CONSTRUCTING A SUMMARY MEASURE OF INCOME MOBILITY FROM TRANSITION MATRICES

BY ADAM WARD
DATA SCIENTIST, DOT LOVES DATA

SECTION I

INTRODUCTION

INCOME MOBILITY

- Suppose we have a cohort of 10,000 people living in Sweden and a second cohort of 10,000 people living in the USA.
- Suppose that we were to determine the average monthly income of people in both cohorts in both 2010 and 2015.

Sweden

Individual	Income in 2010 (t=1)	Income in 2015 (t=2)
1	$829	$1,253
2	$1,101	$1,265
3	$914	$1,188
4	$1,016	$1,236
5	$1,002	$1,160
6	$931	$1,232
7	$941	$1,148
8	$1,023	$1,152
9	$824	$1,173
10	$969	$1,095
11	$1,075	$1,236
12	$865	$1,271
13	$1,129	$1,205
14	$1,079	$1,268
⋮	⋮	⋮
10,000	$723	$1,160

Which cohort exhibits a greater degree of income mobility?

USA

Individual	Income in 2010 (t=1)	Income in 2015 (t=2)
1	$1,054	$1,269
2	$979	$1,163
3	$1,151	$1,259
4	$964	$1,130
5	$888	$1,217
6	$1,042	$1,108
7	$1,143	$1,182
8	$970	$1,185
9	$925	$1,276
10	$987	$1,104
11	$1,062	$1,210
12	$929	$1,161
13	$932	$1,197
14	$1,163	$1,218
⋮	⋮	⋮
10,000	$678	$1,205

- There are two types of model used to measure income mobility:
  1. Elasticity Models.
  2. Transition Matrix Models.

SECTION II

ELASTICITY MODELS

ELASTICITY MODELS

- Assume that there is a (log) linear relationship between income at t=1 and t=2 and perform a regression.

