Correlations of Brothers' Earnings and Intergenerational Transmission

Paul Bingley
SFI - The Danish National Centre for Social Research

Lorenzo Cappellari
Università Cattolica Milano

Winter School on Inequality and Social Welfare Theory
Canazei, 15 January 2014

Background

- Research on income inequality and the family of origin has focussed on two dimensions (among others) of such dependence
  1. **Intergenerational**: parent-child transmission. Elasticity (IGE).
  2. **Sibling**: omnibus measure of intergenerational plus any other shared influence (e.g. schools, friends, neighbours). Correlation.

- What is the relative importance of parents' earnings vs. other factors in determining the overall sibling correlation?
- (A lower bound for the share of inter- within intra-generational inequality).

Contribution

- Existing evidence based on models with homogeneous IGE suggests a limited impact.
- We develop a model of life-cycle earnings for siblings and their fathers allowing for heterogeneity of intergenerational transmission across families.
- We find that the intergenerational correlation accounts for almost all of sibling similarities.

What we do

- Use Danish population data on annual earnings of men grouped in Father-Son1-Son2 triplets.
- Develop a multi-person model of earnings dynamics that
  - Distinguishes permanent earnings from transitory shocks.
  - Allows for life-cycle effects in both.
  - Distinguishes individual-specific effects from sibling-specific effects within permanent earnings.
  - **Decomposes sibling effects into intergenerational and residual sibling effects.**

Findings

Core results

- Intergenerational is *most* of sibling correlation: 50-95% depending on age (Previous DK estimate = 6% using decomposition method with homogeneous IGE).

- Sibling correlation u-shaped in age: 0.5 at 25, 0.15 at 37, 0.2 at 45, 0.23 on average (Previous DK estimate = 0.23 without age effects, for brothers aged 25-42).

Moreover

- Cross-person correlation in transitory earnings : Significant but small.
- Differential transmission by birth order: Mild evidence of larger correlation with later born.

Outline

1. Literature on sibling correlations

2. Data

3. Sibling correlations and IGE heterogeneity

4. Earnings dynamics and estimating issues

5. Model

6. Results

Sibling correlations of incomes

- Omnibus measure of family and community effects (Corcoran et al., 1976; Solon et al., 1991; Altonji and Dunn, 1991).

- Models of sibling effects in permanent incomes:

$$y_{ij} = a_{ij} + f_j \,,\; a_{ij} \sim \big(0, var(a)\big);\; f_j \sim \big(0, var(f)\big)$$
$$r^S = \frac{var(f)}{var(a) + var(f)}$$

- Share of inequality in permanent incomes accounted for by factors shared by siblings (loosely speaking inequality 'between families').
- r^S=0.35 - 0.40 in the US (Solon et al. 1991; Mazumder, 2008), 0.35 in SWE (Björklund et al., 2009), 0.23 in DK (Björklund and Jannti, 2012), 0.43 in GER (Schnitzlein, 2014).

Sibling correlations and IGE (1)

- Solon (1999) allows the shared component to depend on father's permanent earnings (y_j^F) and an orthogonal residual component (ξ_j):

$$f_j = \eta y_j^F + \xi_j \;\; ; \xi_j \sim (0, var(\xi))$$

- Assuming stationarity across generations he provides the analytical link with the intergenerational elasticity (IGE):

$$r^S = IGE^2 + residual\ correlation$$

- Using the formula in calibration he shows that in the US 40% of the correlation accounted for by parental earnings.

- Björklund and Jannti (2009) apply the decomposition to Danish data (among others) and find that IGE accounts for 6% of sibling correlation.

Sibling correlations and IGE (2)

- Mazumder (2008) uses REML to estimate correlations before and after conditioning on observables, parental income accounts for 40%.

- If father's income is the only regressor, the method is equivalent to Solon's decomposition, but without assuming intergenerational stationarity of the earnings distribution.

- Björklund et al. (2010) use a similar methodology finding that parental income accounts for 13% of the sibling correlation in Sweden.

Sibling correlations and other shared influences

- Limited effects of IGE suggest that there must be some other factors at play
- Page and Solon (2003 a,b) compare sibs with neighboring boys and girls but find only small role of neighbours.
- Björklund et al. (2010) find that parental attitudes seem to matter, less so the structure of the family.
- Bingley et al. (2014) find that schools and neighbours matter little and mostly before age 30.

Outline

1. Literature on sibling correlations
2. **Data**
3. Sibling correlations and IGE heterogeneity
4. Earnings dynamics and estimating issues
5. Model
6. Results

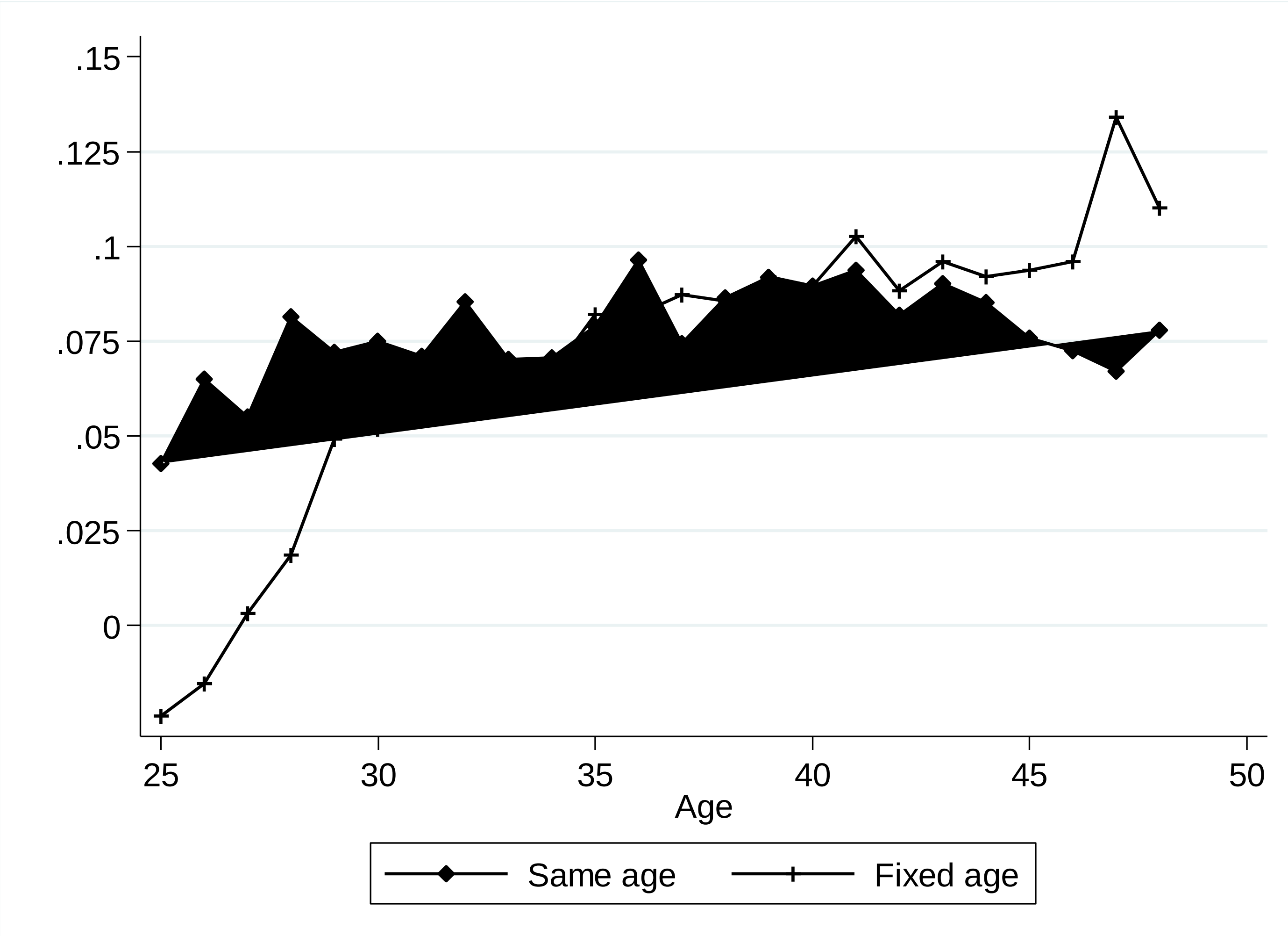
Data – origin & construction

- Danish administrative registers 1980 – 2011
- Gross annual labour income
- Fathers born 1935 – 1964 (aged 25-60)
- Sons born 1959 – 1982 (aged 25-51)
- Registered parents at birth – drop adoptions
- Full fatherhood history (our first son is his first son)
- Full biological brothers
- Age spacing 1-12 – drop twins
- Also use families without a second son
- 5+ years continuous earnings, otherwise missing at random
- drop top & bottom 0.5% earnings by year & person type
- 740k persons, 326k families, 88k triplets, 12m obs.
- Group into 3-year birth cohorts

