$	$(\beta_F, \beta_M, \rho, a)$
Prop. 6	(S, F, MU, MGF)			$(\beta_F, \beta_M, \lambda, \rho, \gamma)$
Prop. 7	(S, F, MU, PGF)			$(\beta_F, \beta_M, \lambda, \rho, \gamma)$
Prop. 8	(S, F, MU)	$\gamma = 0$	$\lambda = \rho$	(β_F, β_M, ρ)

Extended Results

Table 2: Summary of Extended Identification results.

Prop.	Data	Nuisance Assumptions	Structural Assumptions	Point Identified Parameters
Identification allowing heterogeneous effects by gender				
Prop. 9	(S, F, PGF, MGF)	$a = b = c = d$		$(\beta_F^S, \beta_M^S, \beta_F^D, \beta_M^D, \lambda, \rho)$
Prop. 10	(S, F, PGF, MGF)	$b = d = 0;\ a = c$		$(\beta_F^S, \beta_M^S, \beta_F^D, \beta_M^D, \lambda, \rho)$
Identification allowing heterogeneous effects by generation				
Prop. 11	(S, F, PGF, MGF)	$b = d$		$(\beta_F, \beta_M, \rho, \alpha)$
Cor. 1	(S, F, PGF, MGF)	$a = b = d$	$\alpha_F = \alpha_M$	$(\beta_F, \beta_M, \alpha_F, \lambda, \rho)$
Prop. 12	(S, F, PGF, MGF)	$b = d = 0;\ a = c$		$(\beta_F, \beta_M, \alpha_F, \alpha_M, \lambda, \rho)$
Prop. 13	(S, F, PGF, MGF)	$a = b = d$	$\lambda = \rho$	$(\beta_F, \beta_M, \alpha_F, \alpha_M, \rho)$
Prop. 14	(S, F, PGF, MGF)	$b = d = \sqrt{ac}$		$(\beta_F, \beta_M, \rho, \tilde{\alpha})$
Cor. 2	(S, F, PGF, MGF)	$b = d = \sqrt{ac}$	$\alpha_F = \alpha_M$	$(\beta_F, \beta_M, \rho, \tilde{\alpha})$

Finding Women's pre-marriage names

Using the NUMIDENT

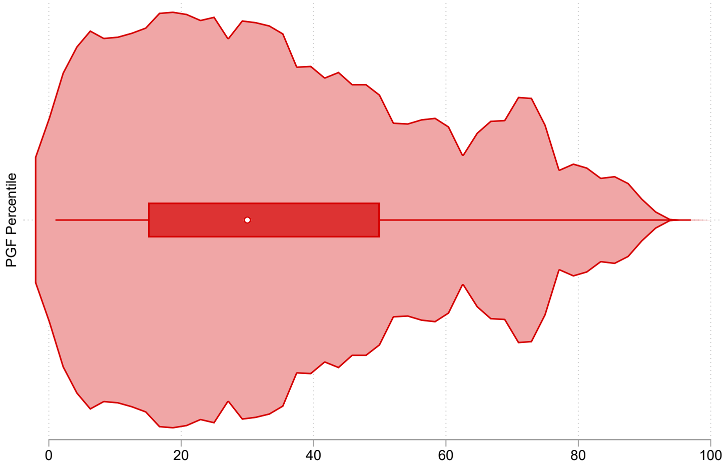
- Standard Linking across three censuses
  - Grandparents (1900)
  - Parents (1920)
  - Child (1940)
- To come: Buckles, Haws, Price and Wilbert (2023)
- Social Security NUMIDENT (Numerical Identification) file
  - Individuals dead (or over 110 years old) by December 31, 2007
  - Includes mother's pre-marriage surname
  - Collected when entering employment (farmers are excluded)

Measuring Socioeconomic Status

Beyond OCCSCORES

- Using OCCSCORES
  - Income information only available after 1950
  - Occupation available 1900-1940
  - OCCSCORES: median income for an occupation in 1950
- Problems with OCCSCORES
  - No variation over space
  - No variation over time
  - Reversal of fortunes for farmers
    - 60% of males are farmers in 1900
    - 64% of them have a son that is a farmer in 1920
- Solution
  - Add variation over space and time
  - New estimates of farmer's income by State-decade

Farmers Income in 1900



Chadwick and Solon (2002)

- Clever way to estimate ρ using (F_i, PGF_i, MGF_i). Assumptions
  - Mother's income relates to Father's income: $F_i = \rho_0 M_i + \nu_i$
  - Equation for Father's income is then: $F_i = \beta_F PGF_i + \beta_M PGM_i + \rho_0 M_i + \varepsilon_i$
  - Moment: No relation among grandparents: $D = \beta_F a + \beta_M b + \rho_0 B$
- Elegant solution (if $a = b = 0$)
  - $\rho_0 = D/B$
- Why not $M_i = \rho_1 F_i + \nu_i'$?
- Summary
  - CS: $F_i = \rho_0 M_i + \varepsilon_i^0$, the estimator is $\rho_0 = D/B$.
  - Reversed CS: $M_i = \rho_1 F_i + \varepsilon_i^1$, the estimator is $\rho_1 = B/D$.
  - Correlational (Prop. 2): $\mathbb{E}[F_i M_i] = \rho$ and $\beta_M = 0$, the estimator is $\rho = BD$.