$$\ln Y_{i,t=2} = \alpha + \beta \ln Y_{i,t=1} + \gamma Z_i + \varepsilon_i$$

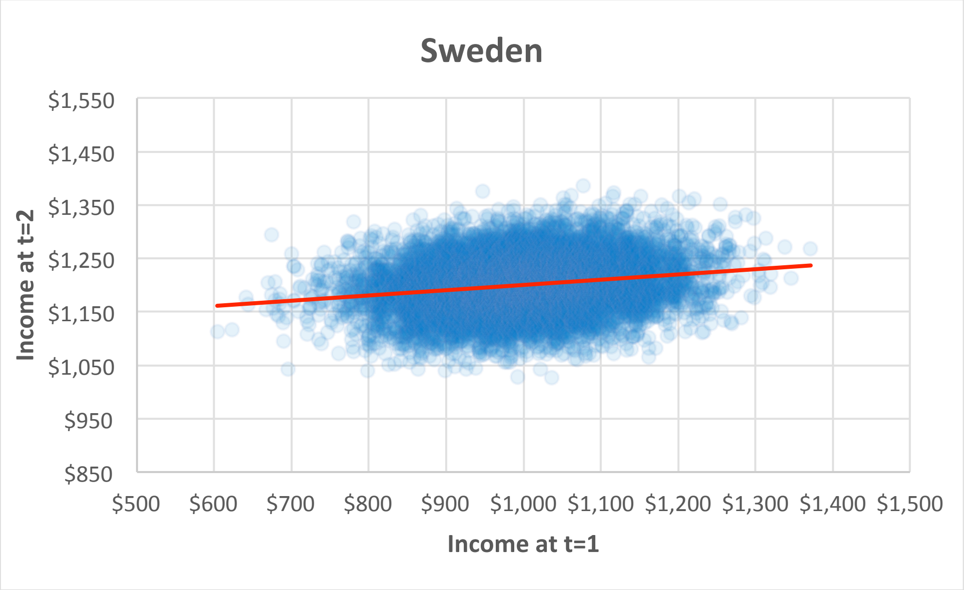
= Income of person i at time j

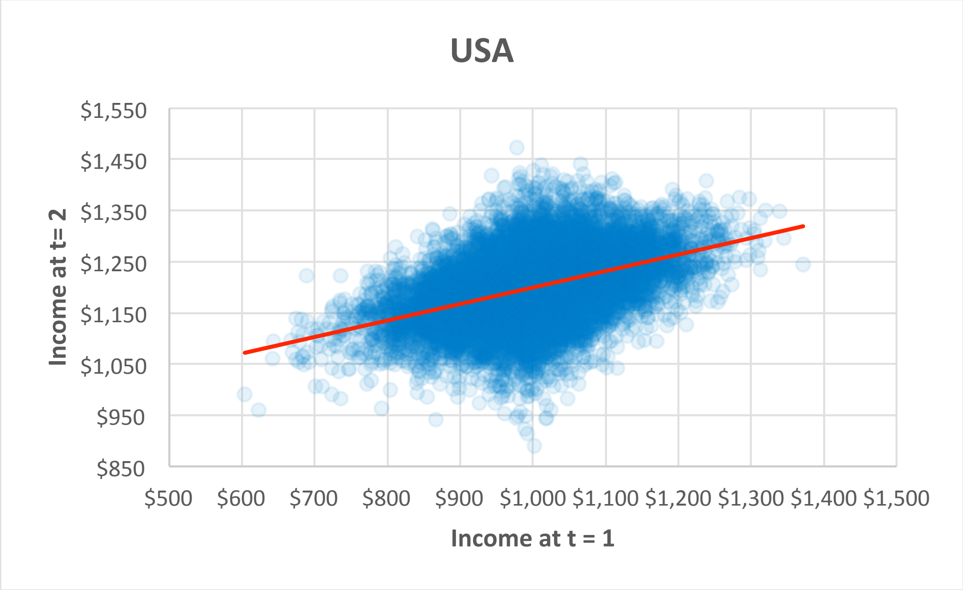
α = Constant term

Z_i = Control variables (e.g. education) relating to person i

ε_i = Error term relating to person i

- β is an **elasticity** measuring the average percentage change in $\ln(Y_{t=2})$ resulting from a 1% change in $\ln(Y_{t=1})$.
- The **greater** β is the **greater** the association between the income of an individual at t=1 and t=2 and the **lower** the level of mobility.





Measure	Sweden	USA
β	0.0804	0.2644
ρ	0.1935	0.4338
R^2	0.0375	0.1882

The individuals in Sweden exhibits a greater level of mobility than individuals in the USA.

ELASTICITY MODELS

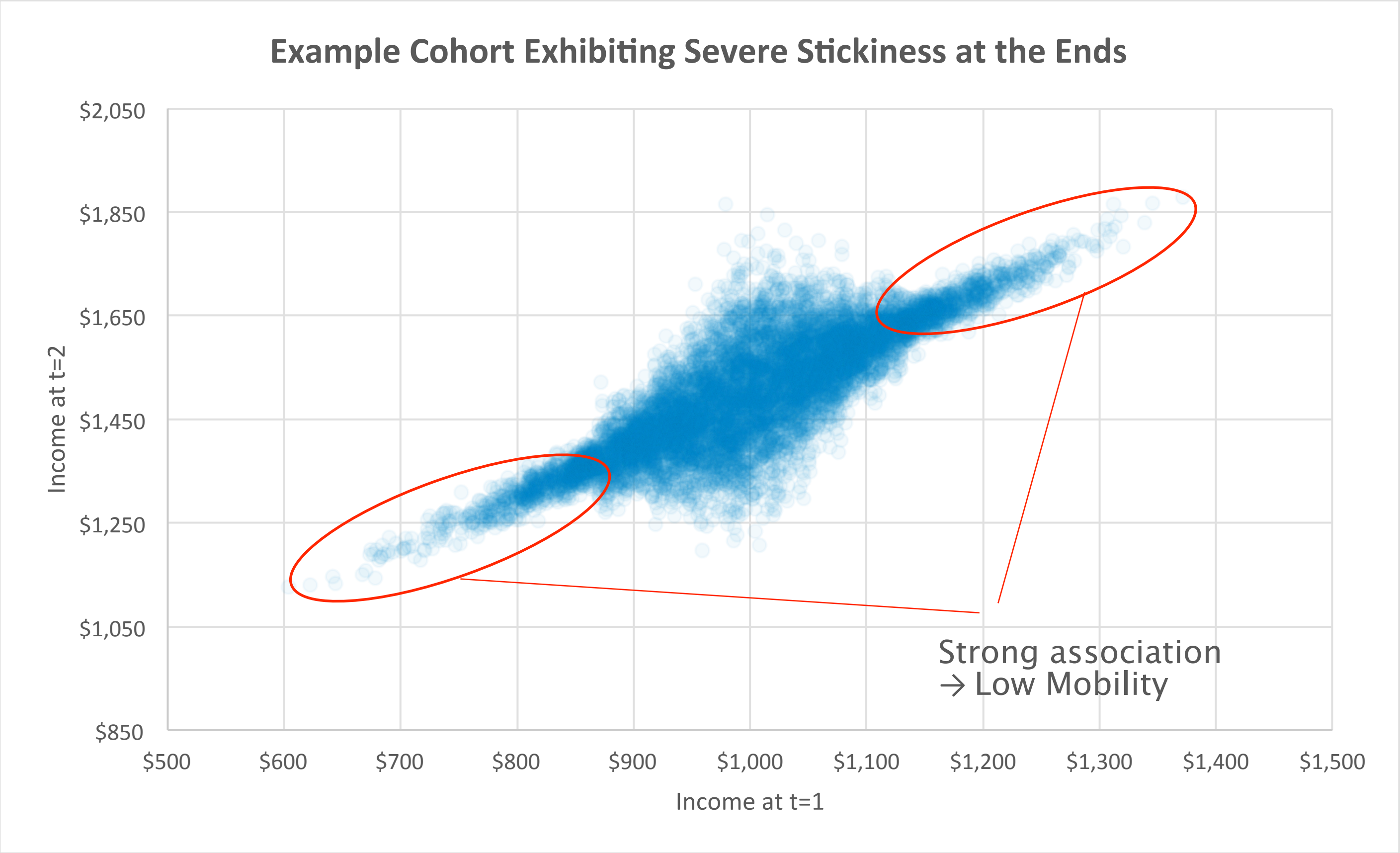
Problem 1- Subgroups

- Elasticity models cannot be used to compare mobility levels between subgroups of the population **relative to the population as a whole.**

- People at the extremes of the income distribution typically exhibit lower levels of mobility than people in the middle of the distribution.