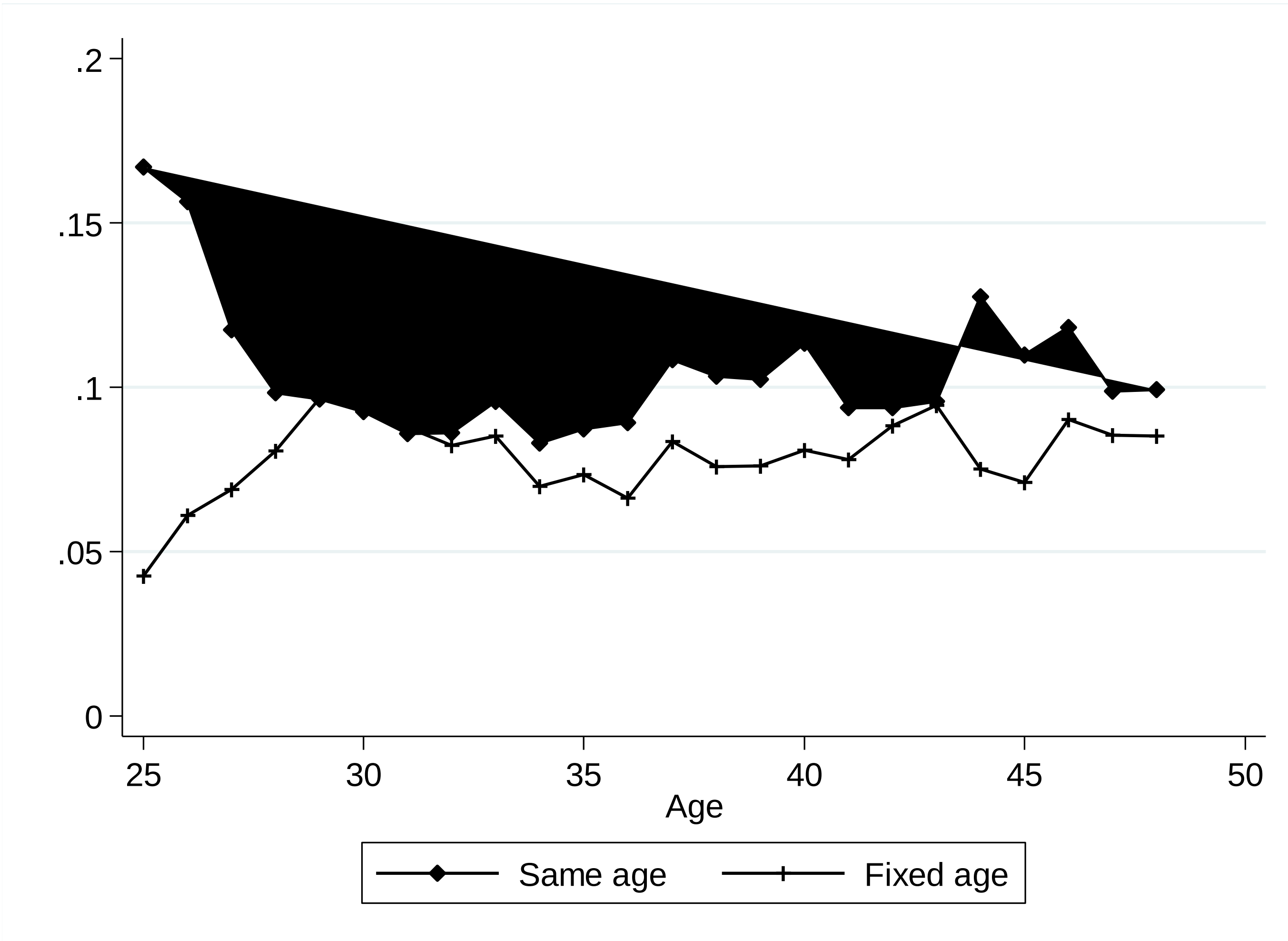
Raw log earnings correlations by age

- Residuals from regressing log real annual earnings on age, age^2 & year dummies by birth cohort (3-year) group
  - Drop small cells throughout (based on <100 cases)
- Father-son
  - Contrast same age with fixed father age (40)
  - Comparing F-S at different points in their life-cycle
- Brother-brother
  - Contrast same age with fixed older brother age (30)
  - Comparing S1-S2 at different points in their life-cycle

