Mobility Across Multiple Generations: The Iterated Regression Fallacy

Jan Stuhler*

*University College London, Centre for Research and Analysis of Migration, IZA

16th May 2014

Inter- vs. Multigenerational Mobility

We know much about *intergenerational* mobility in socio-econ. outcomes.

Galton, Conlisk, Goldberger, Becker and Tomes. Large differences across countries; persistence much higher than previously thought in some (US income elasticity $\beta \approx 0.5$, not ≈ 0.2); trends (Chetty et al.)

But little evidence on *long-run* mobility across *multiple* generations.
Hypotheses instead derived from intergenerational evidence.

This paper studies/argues theoretically (with empirical illustration):

- relationship between inter- and multigenerational mobility
  indirect transmission / multiplicity of skills / grandparents
- standard extrapolation from intergenerational evidence not valid
  the "iterated regression fallacy"
- multigenerational persistence higher than commonly claimed

Inter- vs. Multigenerational Mobility

We know much about *intergenerational* mobility in socio-econ. outcomes.

Galton, Conlisk, Goldberger, Becker and Tomes. Large differences across countries; persistence much higher than previously thought in some (US income elasticity $\beta \approx 0.5$, not ≈ 0.2); trends (Chetty et al.)

But little evidence on *long-run* mobility across *multiple* generations.
Hypotheses instead derived from intergenerational evidence.

This paper studies/argues theoretically (with empirical illustration):

- relationship between inter- and multigenerational mobility
  indirect transmission / multiplicity of skills / grandparents
- standard extrapolation from intergenerational evidence not valid
  the "iterated regression fallacy"
- multigenerational persistence higher than commonly claimed

Standard extrapolation procedure ...

Without direct evidence, we rely on extrapolation from parent-child elasticities. For example, Hertz (2006):

> *"Consider a rich and a poor family [...] and ask how much of the difference in the parents' incomes would be transmitted, on average, to their grandchildren. In the United States this would be $(0.47)^2$ or 22 percent;"*

Extrapolation-by-exponentiation is very prevalent, featuring in policy reports, textbooks (Borjas, 2009), survey articles (Piketty, 2000) ...

It has important implications. Becker and Tomes (1986):

> *"Almost all earnings advantages and disadvantages of ancestors are wiped out in three generations. Poverty would not [...] persist for several generations."*

Standard extrapolation procedure ...

Without direct evidence, we rely on extrapolation from parent-child elasticities. For example, Hertz (2006):

> *"Consider a rich and a poor family [...] and ask how much of the difference in the parents' incomes would be transmitted, on average, to their grandchildren. In the United States this would be $(0.47)^2$ or 22 percent;"*

Extrapolation-by-exponentiation is very prevalent, featuring in policy reports, textbooks (Borjas, 2009), survey articles (Piketty, 2000) ...

It has important implications. Becker and Tomes (1986):

> *"Almost all earnings advantages and disadvantages of ancestors are wiped out in three generations. Poverty would not [...] persist for several generations."*

... implies high long-run mobility:

Extrapolation thus provides ammunition for a contrarian standpoint that disputes the significance of low intergenerational mobility.

Mankiw (2006):

> *I am struck by how much "spin" there is here [...] one can just as easily put the point in a different light: "How much does income inequality persist from generation to generation? After two generations, 78 percent of the benefit of being born into a wealthy family has dissipated." I think many people would find this to be a surprisingly small degree of persistence.*

The Iterated Regression Fallacy

Intergenerational income elasticity: slope coef. in linear regression of offspring on parental log lifetime income (family i, generation t)

$$y_{it} = \beta_{-1} y_{it-1} + \varepsilon_{it}. \tag{1}$$

Extrapolation may seem natural: if β_{-1} measures how parental deviations from the mean are passed to their children then $(\beta_{-1})^2$ measures what remains after being passed twice from parents to children?

$$\beta_{-2} = \frac{Cov(y_{it}, y_{it-2})}{Var(y_{it-2})} = \frac{Cov(\beta_{-1} y_{it-1} + \varepsilon_{it}, y_{it-2})}{Var(y_{it-2})} = (\beta_{-1})^2.$$

Error in last step: ε_{it} is uncorrelated with parental income y_{it-1} (by construction), but not necessarily with grandparental income y_{it-2}.