- This phenomenon is referred to as **stickiness at the ends.**

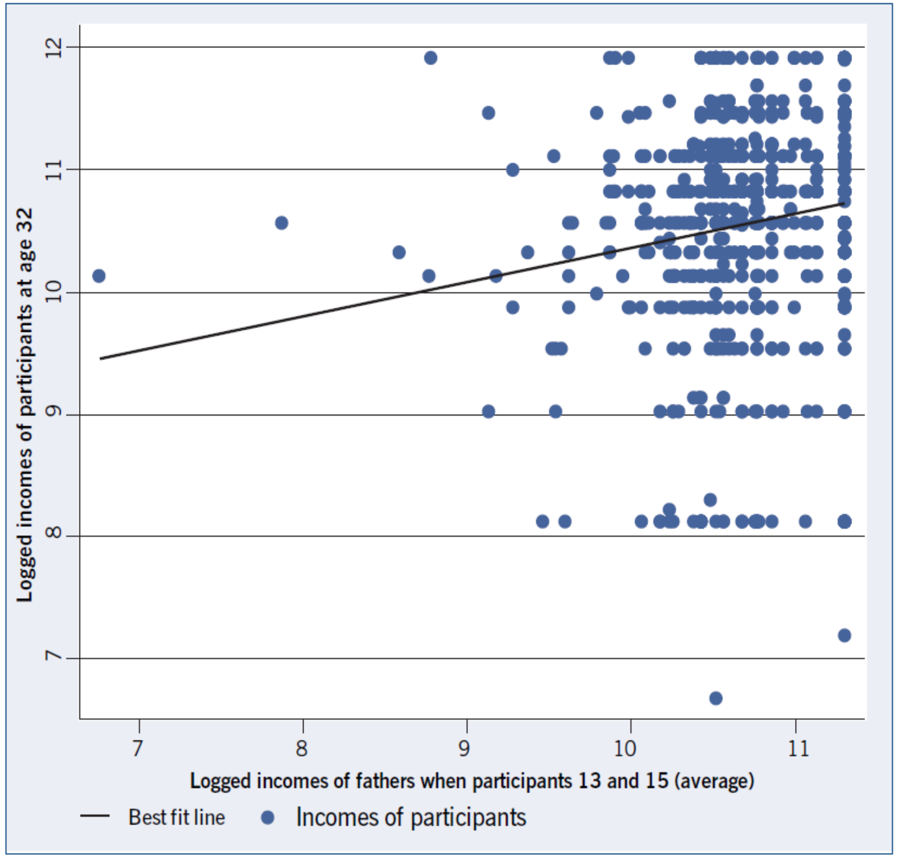
- Elasticity models can mask considerable differences in mobility between subgroups.