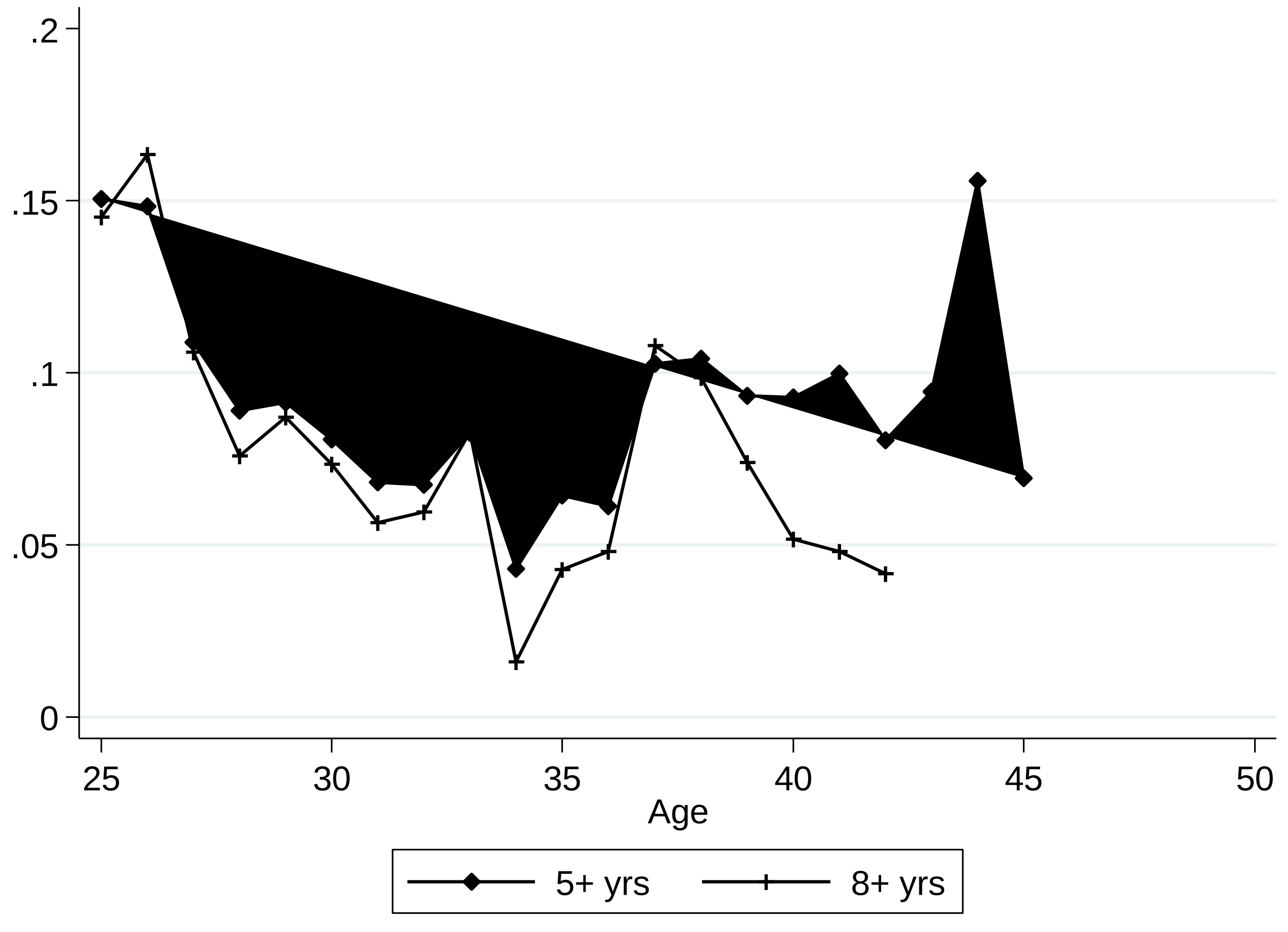
Intergenerational



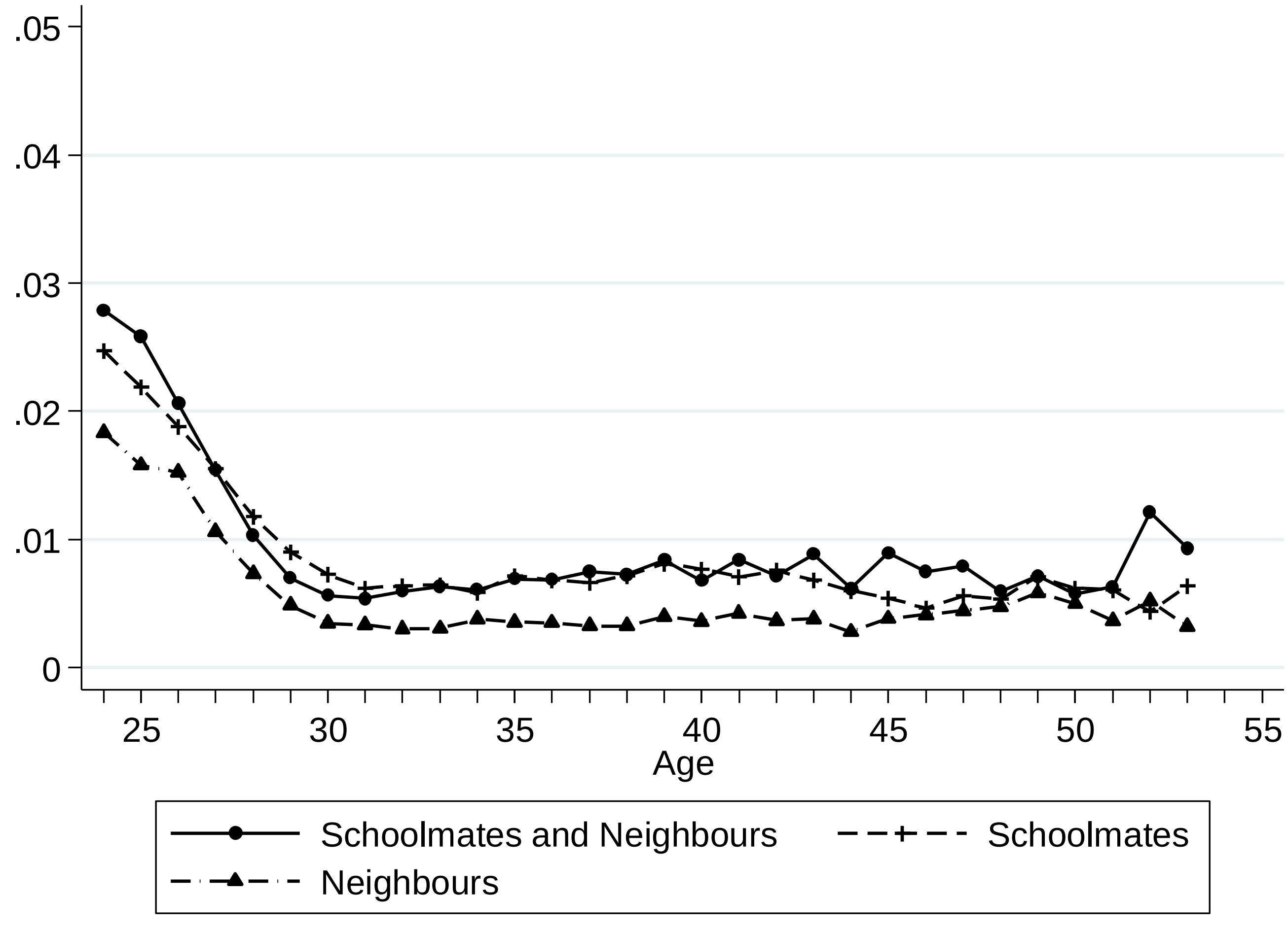
Sibling



Siblings born 5 or 8 yrs apart

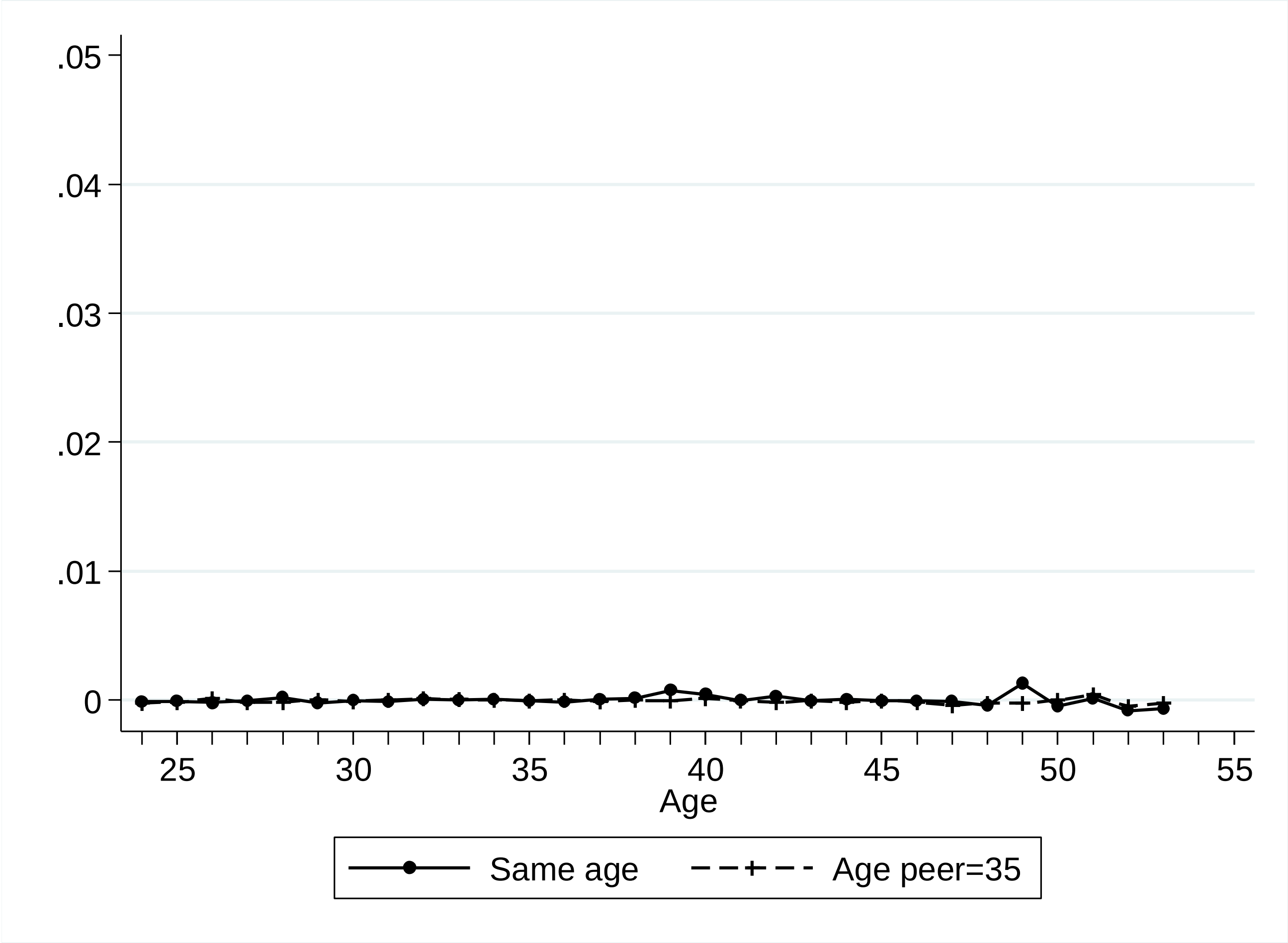


Neighbours & Schoolmates



Source: Bingley et al (2014)

Unrelated peers



Source: Bingley et al (2014)

Outline

1. Literature on sibling correlations
2. Data
3. **Sibling correlations and IGE heterogeneity**
4. Earnings dynamics and estimating issues
5. Model
6. Results

What happens if allow IGE to differ across families?

- To capture the idea that the intensity of intergenerational links varies across the distribution. (Stronger at the top? ..but assume independence)

$$f_j = \eta_j y_j^F + \bar{\eta} y_j^F + \xi_j;$$
$$\eta_j \sim (0, var(\eta)), cov(\eta_j y_j^F) = 0$$

- Then we get an extra term (>0) in the decomposition of the sibling correlation:

$$r^S = var(IGE) + IGE^2 + residual\ correlation$$

	Coeff.	S.E.	%
Decompositions with homogeneous IGE			
Solon (1999) decomposition			
$var(a)$	0.2358	0.0010	
$var(f)$	0.0550	0.0010	
IGE	0.0757	0.0015	
r^S	0.1892	0.0034	
Share of r^S explained by y_j^F			3.02
Sequential conditioning			
$var(a)$ after conditioning on y_j^F	0.2359	0.0010	
$var(f)$ after conditioning on y_j^F	0.0527	0.0010	
r^S after conditioning on y_j^F	0.1828	0.0034	
Share of r^S explained by y_j^F			3.36

Decompositions with heterogeneous IGE	Coeff.	S.E.	%
$var(a)$	0.2354	0.0010	
IGE	0.0912	0.0016	
$var(IGE)$	0.0307	0.0011	
$var(\xi)$	0.0422	0.0010	
r^S	0.1953	0.0034	
Share of r^S explained by y_j^F			
Assuming stationarity			20.00
Without assuming stationarity ($var(y_j^F)$=0.3824)			26.15

Outline

1. Literature on sibling correlations
2. Data
3. Sibling correlations and IGE heterogeneity
4. **Earnings dynamics and estimating issues**
5. Model
6. Results

Estimation issues (1): Transitory shocks

- Downward bias (Solon, 1992, and Zimmerman, 1992). Argue in favour of multiperiod averages of earnings to integrate out transitory shocks.