The Iterated Regression Fallacy

Intergenerational income elasticity: slope coef. in linear regression of offspring on parental log lifetime income (family i, generation t)

$$y_{it} = \beta_{-1} y_{it-1} + \varepsilon_{it}. \tag{1}$$

Extrapolation may seem natural: if β_{-1} measures how parental deviations from the mean are passed to their children then $(\beta_{-1})^2$ measures what remains after being passed twice from parents to children?

$$\beta_{-2} = \frac{Cov(y_{it}, y_{it-2})}{Var(y_{it-2})} = \frac{Cov(\beta_{-1} y_{it-1} + \varepsilon_{it}, y_{it-2})}{Var(y_{it-2})} = (\beta_{-1})^2.$$

Error in last step: ε_{it} is uncorrelated with parental income y_{it-1} (by construction), but not necessarily with grandparental income y_{it-2}.

The Iterated Regression Fallacy

A "classic" regression fallacy?*

- Francis Galton fell fault of it (Bulmer, 2003)
- the intergenerational literature
- other literatures: the "convergence hypothesis" of the neo-classical growth model (Bernard and Durlauf, 1996)
- other disciplines (Nesselroade et al., 1980)

*such as: regression to the mean does not imply convergence to the mean (Friedman, 1992); or the failure to account for it in comparisons over time (Jerrim and Vignoles, 2012).

Inter- vs. Multigenerational Mobility, Swedish Registers:

	Child		Father		Grandfather
two generations		0.238***		0.406***	
		(0.002)		(0.004)	
three gen. (prediction)			0.096***		
			(0.001)		
three gen. (actual)			0.137***		
			(0.003)		

	Child		Mother		Grandmother
two generations		0.267***		0.301***	
		(0.002)		(0.005)	
three gen. (prediction)			0.080***		
			(0.002)		
three gen. (actual)			0.152***		
			(0.004)		

Notes: Slope coefficients from separate regressions of years of schooling of offspring on years of schooling of family member in older generation. N=145,590 observations for panel A (fathers/grandfathers), N=156,847 for panel B (mothers/grandmothers). Standard errors (in parantheses) are clustered on fathers (panel A) or mothers (panel B).

A simple model

Simplified one-parent one-offspring family structure:

$$y_{it} = \rho e_{it} + u_{it} \tag{2}$$

$$e_{it} = \lambda e_{it-1} + v_{it}, \tag{3}$$

- y_{it}: log lifetime income in generation t of family i
- e_{it}: human capital
- ρ: transferability; λ: heritability
- noise terms u_{it} and v_{it}: market and endowment luck (uncorrelated with each other and past values)

Assume throughout that variables are measured as trendless indices with mean zero and variance one.

Indirect transmission

Given equations (2) and (3) the intergenerational elasticity equals

$$\begin{aligned} \beta_{-1} &= Cov(y_t, y_{t-1}) \\ &= \rho^2 \lambda, \end{aligned} \tag{4}$$

and across three generations

$$\begin{aligned} \beta_{-2} &= Cov(y_t, y_{t-2}) \\ &= \rho^2 \lambda^2. \end{aligned} \tag{5}$$

The extrapolation error from exponentiating (4) equals

$$\begin{aligned} \Delta &= (\beta_{-1})^2 - \beta_{-2} \\ &= (\rho^2 - 1)\rho^2 \lambda^2 \end{aligned} \tag{6}$$

which is negative for $0 < \rho < 1$ and $0 < \lambda < 1$, that is as long as the intergenerational transmission of human capital and its transformation into income are not perfect.

Indirect transmission

Given equations (2) and (3) the intergenerational elasticity equals

$$\begin{aligned} \beta_{-1} &= Cov(y_t, y_{t-1}) \\ &= \rho^2 \lambda, \end{aligned} \tag{4}$$

and across three generations

$$\begin{aligned} \beta_{-2} &= Cov(y_t, y_{t-2}) \\ &= \rho^2 \lambda^2. \end{aligned} \tag{5}$$

The extrapolation error from exponentiating (4) equals

$$\begin{aligned} \Delta &= (\beta_{-1})^2 - \beta_{-2} \\ &= (\rho^2 - 1)\rho^2 \lambda^2 \end{aligned} \tag{6}$$

which is negative for $0 < \rho < 1$ and $0 < \lambda < 1$, that is as long as the intergenerational transmission of human capital and its transformation into income are not perfect.

Indirect transmission

- Both heritability and transferability of traits affect persistence
- But long-run persistence depends more on heritability
  $\beta_{-2} = \beta_{-1}\lambda$ across two generations, $\beta_{-3} = \beta_{-1}\lambda^2$ across three, ...