ELASTICITY MODELS

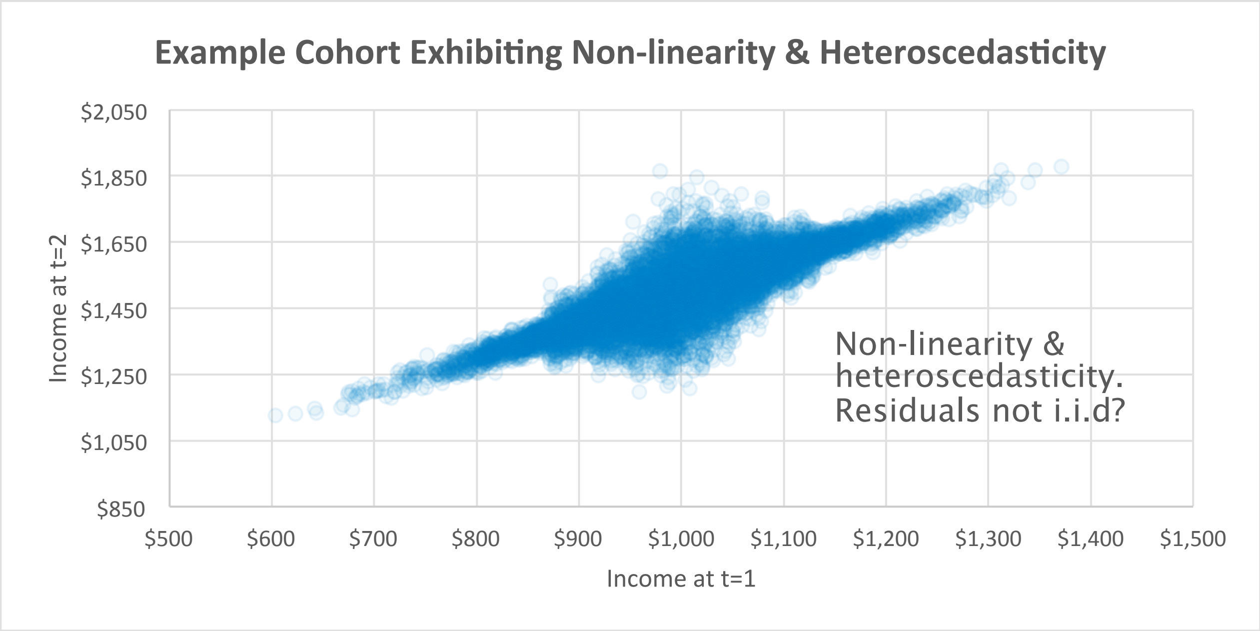
Problem 2 - Parametric

- Elasticity models **assume** that the relationship between $Y_{t=2}$ and $Y_{t=1}$ is (log) linear.



Intergenerational Economic Mobility in New Zealand: Gibbons 2008 [1]

- Often this is simply not the case.
- In such circumstances, one can question whether calculated elasticities are an appropriate measure of mobility.



SECTION III

TRANSITION MATRIX MODELS

TRANSITION MATRIX MODELS

- Rank individuals into *k* groups at t=1 and t=2 based on their income. By calculating the relative frequency of transition between groups we derive a transition matrix.

2015 (columns) / 2010 (rows)

USA	1	2	3	4	5	6	7	8	9	10
1	0.206	0.235	0.177	0.142	0.104	0.071	0.040	0.014	0.007	0.004
2	0.141	0.147	0.154	0.127	0.129	0.110	0.093	0.059	0.027	0.013
3	0.124	0.132	0.145	0.118	0.124	0.103	0.083	0.074	0.065	0.032
4	0.132	0.122	0.111	0.100	0.117	0.098	0.100	0.079	0.073	0.068
5	0.138	0.092	0.102	0.093	0.096	0.085	0.104	0.089	0.090	0.111
6	0.130	0.096	0.078	0.097	0.099	0.095	0.090	0.089	0.099	0.127
7	0.074	0.078	0.093	0.115	0.092	0.083	0.097	0.114	0.117	0.137
8	0.039	0.063	0.074	0.086	0.086	0.125	0.110	0.151	0.129	0.137
9	0.013	0.029	0.049	0.084	0.094	0.128	0.138	0.147	0.165	0.153
10	0.003	0.006	0.017	0.038	0.059	0.102	0.145	0.184	0.228	0.218

TRANSITION MATRIX MODELS

- Rank individuals into *k* groups at t=1 and t=2 based on their income. By calculating the relative frequency of transition between groups we derive a transition matrix.

2015

2010

USA	1	2	3	4	5	6	7	8	9	10
1	0.206	0.235	0.177	0.142	0.104	0.071	0.040	0.014	0.007	0.004
2	0.141	0.147	0.154	0.127	0.129	0.110	0.093	0.059	0.027	0.013
3	0.124	0.132	0.145	0.118	0.124	0.103	0.083	0.074	0.065	0.032
4	0.132	0.122	0.111	0.100	0.117	0.098	0.100	0.079	0.073	0.068
5	0.138	0.092	0.102	0.093	0.096	0.085	0.104	0.089	0.090	0.111
6	0.130	0.096	0.078	0.097	0.099	0.095	0.090	0.089	0.099	0.127
7	0.074	0.078	0.093	0.115	0.092	0.083	0.097	0.114	0.117	0.137
8	0.039	0.063	0.074	0.086	0.086	0.125	0.110	0.151	0.129	0.137
9	0.013	0.029	0.049	0.084	0.094	0.128	0.138	0.147	0.165	0.153
10	0.003	0.006	0.017	0.038	0.059	0.102	0.145	0.184	0.228	0.218