- Serially correlated shocks may resist multiperiod averaging on short time windows (Mazumder, 2005).

- Solutions: average over long strings of earnings or model shock correlation (Björklund et al. 2009).

Estimation issues (2): Life cycle bias

- We may still miss permanent income if data are sampled in the “wrong” phase of the life-cycle (Jenkins, 1987).

- Haider and Solon (2006): Heterogeneous income growth causes life-cycle bias
  - Show how the bias varies over age
  - Find that for men bias is minimised between ages of 30 and 40

- Björklund et al. (2009) use incomes averaged in the 30-40 age range to estimate sibling correlations of permanent incomes with bias minimised.

- Nybom and Stuhler (2013) stress the need of a better assessment of within-family correlation of earnings profile heterogeneity.

Estimation issues (1) vs (2)

- **Key tension:** transitory shocks require long series of individual incomes, life-cycle bias calls for concentrating on ten years.
- Our model allows for serially correlated transitory shocks and within family correlations of individual earnings profiles.

Earnings dynamics

- Lillard and Willis (1978), MaCurdy (1982), Meghir and Pistaferri (2011), Moffitt and Gottschalk (2012).
- Few examples of multi-person modelling (Hyslop, 2001; Ostrowsky, 2012; Blundell et al. 2012). Couples.
- Permanent and transitory components.
- Transitory earnings as ARMA processes.
- Useful in our context due to estimation issue (1).
- Baker and Solon (2003) show transitory shocks u-shaped in age.

Dynamics of permanent earnings

- Allow for life cycle variation.

- Useful in our context due to estimation issue (2).

- Models of permanent earnings:
  - Random Growth (RG, aka HIP)

$$y_{it} = a_i + b_i A_{it}; \quad var(a), var(b), cov(ab)$$

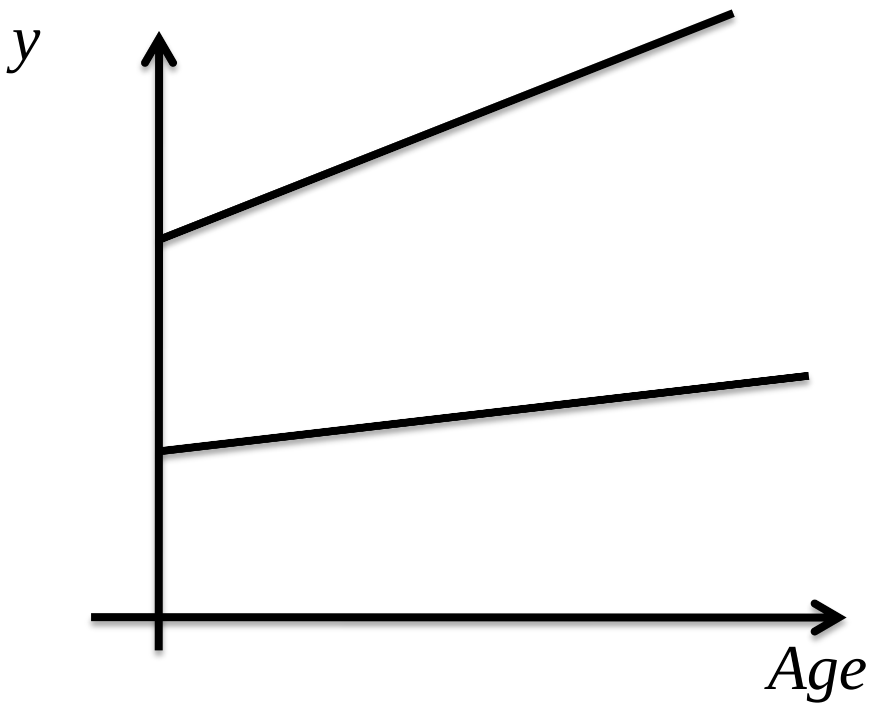
  - Random Walk (RW, aka RIP)