Substantial extrapolation error possible. Assume $\beta_{-1} = 0.5$

- extrapolation implies $\{\beta_{-1}, \beta_{-2}, \beta_{-3}\} = \{0.5, 0.25, 0.125\}$
- if $\rho = 0.8$ (market luck explains one third of income variance) then instead $\{\beta_{-1}, \beta_{-2}, \beta_{-3}\} = \{0.5, 0.39, 0.31\}$

Implications:

1. Difference between intergenerational and long-run mobility smaller if the former more due to imperfect transferability than low heritability.
2. Do cross-country differences in mobility extend to long-run?
   Example: high intergenerational mobility in Nordic country may not extend to long-run if due to policies that interfere with formation of market prices for traits

Indirect transmission

- Both heritability and transferability of traits affect persistence
- But long-run persistence depends more on heritability
  $\beta_{-2} = \beta_{-1}\lambda$ across two generations, $\beta_{-3} = \beta_{-1}\lambda^2$ across three, ...

Substantial extrapolation error possible. Assume $\beta_{-1} = 0.5$

- extrapolation implies $\{\beta_{-1}, \beta_{-2}, \beta_{-3}\} = \{0.5, 0.25, 0.125\}$
- if $\rho = 0.8$ (market luck explains one third of income variance) then instead $\{\beta_{-1}, \beta_{-2}, \beta_{-3}\} = \{0.5, 0.39, 0.31\}$

Implications:

1. Difference between intergenerational and long-run mobility smaller if the former more due to imperfect transferability than low heritability.
2. Do cross-country differences in mobility extend to long-run?
   Example: high intergenerational mobility in Nordic country may not extend to long-run if due to policies that interfere with formation of market prices for traits

Indirect transmission

- Both heritability and transferability of traits affect persistence
- But long-run persistence depends more on heritability
  $\beta_{-2} = \beta_{-1}\lambda$ across two generations, $\beta_{-3} = \beta_{-1}\lambda^2$ across three, ...

Substantial extrapolation error possible. Assume $\beta_{-1} = 0.5$

- extrapolation implies $\{\beta_{-1}, \beta_{-2}, \beta_{-3}\} = \{0.5, 0.25, 0.125\}$
- if $\rho = 0.8$ (market luck explains one third of income variance) then instead $\{\beta_{-1}, \beta_{-2}, \beta_{-3}\} = \{0.5, 0.39, 0.31\}$

Implications:

1. Difference between intergenerational and long-run mobility smaller if the former more due to imperfect transferability than low heritability.
2. Do cross-country differences in mobility extend to long-run?
   Example: high intergenerational mobility in Nordic country may not extend to long-run if due to policies that interfere with formation of market prices for traits

An additional factor

Introduce an additional *factor* into our starting model,

$$y_t = \rho_1 e_t + \rho_2 a_t + u_t \tag{7}$$

$$e_t = \lambda_1 e_{t-1} + v_t \tag{8}$$

$$a_t = \lambda_2 a_{t-1} + w_t. \tag{9}$$

Parents inherit two characteristics according to heritability parameters λ_1 and λ_2. Assume $0 < \rho_1 < 1$ and $0 < \rho_2 < 1$.

The parent-child elasticity then equals

$$\beta_{-1} = \rho_1^2 \lambda_1 + \rho_2^2 \lambda_2, \tag{10}$$

and the grandparent-grandchild elasticity equals

$$\beta_{-2} = \rho_1^2 \lambda_1^2 + \rho_2^2 \lambda_2^2. \tag{11}$$

An additional factor

Introduce an additional *factor* into our starting model,

$$y_t = \rho_1 e_t + \rho_2 a_t + u_t \tag{7}$$
$$e_t = \lambda_1 e_{t-1} + v_t \tag{8}$$
$$a_t = \lambda_2 a_{t-1} + w_t. \tag{9}$$

Parents inherit two characteristics according to heritability parameters λ_1 and λ_2. Assume $0 < \rho_1 < 1$ and $0 < \rho_2 < 1$.