Advantages

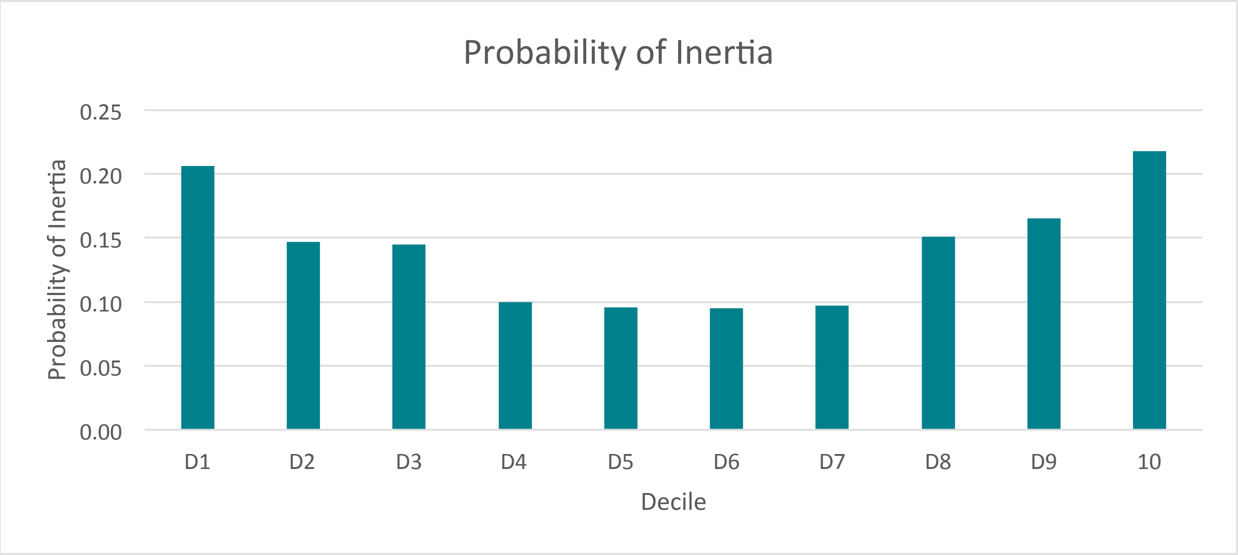
Non-parametric

No assumptions on the nature of the relationship between $Y_{t=2}$ and $Y_{t=1}$.

Subgroups

Probability of Inertia = Probability an individual is in the same group at t=1 and t=2.

Calculating the probability of inertia shows differences in mobility levels between individuals in different subgroups of the income distribution.

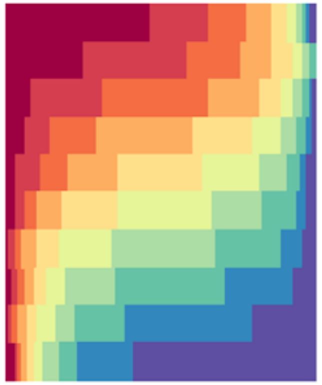


TRANSITION MATRIX MODELS

Problem 1 - Lack of a Summary Measure

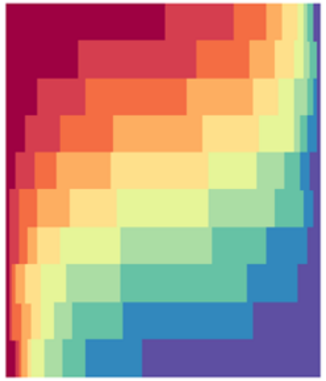
- Given two transition matrices….

USA	1	2	3	4	5	6	7	8	9	10
1	0.206	0.235	0.177	0.142	0.104	0.071	0.040	0.014	0.007	0.004
2	0.141	0.147	0.154	0.127	0.129	0.110	0.093	0.059	0.027	0.013
3	0.124	0.132	0.145	0.118	0.124	0.103	0.083	0.074	0.065	0.032
4	0.132	0.122	0.111	0.100	0.117	0.098	0.100	0.079	0.073	0.068
5	0.138	0.092	0.102	0.093	0.096	0.085	0.104	0.089	0.090	0.111
6	0.130	0.096	0.078	0.097	0.099	0.095	0.090	0.089	0.099	0.127
7	0.074	0.078	0.093	0.115	0.092	0.083	0.097	0.114	0.117	0.137
8	0.039	0.063	0.074	0.086	0.086	0.125	0.110	0.151	0.129	0.137
9	0.013	0.029	0.049	0.084	0.094	0.128	0.138	0.147	0.165	0.153
10	0.003	0.006	0.017	0.038	0.059	0.102	0.145	0.184	0.228	0.218



… which one represents the greater level of overall mobility?

Sweden	1	2	3	4	5	6	7	8	9	10
1	0.180	0.123	0.109	0.106	0.096	0.116	0.079	0.088	0.056	0.047
2	0.136	0.129	0.124	0.085	0.120	0.094	0.091	0.073	0.073	0.075
3	0.119	0.120	0.108	0.112	0.096	0.096	0.102	0.099	0.082	0.066
4	0.102	0.112	0.119	0.094	0.090	0.112	0.095	0.108	0.083	0.085
5	0.099	0.110	0.106	0.100	0.091	0.099	0.101	0.100	0.095	0.099
6	0.087	0.093	0.102	0.109	0.095	0.092	0.098	0.115	0.106	0.103
7	0.095	0.089	0.087	0.093	0.105	0.097	0.105	0.107	0.113	0.109
8	0.082	0.084	0.082	0.111	0.100	0.095	0.109	0.106	0.119	0.112
9	0.062	0.068	0.077	0.091	0.106	0.110	0.117	0.094	0.132	0.143
10	0.038	0.072	0.086	0.099	0.101	0.089	0.103	0.110	0.141	0.161



Have to analyze each group individually!

TRANSITION MATRIX MODELS

Problem 2 - Disaggregate Nature

- Different transition matrices (of different dimensions) will be produced depending on whether individuals are ranked into quartiles, quintiles, deciles, percentiles etc.
- In general best to rank individuals into as many groups as possible to capture the greatest amount of movement, i.e. use high dimensional matrices.
- However, lack of a summary measure means analyzing mobility in each group separately is an onerous task.