$$y_{it} = y_{it-1} + r_{it}; \quad var\left(y_{it(A_0)}\right), var(r)$$

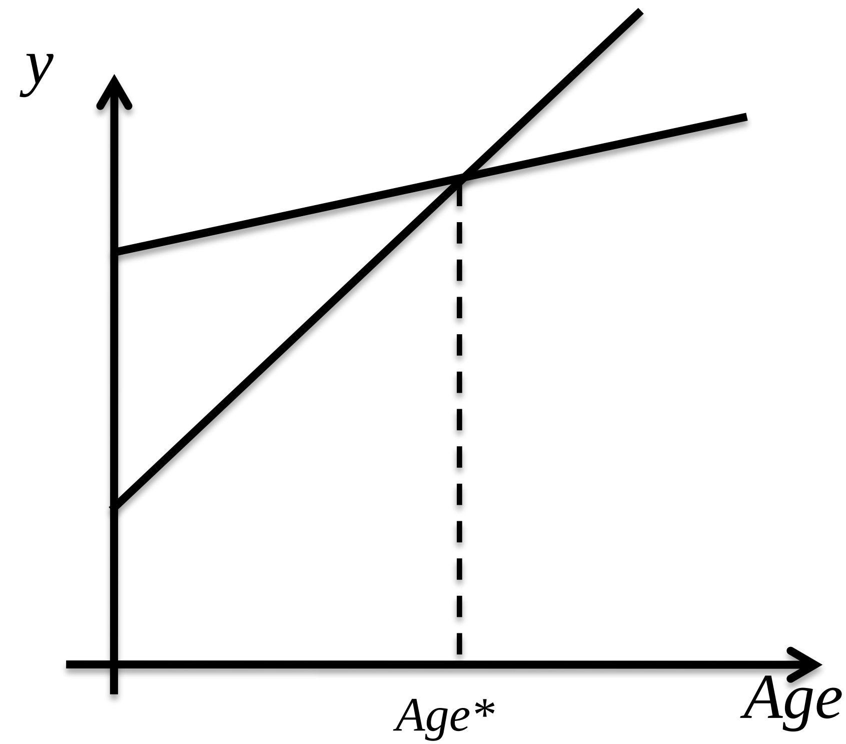
Earnings dynamics – RG & RW

- Random Growth
  - Grounded in Mincerian human capital model.
  - Can capture u-shaped pattern of life cycle variance (recurrent stylised fact; human capital model, Rubinstein and Weiss, 2006).
  - Needs learning foundation in rational expectations settings (Guvenen 2007).
- Random Walk
  - Predicts always-increasing life cycle variance.
  - Fits well in rational expectations models.

Earnings dynamics – RG illustration



Cov(ab)>0



Cov(ab)<0

With Cov(ab)<0: Mincerian cross-overs at Age*. Intragenerational income mobility increases up to Age*, then decreases.

Outline

1. Literature on sibling correlations
2. Data
3. Sibling correlations and IGE heterogeneity
4. Earnings dynamics and estimating issues
5. **Model**
6. Results

Model

- Distinguish transitory from permanent earnings using orthogonal decomposition:

$$w_{ijt}^{h} = y_{ijt}^{h} + v_{ijt}^{h}; E\left(y_{ijt}^{h}, v_{ijt}^{h}\right) = 0, \qquad h = F, S1, S2.$$

- Log-deviations from period- cohort- member-specific means.

- Earnings components orthogonal by definition, but correlated within the family.

- Earnings processes with desirable properties:
  - Life-cycle effects in permanent earnings.
  - Serial correlation in transitory shocks.

Model - overview

- We innovate the canonical sibling model in two key directions:
  1. We split the sibling component into the intergenerational effect and a residual effect (direct decomposition).
  2. We introduce life-cycle effects.

- We achieve #2 using a mixture of RG and RW, plus age-based heteroskedasticity of transitory shocks.

- RG for shared components of permanent earnings:
  – Life-cycle biases.
  – Empirical patterns u-shaped.

- RW for idiosyncratic components of permanent earnings.

Model – sons' permanent earnings

$$y_{ijt}^{h} = \left(\left(\mu_j^I + \mu_j^R\right) + \left(\lambda_j^I + \lambda_j^R\right)A_{it} + \omega_{ijt}^{h}\right)\pi_t$$

$$\omega_{ijt}^{h} = \omega_{ijt-1}^{h} + \phi_{ijt}^{h}$$

$$\left(\omega_{ijt(A_0)}^{h}, \phi_{ijt}^{h}\right) \sim \left(0,0; \sigma_{\omega 0h}^2, \sigma_{\phi h}^2\right),$$

$$(\mu_j^I, \lambda_j^I) \sim \left(0,0;\ \sigma_{\mu I}^2, \sigma_{\lambda I}^2, \sigma_{\mu\lambda I}\right),$$

$$(\mu_j^R, \lambda_j^R) \sim \left(0,0;\ \sigma_{\mu R}^2, \sigma_{\lambda R}^2, \sigma_{\mu\lambda R}\right)$$

Model – fathers' permanent earnings

- Identification of intergenerational component requires father's earnings to be modelled jointly with sons' ones.

$$y_{ijt}^{F} = (\mu_j^I + \lambda_j^I A_{it} + \omega_{ijt}^{F})\pi_t$$

Model – transitory earnings

- Type-specific AR(1) with age-based heteroskedasticity and cross-person correlation of shocks:

$$v_{ijt}^{h} = \tau_t u_{ijt} = \tau_t \left(\rho_h u_{ijt-1} + \varepsilon_{ijt} \right),$$
$$\varepsilon_{ijt} \sim \left(0, \sigma_{\varepsilon hA}^2\right), \sigma_{\varepsilon hA}^2 = \sigma_{\varepsilon h}^2 \exp(g_h(A_{it}))$$
$$u_{ijs} \sim \left(0, \eta_c^{d(s=t_0)} \sigma_{sh}^2\right),$$
$$E\left(\varepsilon_{ijt}\varepsilon_{kjt}\right) = \sigma_{hl}$$

Model - decomposition

- Use parameter estimates to decompose the sibling correlation of permanent earnings over the life-cycle (ρ^S) into its intergenerational (ρ^I) and residual sibling (ρ^R) components:

$$\rho^S(A) = \rho^I(A) + \rho^R(A)$$

Model - estimation

- The model yields restrictions on second moments of the earnings distribution, both between and within persons.

- A non-linear function of the parameters of interest (RG and RW variances and covariances, person-specific AR(1) parameters, period factor loadings on permanent and transitory earnings).

- Match these to empirical earnings moments via GMM (Minimum Distance Estimation).

Model – moment restrictions

1. (Moments decay over lags: permanent vs transitory.)
2. Earnings moments of each brother (idyosincratic + intergenerational + residual sibling).
3. Earnings moments of fathers (idyosincratic + intergenerational).
4. Earnings moments between brothers (intergenerational + residual sibling).
5. Earnings moments between father and sons (intergenerational).
6. $\cong$ 44k moment restrictions.

Outline

1. Literature on sibling correlations
2. Data
3. Sibling correlations and IGE heterogeneity
4. Earnings dynamics and estimating issues
5. Model
6. **Results**