Chadwick and Solon (2002)

- Clever way to estimate ρ using (F_i, PGF_i, MGF_i). Assumptions
  - Mother's income relates to Father's income: $F_i = \rho_0 M_i + \nu_i$
  - Equation for Father's income is then: $F_i = \beta_F PGF_i + \beta_M PGM_i + \rho_0 M_i + \varepsilon_i$
  - Moment: No relation among grandparents: $D = \beta_F a + \beta_M b + \rho_0 B$
- Elegant solution (if $a = b = 0$)
  - $\rho_0 = D/B$
- Why not $M_i = \rho_1 F_i + \nu_i'$?
- Summary
  - CS: $F_i = \rho_0 M_i + \varepsilon_i^0$, the estimator is $\rho_0 = D/B$.
  - Reversed CS: $M_i = \rho_1 F_i + \varepsilon_i^1$, the estimator is $\rho_1 = B/D$.
  - Correlational (Prop. 2): $\mathbb{E}[F_i M_i] = \rho$ and $\beta_M = 0$, the estimator is $\rho = BD$.

Chadwick and Solon (2002)

- Clever way to estimate ρ using (F_i, PGF_i, MGF_i). Assumptions
  - Mother's income relates to Father's income: $F_i = \rho_0 M_i + \nu_i$
  - Equation for Father's income is then: $F_i = \beta_F PGF_i + \beta_M PGM_i + \rho_0 M_i + \varepsilon_i$
  - Moment: No relation among grandparents: $D = \beta_F a + \beta_M b + \rho_0 B$
- Elegant solution (if $a = b = 0$)
  - $\rho_0 = D/B$
- Why not $M_i = \rho_1 F_i + \nu_i'$?
- Summary
  - CS: $F_i = \rho_0 M_i + \varepsilon_i^0$, the estimator is $\rho_0 = D/B$.
  - Reversed CS: $M_i = \rho_1 F_i + \varepsilon_i^1$, the estimator is $\rho_1 = B/D$.
  - Correlational (Prop. 2): $\mathbb{E}[F_i M_i] = \rho$ and $\beta_M = 0$, the estimator is $\rho = BD$.

Main Results

Table 1: Summary of Main Identification results.

Prop.	Data	Nuisance Assumptions	Structural Assumptions	Point Identified Parameters
Identification using two generations				
Prop. 1	(F, PGF, MGF)	$a = b = c = d$	$\lambda = \rho$	(β_F, β_M, ρ)
Prop. 2	(F, PGF, MGF)		$\beta_M = 0$	(β_F, ρ)
Prop. 3	(F, PGF, MGF)	$a = b = c = d$	$\beta_F = \beta_M$	(β_F, λ, ρ)
Identification using three generations				
Prop. 4	(S, F, PGF, MGF)	$b = d$		$(\beta_F, \beta_M, \lambda, \rho)$
Identification from maternal uncles				
Prop. 5	(S, F, MGF)	$a = b = c = d$	$\lambda = \rho$	$(\beta_F, \beta_M, \rho, a)$
Prop. 6	(S, F, MU, MGF)			$(\beta_F, \beta_M, \lambda, \rho, \gamma)$
Prop. 7	(S, F, MU, PGF)			$(\beta_F, \beta_M, \lambda, \rho, \gamma)$
Prop. 8	(S, F, MU)	$\gamma = 0$	$\lambda = \rho$	(β_F, β_M, ρ)

Two Generations Empirical Results

Table 3: Identification using Two Generations

Parameter	Estimate		
	Prop. 1	Prop. 2	Prop. 3
β_F	0.167	0.930	0.465
	(0.025)	(0.015)	(0.007)
β_M	0.763		
	(0.039)		
ρ	0.416	0.416	0.416
	(0.012)	(0.012)	(0.007)

Model using Generational Effects

Allowing for gendered effects the model becomes

$$S_i = \beta_F F_i + \beta_M M_i + \varepsilon_i^S$$

$$F_i = \alpha_F PGF_i + \alpha_M PGM_i + \varepsilon_i^F$$

$$M_i = \alpha_F MGF_i + \alpha_M MGM_i + \varepsilon_i^M$$

where

- β_F is the effect of the father on a child in the second generation
- β_M is the effect of the mother on a child in the second generation
- α_F is the effect of the father on a child in the first generation
- α_M is the effect of the mother on a child in the first generation

Extended Results

Table 2: Summary of Extended Identification results.