2x2 Matrix

USA	1	2
1	0.661	0.338
2	0.339	0.661

5x5 Matrix

USA	1	2	3	4	5
1	0.382	0.306	0.192	0.094	0.026
2	0.260	0.236	0.193	0.185	0.127
3	0.214	0.187	0.187	0.185	0.228
4	0.119	0.168	0.222	0.237	0.255
5	0.025	0.103	0.207	0.300	0.365

10x10 Matrix

USA	1	2	3	4	5	6	7	8	9	10
1	0.206	0.235	0.177	0.142	0.104	0.071	0.040	0.014	0.007	0.004
2	0.141	0.147	0.154	0.127	0.129	0.110	0.093	0.059	0.027	0.013
3	0.124	0.132	0.145	0.118	0.124	0.103	0.083	0.074	0.065	0.032
4	0.132	0.122	0.111	0.100	0.117	0.098	0.100	0.079	0.073	0.068
5	0.138	0.092	0.102	0.093	0.096	0.085	0.104	0.089	0.090	0.111
6	0.130	0.096	0.078	0.097	0.099	0.095	0.090	0.089	0.099	0.127
7	0.074	0.078	0.093	0.115	0.092	0.083	0.097	0.114	0.117	0.137
8	0.039	0.063	0.074	0.086	0.086	0.125	0.110	0.151	0.129	0.137
9	0.013	0.029	0.049	0.084	0.094	0.128	0.138	0.147	0.165	0.153
10	0.003	0.006	0.017	0.038	0.059	0.102	0.145	0.184	0.228	0.218

TRANSITION MATRIX MODELS

"A summary measure of mobility across relative positions would be useful to consolidate the information provided within transition matrices"

Bhattacharya & Mazumder [2]

Our aim is to derive a summary measure of mobility from a transition matrix which:

1. Condenses all of the information therein into a single real number, contained in the interval [0,1), describing the overall level of mobility.

2. Is applicable to high dimensional matrices.

SECTION IV

KEY ASSUMPTION

ASSUMPTION

In order to derive our summary measure of mobility we make the following assumption:

The movement of individuals between groups will continue to evolve according to the probabilities described within the transition matrix.

The future evolution of individuals between groups can be modelled according to the Markov chain defined by the observed transition matrix.

The individuals described by the transition matrix are engaged in a random walk along the directed, weighted graph defined by the transition matrix.

2015 (columns); 2010 (rows)

USA	1	2	3	4	5	6	7	8	9	10
1	0.206	0.235	0.177	0.142	0.104	0.071	0.040	0.014	0.007	0.004
2	0.141	0.147	0.154	0.127	0.129	0.110	0.093	0.059	0.027	0.013
3	0.124	0.132	0.145	0.118	0.124	0.103	0.083	0.074	0.065	0.032
4	0.132	0.122	0.111	0.100	0.117	0.098	0.100	0.079	0.073	0.068
5	0.138	0.092	0.102	0.093	0.096	0.085	0.104	0.089	0.090	0.111
6	0.130	0.096	0.078	0.097	0.099	0.095	0.090	0.089	0.099	0.127
7	0.074	0.078	0.093	0.115	0.092	0.083	0.097	0.114	0.117	0.137
8	0.039	0.063	0.074	0.086	0.086	0.125	0.110	0.151	0.129	0.137
9	0.013	0.029	0.049	0.084	0.094	0.128	0.138	0.147	0.165	0.153
10	0.003	0.006	0.017	0.038	0.059	0.102	0.145	0.184	0.228	0.218

ASSUMPTION

This is a BIG assumption…

1. Treats all individuals within each group as homogenous.

2. Assumes that the movement of individuals between groups is memoryless, i.e. is not affected by their previous positions in the income distribution.

ASSUMPTION

This is a BIG assumption…

1. Treats all individuals within each group as homogenous.

2. Assumes that the movement of individuals between groups is memoryless, i.e. is not affected by their previous positions in the income distribution.

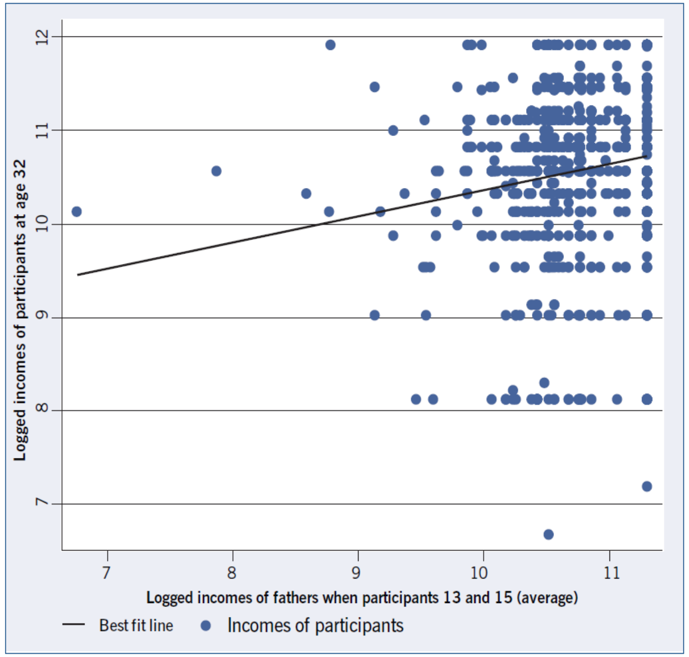
This is a BIG assumption… but …

1. We are applying the same assumption to both cohorts.

2. The greater the number of groups (the higher the dimension of the matrix) the more homogenous individuals within the same group are likely to be.