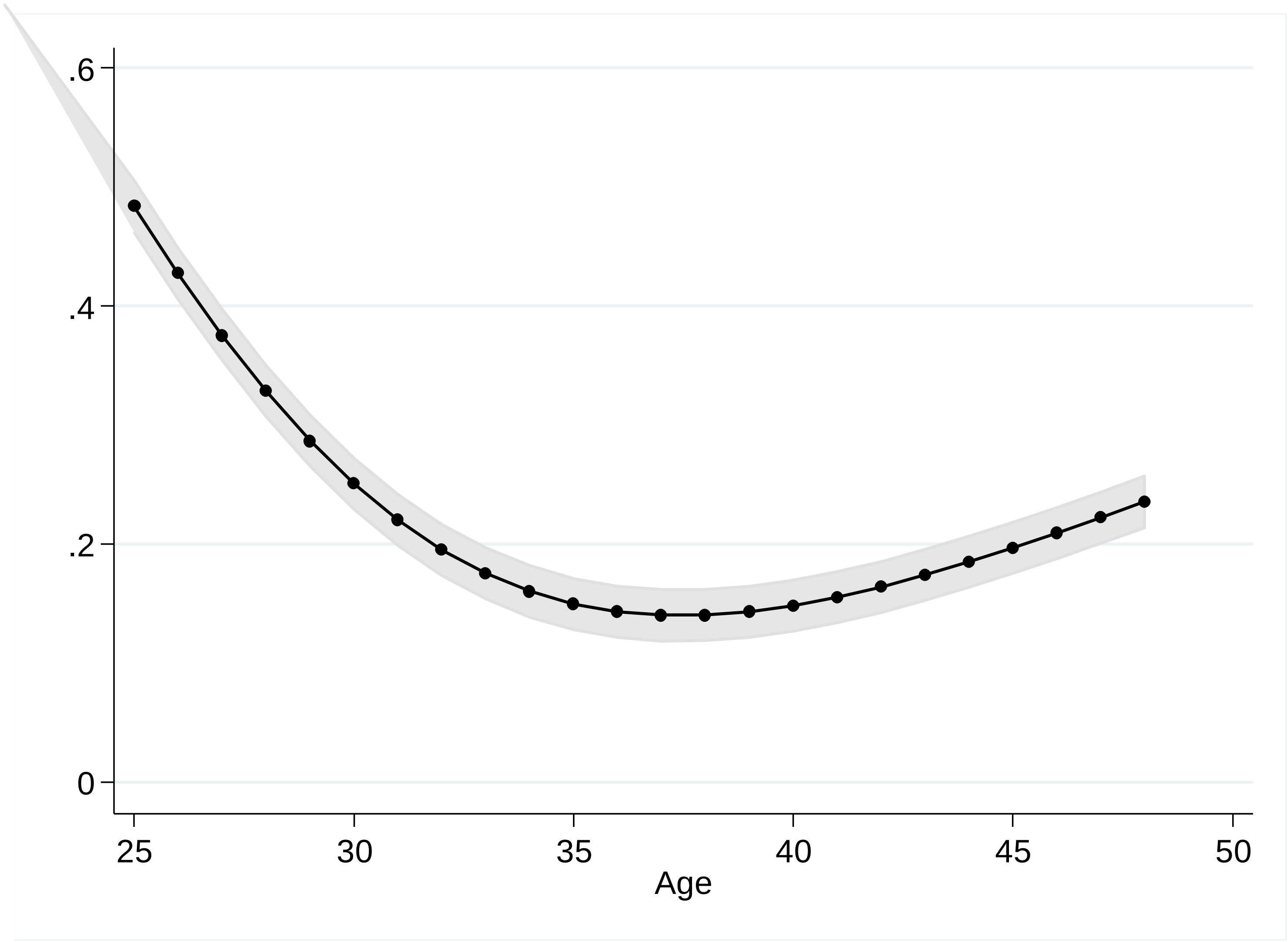
Results

1. Main model
2. Nested models
3. Accounting for sisters
4. Differential IG transmission

Table 2: Estimates of parameters of permanent earnings

	Coeff.	S.E.
Shared components		
Variance of initial earnings		
$\sigma^2_{\mu I}$ (Intergenerational)	0.0339	0.0015
$\sigma^2_{\mu R}$ (Residual Sibling)	0.0243	0.0029
Variance of earnings growth rates		
$\sigma^2_{\gamma I}$ (Intergenerational)	0.0002	0.00001
$\sigma^2_{\gamma R}$ (Residual Sibling)	0.0002	0.00001
Covariance		
$\sigma_{\mu\gamma I}$ (Intergenerational)	-0.0014	0.0001
$\sigma_{\mu\gamma R}$ (Residual Sibling)	-0.0018	0.0002
Idiosyncratic component		
Variance of initial earnings		
$\sigma^2_{\omega 0F}$ (Father)	0.0697	0.0043
$\sigma^2_{\omega 0S1}$ (Son 1)	0.0711	0.0051
$\sigma^2_{\omega 0S2}$ (Son 2)	0.0531	0.0048
Variance of shocks		
$\sigma^2_{\phi F}$ (Father)	0.0021	0.0006
$\sigma^2_{\phi S1}$ (Son 1)	0.0071	0.0007
$\sigma^2_{\phi S2}$ (Son 2)	0.0082	0.0009