The parent-child elasticity then equals

$$\beta_{-1} = \rho_1^2 \lambda_1 + \rho_2^2 \lambda_2, \tag{10}$$

and the grandparent-grandchild elasticity equals

$$\beta_{-2} = \rho_1^2 \lambda_1^2 + \rho_2^2 \lambda_2^2. \tag{11}$$

An additional factor

The extrapolation error equals

$$\Delta = (\rho_1^2 - 1)\rho_1^2\lambda_1^2 + (\rho_2^2 - 1)\rho_2^2\lambda_2^2 + 2\rho_1^2\rho_2^2\lambda_1\lambda_2. \tag{12}$$

Assume inherited characteristics are indeed *perfectly* transmitted into income, such that $\rho_1^2 + \rho_2^2 = 1$ and $Var(u_t) = 0$. Can rewrite ...

$$\Delta = \rho_1^2(\rho_1^2 - 1)(\lambda_1 - \lambda_2)^2. \tag{13}$$

Expression is negative for $\lambda_1 \neq \lambda_2$. *Jensen's inequality*: square of average heritability is smaller than the average of squared heritabilities. Intuition:

- intergenerational persistence of highly inheritable traits diminish slowly; explain increasingly larger share of long-run persistence
- long-run elasticities never converge to zero if some characteristics are perfectly transmitted.
  e.g. ethnicity may be highly persistent if interracial marriage is rare

An additional factor

The extrapolation error equals

$$\Delta = (\rho_1^2 - 1)\rho_1^2\lambda_1^2 + (\rho_2^2 - 1)\rho_2^2\lambda_2^2 + 2\rho_1^2\rho_2^2\lambda_1\lambda_2. \qquad (12)$$

Assume inherited characteristics are indeed *perfectly* transmitted into income, such that $\rho_1^2 + \rho_2^2 = 1$ and $Var(u_t) = 0$. Can rewrite ...

$$\Delta = \rho_1^2(\rho_1^2 - 1)(\lambda_1 - \lambda_2)^2. \qquad (13)$$

Expression is negative for $\lambda_1 \neq \lambda_2$. *Jensen's inequality*: square of average heritability is smaller than the average of squared heritabilities. Intuition:

- intergenerational persistence of highly inheritable traits diminish slowly; explain increasingly larger share of long-run persistence
- long-run elasticities never converge to zero if some characteristics are perfectly transmitted.
  e.g. ethnicity may be highly persistent if interracial marriage is rare

An additional generation

Results did not rely on the assumption of independent higher-order causal effects (e.g. from grandparents on their grandchildren).

Higher-order effects do raise long-run persistence (see paper).

Grandparents

But from the observation that $(\beta_{-1})^2 < \beta_{-2}$ we cannot conclude that the interg. transmission process has a memory of more than one generation.

An additional generation

	Years of schooling - Child				
	(1)	(2)	(3)	(4)	(5)
Parents:					
schooling father	0.222***	0.159***	0.135***	saturated	saturated
	(0.0025)	(0.0026)	(0.0027)		
schooling mother		0.182***	0.169***	saturated	saturated
		(0.00289)	(0.0029)		
income father			0.546***	0.461***	0.407***
			(0.0171)	(0.0169)	(0.0245)
income mother			-0.0176	-0.0021	0.0298*
			(0.0095)	(0.0094)	(0.0152)
Grandparents:					
schooling grandfather	**0.0456*****	**0.0259*****	**0.0183*****	**0.0083****	**0.0029**
(paternal)	(0.0031)	(0.0030)	(0.0030)	(0.0030)	(0.0047)
schooling grandmother					0.0069
(paternal)					(0.0060)
schooling grandfather					0.0061
(maternal)					(0.0046)
schooling grandmother					0.0069
(maternal)					(0.0059)
# obs.	104,904	104,904	104,904	104,904	47,797

Notes: Slope coefficients from separate regressions of years of schooling of offspring on characteristics of parents and grandparents. Standard errors (in parantheses) are clustered on mothers.

Parental investment

Previous discussion implies that extrapolations from intergenerational elasticities understate long-run persistence.

Can consider model in which multigenerational persistence is *below* extrapolation, e.g. if parental income has a strong and direct causal effect

- but channel seems speculative, while relevance of indirect transmission and multiplicity of skills does not.
- causal effect of parental income probably small (Björklund and Jäntti, 2009), part of it may work indirectly.

Conclusion: long-run persistence is higher, maybe much higher than implied by the standard interpretation of intergenerational elasticities.

Parental investment

Previous discussion implies that extrapolations from intergenerational elasticities understate long-run persistence.

Can consider model in which multigenerational persistence is *below* extrapolation, e.g. if parental income has a strong and direct causal effect

- but channel seems speculative, while relevance of indirect transmission and multiplicity of skills does not.
- causal effect of parental income probably small (Björklund and Jäntti, 2009), part of it may work indirectly.

Conclusion: long-run persistence is higher, maybe much higher than implied by the standard interpretation of intergenerational elasticities.