3. The assumption is no more or less valid than the assumption the relationship between income at t=1 and t=2 is log linear.

4. Could apply technique to other forms of mobility (wealth, earnings, socio-economic) where this assumption is more valid.



SECTION V

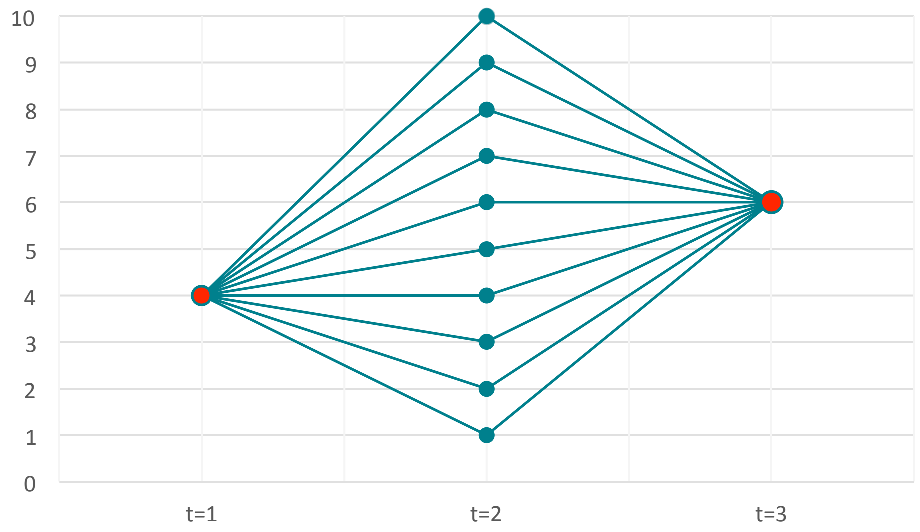
DOT

THEORY

N-STEP TRANSITION PROBABILITIES

- Given that we have assumed that the movement of individuals between groups will continue to evolve according to the probabilities described within the transition matrix, what is the probability that an individual moves from group i to group j
  - in 5 years = 1 transition $P_{i,j}^{(1)}$
  - in 10 years = 2 transitions $P_{i,j}^{(2)}$
  - in 15 years = 3 transitions $P_{i,j}^{(3)}$
  - in 5n years = n transitions $P_{i,j}^{(n)}$

Graphical Illustration of the Kolmogorov-Chapman Equation



- The well known Kolmogorov-Chapman equation states that:

$$P_{i,j}^{(2)} = \sum_{s=1}^{k} P_{i,s} * P_{s,j} = [P^2]_{i,j}$$

i.e. $P_{i,j}^{(2)}$ is equal to the (i,j)th element of the matrix *P* raised to the power of 2.

- By induction…

$$P_{i,j}^{(n)} = \sum_{s=1}^{k} P_{i,s}^{(n-1)} * P_{s,j} = [P^n]_{i,j}$$

i.e. $P_{i,j}^{(n)}$ is equal to the (i,j)th element of the matrix *P* raised to the power of *n*.

- **The probability that an individual goes from group *i* to group *j* in *n* transitions is equal to the i,jth element of the matrix P^n.**

LIMITING BEHAVIOR

- What happens to $P_{i,j}^{(n)}$ as n becomes very large?
- If the transition matrix is **ergodic, doubly stochastic** and **diagonalizable** then $P_{i,j}^{(n)} \rightarrow 1/k$ as $n \rightarrow \infty$ where k is the number of groups.

USA	1	2	3	4	5	6	7	8	9	10
1	0.206	0.235	0.177	0.142	0.104	0.071	0.040	0.014	0.007	0.004
2	0.141	0.147	0.154	0.127	0.129	0.110	0.093	0.059	0.027	0.013
3	0.124	0.132	0.145	0.118	0.124	0.103	0.083	0.074	0.065	0.032
4	0.132	0.122	0.111	0.100	0.117	0.098	0.100	0.079	0.073	0.068
5	0.138	0.092	0.102	0.093	0.096	0.085	0.104	0.089	0.090	0.111
6	0.130	0.096	0.078	0.097	0.099	0.095	0.090	0.089	0.099	0.127
7	0.074	0.078	0.093	0.115	0.092	0.083	0.097	0.114	0.117	0.137
8	0.039	0.063	0.074	0.086	0.086	0.125	0.110	0.151	0.129	0.137
9	0.013	0.029	0.049	0.084	0.094	0.128	0.138	0.147	0.165	0.153
10	0.003	0.006	0.017	0.038	0.059	0.102	0.145	0.184	0.228	0.218

LIMITING BEHAVIOR

- What happens to $P_{i,j}^{(n)}$ as n becomes very large?
- If the transition matrix is **ergodic, doubly stochastic** and **diagonalizable** then $P_{i,j}^{(n)} \to 1/k$ as $n \to \infty$ where k is the number of groups.