Prop.	Data	Nuisance Assumptions	Structural Assumptions	Point Identified Parameters
Identification allowing heterogeneous effects by gender				
Prop. 9	(S, F, PGF, MGF)	$a = b = c = d$		$(\beta_F^S, \beta_M^S, \beta_F^D, \beta_M^D, \lambda, \rho)$
Prop. 10	(S, F, PGF, MGF)	$b = d = 0;\ a = c$		$(\beta_F^S, \beta_M^S, \beta_F^D, \beta_M^D, \lambda, \rho)$
Identification allowing heterogeneous effects by generation				
Prop. 11	(S, F, PGF, MGF)	$b = d$		$(\beta_F, \beta_M, \rho, \alpha)$
Cor. 1	(S, F, PGF, MGF)	$a = b = d$	$\alpha_F = \alpha_M$	$(\beta_F, \beta_M, \alpha_F, \lambda, \rho)$
Prop. 12	(S, F, PGF, MGF)	$b = d = 0;\ a = c$		$(\beta_F, \beta_M, \alpha_F, \alpha_M, \lambda, \rho)$
Prop. 13	(S, F, PGF, MGF)	$a = b = d$	$\lambda = \rho$	$(\beta_F, \beta_M, \alpha_F, \alpha_M, \rho)$
Prop. 14	(S, F, PGF, MGF)	$b = d = \sqrt{ac}$		$(\beta_F, \beta_M, \rho, \tilde{\alpha})$
Cor. 2	(S, F, PGF, MGF)	$b = d = \sqrt{ac}$	$\alpha_F = \alpha_M$	$(\beta_F, \beta_M, \rho, \tilde{\alpha})$

Generational Effects Empirical Results

Table 4: Identification Allowing Heterogeneous Effects by Generation

Parameter	Estimate			
	Prop. 11	Prop. 12	Prop. 13	Prop. 14
α	0.465 (0.015)			
α_F		0.568 (0.013)	0.138 (0.295)	
α_M		0.098 (1.094)	0.792 (0.299)	
$\tilde{\alpha}$				0.465 (0.007)
β_F	0.272 (0.080)	0.346 (0.022)	0.272 (0.080)	0.272 (0.080)
β_M	0.220 (0.081)	0.169 (0.024)	0.220 (0.081)	0.220 (0.081)
λ		0.010 (0.125)		0.041 (0.021)
ρ	0.437 (0.207)	0.157 (0.105)	0.437 (0.207)	0.437 (0.207)

Generational Effects Empirical Results

Table 4: Identification Allowing Heterogeneous Effects by Generation

Parameter	Estimate			
	Prop. 11	Prop. 12	Prop. 13	Prop. 14
α	0.465 (0.015)			
α_F		0.568 (0.013)	0.138 (0.295)	
α_M		0.098 (1.094)	0.792 (0.299)	
$\tilde{\alpha}$				0.465 (0.007)
β_F	0.272 (0.080)	0.346 (0.022)	0.272 (0.080)	0.272 (0.080)
β_M	0.220 (0.081)	0.169 (0.024)	0.220 (0.081)	0.220 (0.081)
λ		0.010 (0.125)		0.041 (0.021)
ρ	0.437 (0.207)	0.157 (0.105)	0.437 (0.207)	0.437 (0.207)

Generational Effects Empirical Results

Table 4: Identification Allowing Heterogeneous Effects by Generation

Parameter	Estimate			
	Prop. 11	Prop. 12	Prop. 13	Prop. 14
α	0.465 (0.015)			
α_F		0.568 (0.013)	0.138 (0.295)	
α_M		0.098 (1.094)	0.792 (0.299)	
$\tilde{\alpha}$				0.465 (0.007)
β_F	0.272 (0.080)	0.346 (0.022)	0.272 (0.080)	0.272 (0.080)
β_M	0.220 (0.081)	0.169 (0.024)	0.220 (0.081)	0.220 (0.081)
λ		0.010 (0.125)		0.041 (0.021)
ρ	0.437 (0.207)	0.157 (0.105)	0.437 (0.207)	0.437 (0.207)

Conclusions

- Mobility estimates for Mothers are large
- Assortative Mating is high
  - Rethinking the implication of mobility estimates
  - Interaction with mobility estimates
- Grandparents may affect mobility via Assortative Mating
- Extensions
  - Direct Grandparents effects
  - Mating on unobservables
  - Estimates of nuisance parameters with data on other relatives