Sibling correlation in permanent earnings



Sibling correlation - Decomposition

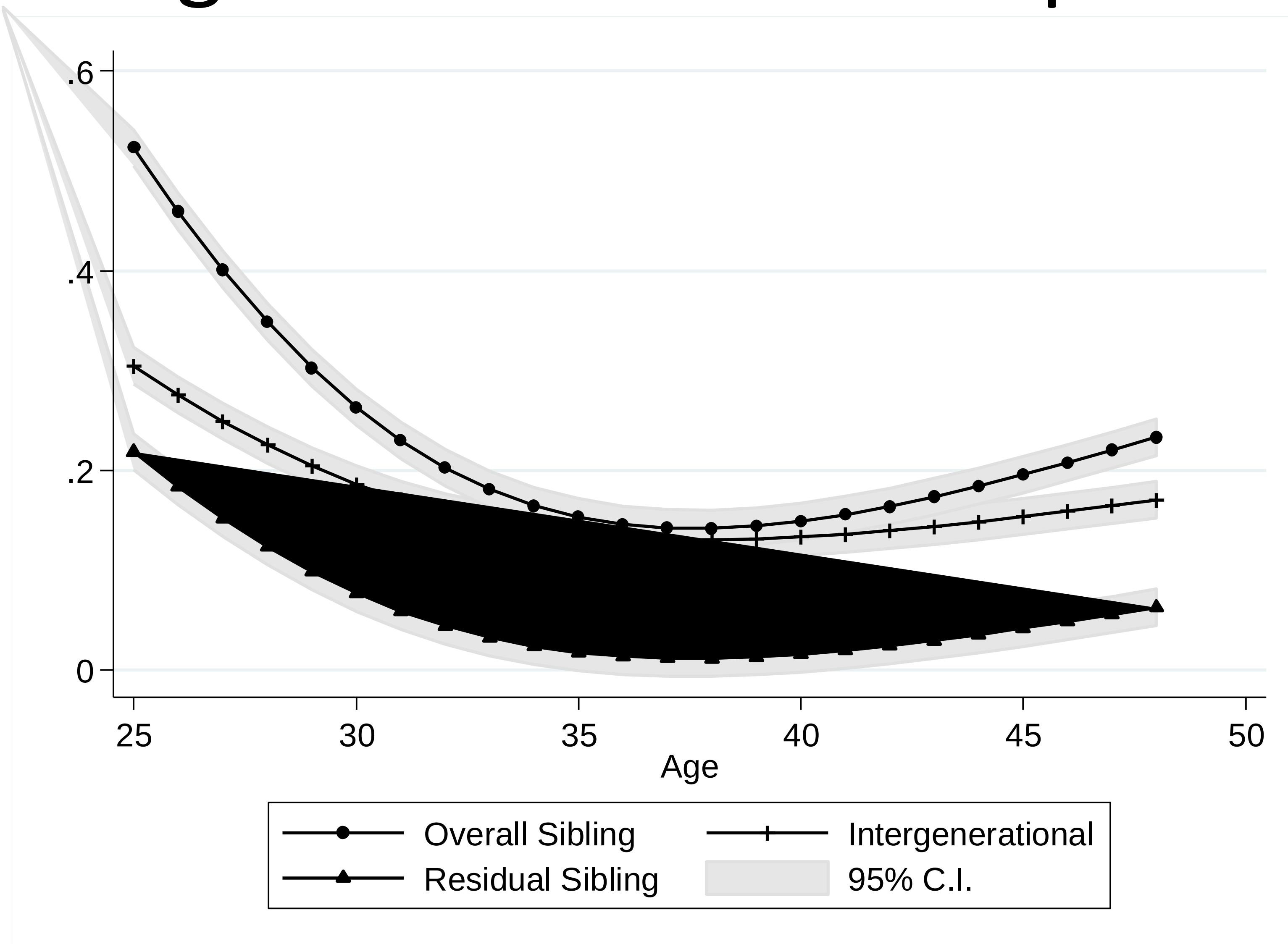


Table 3: Estimates of member-specific AR(1) parameters of transitory earnings

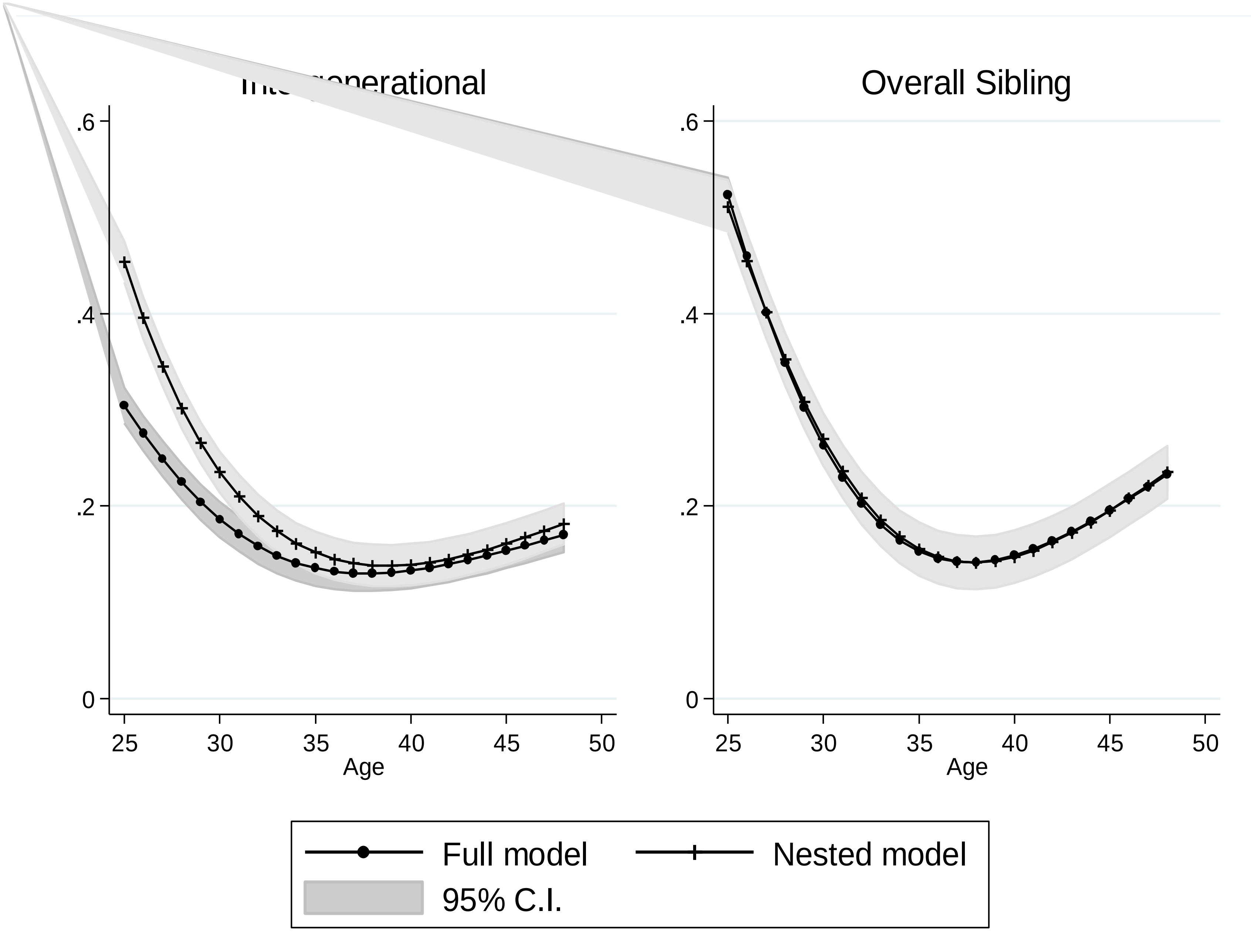
	Father		Son 1		Son 2	
	Coeff.	S.E.	Coeff.	S.E.	Coeff.	S.E.
$\sigma^2_{\varepsilon h}$(Baseline variance)	0.2847	0.0355	0.2474	0.0254	0.2309	0.0246
Age splines						
26-30	-0.1024	0.0476	-0.1357	0.0037	-0.1392	0.0065
31-35	-0.0286	0.0176	-0.0501	0.0034	-0.0644	0.0066
36-40	-0.0263	0.0111	-0.0031	0.0040	-0.0002	0.0082
41-45	0.0010	0.0127	-0.0348	0.0093	-0.0134	0.0197
46-51	-0.0199	0.0055	-0.0301	0.0133	-0.1052	0.0483
52-60	0.0591	0.0029				
ρ_h(Autocorrelation coefficient)	0.5136	0.0102	0.5141	0.0034	0.5213	0.0055
σ^2_{sh}(Baseline initial condition)	0.2558	0.0255	0.4115	0.0419	0.4126	0.0428
η_c(Initial condition shifter for left-censored cohorts, 1953-55=1)						
1935-37	1.3514	0.1982				
1938-40	1.4657	0.1895				
1941-43	1.3005	0.1585				
1944-46	1.0929	0.1257				
1947-49	0.8896	0.0972				
1950-52	0.9384	0.0961				
σ_{hl}(Between-person covariance)						
Father			0.0027	0.0003	0.0030	0.0003
Son1					0.0066	0.0007

Three nested models

- Without life-cycle effects
  - Model underlying Solon decomposition
- Intergenerational-only
  - Constrain residual sibling component to 0
  - Check plausibility of assumed zero correlation of IG & residual sib effects.
- Siblings-only
  - Constrain IG component to 0, use only sib moments
  - Can sib model capture IG effects?

Nested models - findings

- Without life-cycle effects
  - $r^S = 0.22$
  - 43% is accounted for by father's earnings
- Intergenerational-only
  - Over-predicts IG compared to full model
  - By 0.1 at age 25, insignificantly different by 30
  - Bias from omitted correlation modest & for young
- Siblings-only
  - No substantive difference to full model