Conclusions

1. Extrapolation from intergenerational evidence widespread, but not valid: the "Iterated Regression Fallacy".
2. The relation between intergenerational and long-run mobility depends on causal pathways of transmission (interesting both ways).
3. Various *simple* theoretical reasons to expect that multigenerational persistence declines at less than geometric rate.
   market luck and indirect transmission; multiplicity of skills; grandparents(?)

Current wave of empirical papers seems supportive:

- Longitudinal data: Lindahl et al. (2014); Dribe and Helgertz (2013); Boserup et al. (2013)
- Repeated cross-sections: Long and Ferrie (2013); Collado et al. (2013); Clark (2013); Olivetti et al. (2014)

Conclusions

1. Extrapolation from intergenerational evidence widespread, but not valid: the "Iterated Regression Fallacy".
2. The relation between intergenerational and long-run mobility depends on causal pathways of transmission (interesting both ways).
3. Various *simple* theoretical reasons to expect that multigenerational persistence declines at less than geometric rate.
   market luck and indirect transmission; multiplicity of skills; grandparents(?)

Current wave of empirical papers seems supportive:

- Longitudinal data: Lindahl et al. (2014); Dribe and Helgertz (2013); Boserup et al. (2013)
- Repeated cross-sections: Long and Ferrie (2013); Collado et al. (2013); Clark (2013); Olivetti et al. (2014)

Appendix: An additional generation

Assume that offspring human capital depends on both parents and grandparents, such that equation (3) becomes

$$e_t = \lambda_{-1} e_{t-1} + \lambda_{-2} e_{t-2} + v_t, \tag{14}$$

Assuming stationarity the parent-child elasticity equals

$$\beta_{-1} = \rho^2 \left(\frac{\lambda_{-1}}{1 - \lambda_{-2}} \right), \tag{15}$$

Consider parameterizations that yield the same intergenerational elasticity as the previous model, such that $\lambda = \lambda_{-1}/(1 - \lambda_{-2})$. The grandparent-grandchild elasticity,

$$\beta_{-2} = \rho^2 \lambda^2 + \rho^2 \lambda_{-2}(1 - \lambda^2), \tag{16}$$

is then greater than the respective elasticity in the baseline model (assuming $\rho > 0$ and $\lambda < 1$). Back to data

Parental investment: case 1

Indirect effect of parental income, assume

$$y_t = \rho e_t + u_t \tag{17}$$

$$e_t = \theta y_{t-1} + \eta e_{t-1} + v_t. \tag{18}$$

The parent-child and grandparent-grandchild elasticities then equal

$$\beta_{-1} = \rho\theta + \rho^2\eta \quad , \quad \beta_{-2} = (\rho\eta + \rho^2\theta)(\rho\eta + \theta).$$

Consider again parameterizations that yield the same level of β_{-1}, which requires $\eta < \lambda$. The extrapolation error,

$$\Delta = (\rho^2 - 1)\eta\beta_{-1}, \tag{19}$$

is then smaller than the error in our first model (which equals $(\rho^2 - 1)\lambda\beta_{-1}$), but it will still be negative.

Parental investment: case 2

Direct effect of parental income, assume

$$y_t = \phi y_{y-1} + \tau e_t + u_t \tag{20}$$
$$e_t = \lambda e_{t-1} + v_t. \tag{21}$$

The parent-child and grandparent-grandchild elasticities then equal

$$\beta_{-1} = \phi + \frac{\tau^2 \lambda}{1 - \phi\lambda} \quad , \quad \beta_{-2} = \phi^2 + \frac{\tau^2 \lambda (\phi + \lambda)}{1 - \phi\lambda}.$$

The extrapolation error equals

$$\Delta = \left(\frac{\tau^2 \lambda}{1 - \phi\lambda} \right)^2 + (\phi - \lambda) \frac{\tau^2 \lambda}{1 - \phi\lambda}. \tag{22}$$

which may be positive. Intuition:

- short-run persistence affected by the direct income effect ϕ
- but long-run will be dominated by the heritability of ability λ.

Parental investment, some funny implications

If you are a conservative ...
and you believe that offspring from affluent parents tend to fare better because inherited traits and parental investment raise their productive abilities:

- then you should expect that long-run mobility is lower than implied by exponentiated intergenerational elasticities
- but the significance of low intergenerational mobility estimates is dismissed on the right wing precisely by the argument that they nevertheless imply high long-run mobility (e.g. see Mankiw 2006)

If you are a lefty ...
and you believe that intergenerational income correlations are instead due to mechanisms that resemble nepotism (e.g. parental networks)

- then you should expect that long-run mobility is high even when intergenerational mobility is relatively low