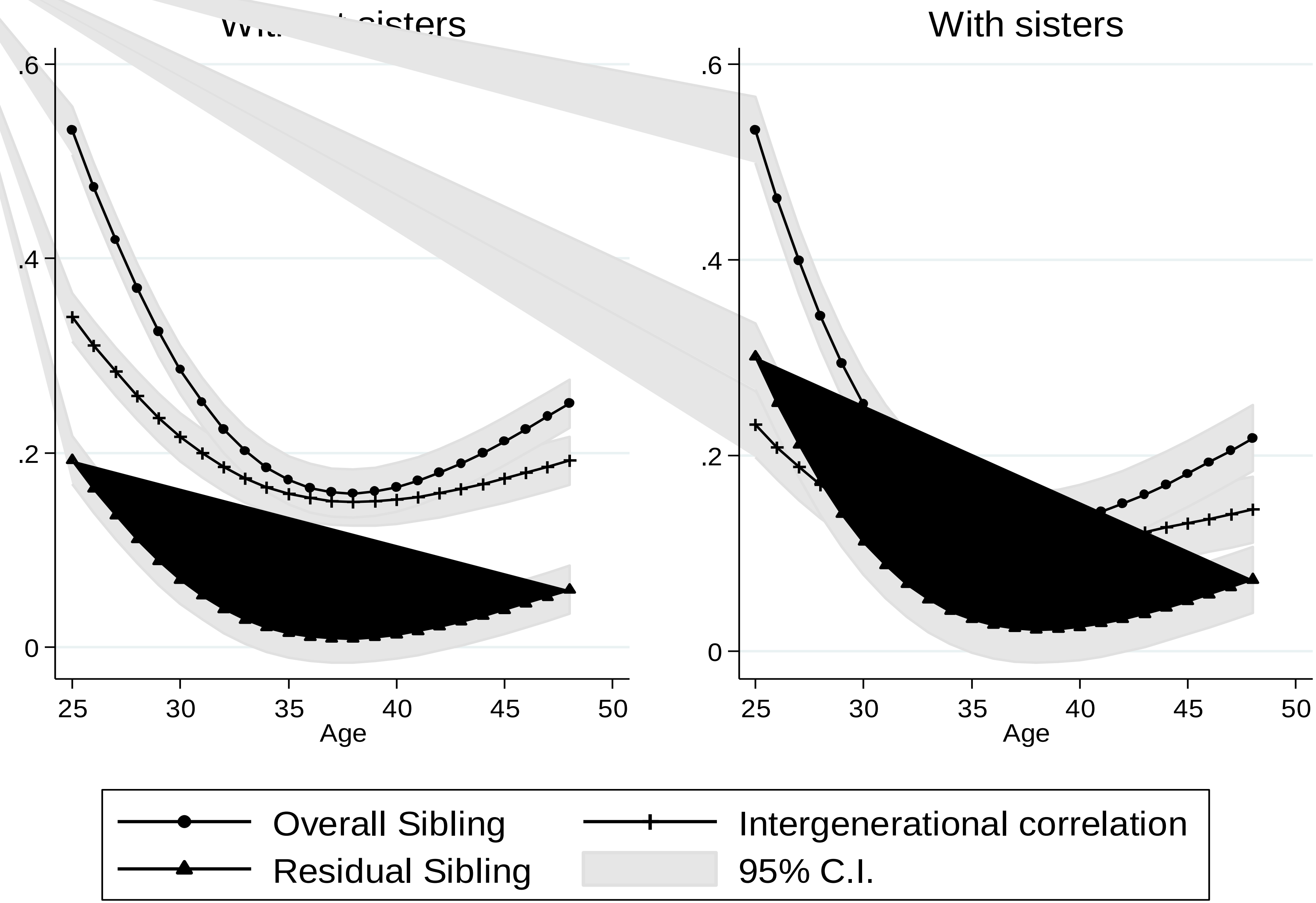

Intergenerational
Overall Sibling
.6
.4
.2
0
25
30
35
40
45
50
Age
Full model
Nested model
95% C.I.

Household Structure

- Families with more than two sons represent a small proportion of the population (<5%)

- We investigate the impact of household structure by focussing on sisters.

- Divide families in two groups:
  1. no sisters
  2. at least one sister


With sisters
.6
.4
.2
0
25
30
35
40
45
50
Age
Age
Overall Sibling
Intergenerational correlation
Residual Sibling
95% C.I.

Differential transmission

- We observe intergenerational transmission to two sons.

- Can model differential transmission:

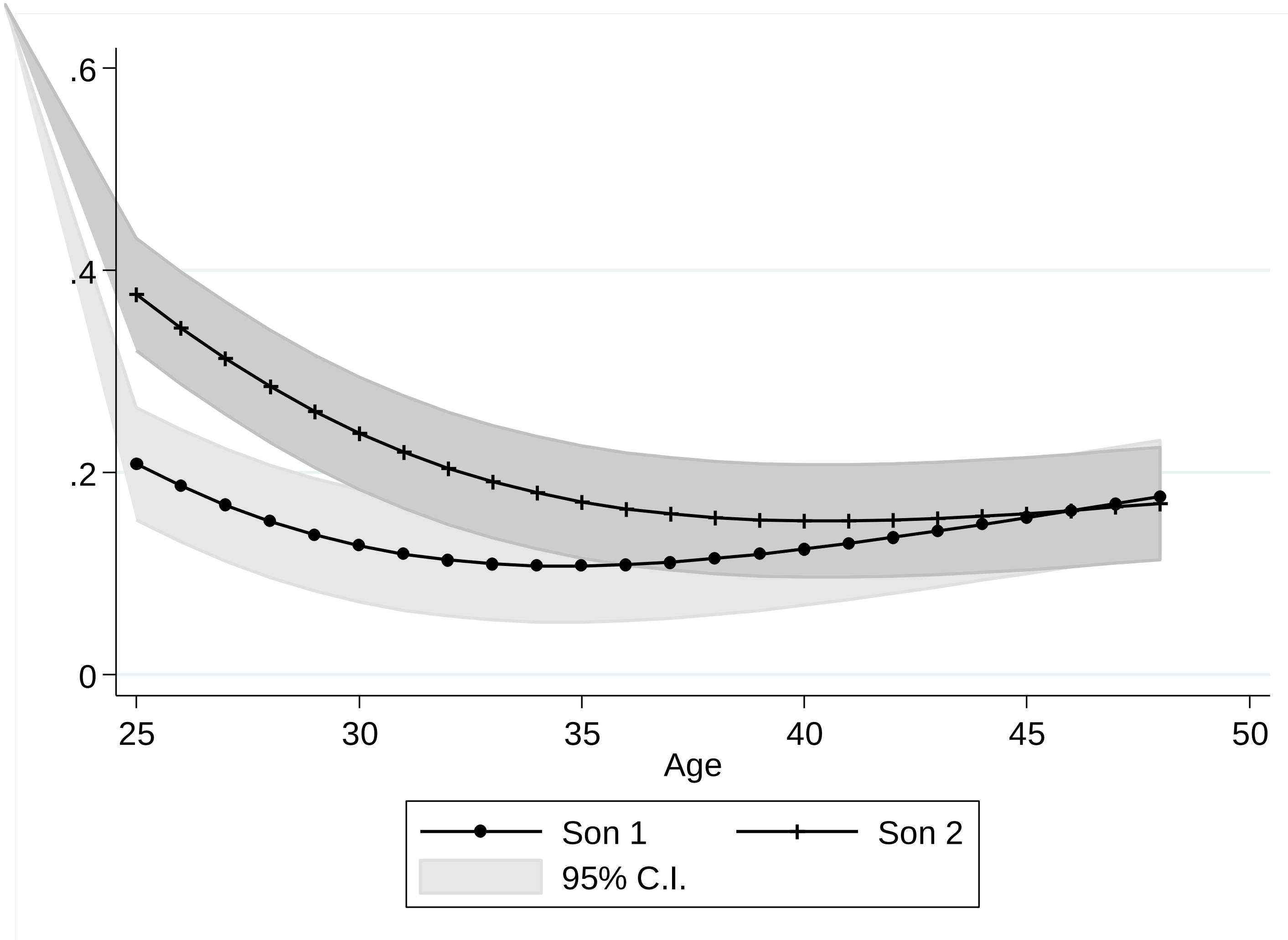
$$y_{ijt}^{S2} = \left(\left(\delta_\mu \mu_j^I + \mu_j^R\right) + \left(\delta_\lambda \lambda_j^I + \lambda_j^R\right) A_{it} + \omega_{ijt}^{S2} \right) \pi_t$$

- Use triplets-only sample.

Differential transmission - estimates

	(3) Differential IG	
	Coeff.	S.E.
Shared components		
Variance of initial earnings		
δ_μ (Intergenerational loading Son 2)	1.3212	0.0188
Variance of earnings growth rates		
δ_γ (Intergenerational loading Son 2)	0.9689	0.0046

Differential IG correlations



Differential transmission - interpretation

- Might seem at odd with findings of birth order studies.
- But looking at two distinct aspects: levels vs correlation.
- Might reflect:
  - experience in parenting
  - more established socio-economic status of parents.
- Can be predicted by birth order model, e.g. poor families investing more in first born and exhausting resources.

Summary

- Demonstrate the value of analysing triplets
- Intergenerational is *most* of sibling correlation
  - Much higher than previous estimates
- Sibling correlation u-shaped in age
  - Especially high for starting wages
- Differential transmission by birth order
  - Mild evidence of larger correlation with later born