USA	1	2	3	4	5	6	7	8	9	10
1	0.206	0.235	0.177	0.142	0.104	0.071	0.040	0.014	0.007	0.004
2	0.141	0.147	0.154	0.127	0.129	0.110	0.093	0.059	0.027	0.013
3	0.124	0.132	0.145	0.118	0.124	0.103	0.083	0.074	0.065	0.032
4	0.132	0.122	0.111	0.100	0.117	0.098	0.100	0.079	0.073	0.068
5	0.138	0.092	0.102	0.093	0.096	0.085	0.104	0.089	0.090	0.111
6	0.130	0.096	0.078	0.097	0.099	0.095	0.090	0.089	0.099	0.127
7	0.074	0.078	0.093	0.115	0.092	0.083	0.097	0.114	0.117	0.137
8	0.039	0.063	0.074	0.086	0.086	0.125	0.110	0.151	0.129	0.137
9	0.013	0.029	0.049	0.084	0.094	0.128	0.138	0.147	0.165	0.153
10	0.003	0.006	0.017	0.038	0.059	0.102	0.145	0.184	0.228	0.218

ERGODICITY:

- **Ergodic** if there exists some $s \in \mathbb{N}$ so that $P_{i,j}^{(s)} > 0$ for all $i,j \in \Omega$.
- There is a non-zero probability of reaching any group from any other group in *exactly* s transition.
- All *real world* transition matrices are ergodic.

	Eigenvalues	Magnitude
λ_1	1	1.000
λ_2	0.435	0.435
λ_3	-0.005 + 0.06i	0.060
λ_4	-0.005 - 0.06i	0.060
λ_5	-0.034	0.034
λ_6	0.015 + 0.011i	0.019
λ_7	0.015 - 0.011i	0.019
λ_8	-0.017	0.017
λ_9	0.017	0.017
λ_{10}	0	0

CONSEQUENCE:

- Largest eigenvalue of matrix is equal to 1. All other eigenvalues are strictly less than 1 in magnitude (Perron-Frobenius theorem [3])

$$\lambda_1 = 1 > |\lambda_2| \geq |\lambda_3| \geq \ldots \geq |\lambda_k|$$

LIMITING BEHAVIOR

- What happens to $P_{i,j}^{(n)}$ as n becomes very large?
- If the transition matrix is **ergodic, doubly stochastic** and **diagonalizable** then $P_{i,j}^{(n)} \to 1/k$ as $n \to \infty$ where k is the number of groups.

USA	1	2	3	4	5	6	7	8	9	10	
1	0.206	0.235	0.177	0.142	0.104	0.071	0.040	0.014	0.007	0.004	**1**
2	0.141	0.147	0.154	0.127	0.129	0.110	0.093	0.059	0.027	0.013	**1**
3	0.124	0.132	0.145	0.118	0.124	0.103	0.083	0.074	0.065	0.032	**1**
4	0.132	0.122	0.111	0.100	0.117	0.098	0.100	0.079	0.073	0.068	**1**
5	0.138	0.092	0.102	0.093	0.096	0.085	0.104	0.089	0.090	0.111	**1**
6	0.130	0.096	0.078	0.097	0.099	0.095	0.090	0.089	0.099	0.127	**1**
7	0.074	0.078	0.093	0.115	0.092	0.083	0.097	0.114	0.117	0.137	**1**
8	0.039	0.063	0.074	0.086	0.086	0.125	0.110	0.151	0.129	0.137	**1**
9	0.013	0.029	0.049	0.084	0.094	0.128	0.138	0.147	0.165	0.153	**1**
10	0.003	0.006	0.017	0.038	0.059	0.102	0.145	0.184	0.228	0.218	**1**
	1	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	

DOUBLE STOCHASTICITY:

- **Stochastic** if all rows sum to 1. All transition matrices are stochastic.
- **Doubly stochastic** if all columns also sum to 1. *In theory* all transition matrices are doubly stochastic.

CONSEQUENCE:

- $\boldsymbol{e} = k^{-1/2}\,(1,1,\ldots,1)^T$ is a **right** eigenvector of P corresponding to the eigenvalue λ = 1

$$P\boldsymbol{e} = k^{-1/2}\,P\boldsymbol{e} = k^{-1/2}\,\boldsymbol{e} = \boldsymbol{e}$$

- $\boldsymbol{e}^{\boldsymbol{T}} = k^{-1/2}\,(1,1,\ldots,1)$ is a **left** eigenvector of P corresponding to the eigenvalue λ = 1

$$\boldsymbol{e}^T P = k^{-1/2}\,\boldsymbol{e}^T P = k^{-1/2}\,\boldsymbol{e}^{\boldsymbol{T}} = \boldsymbol{e}^T$$

LIMITING BEHAVIOR

- What happens to $P_{i,j}^{(n)}$ as n becomes very large?
- If the transition matrix is **ergodic, doubly stochastic** and **diagonalizable** then $P_{i,j}^{(n)} \to 1/k$ as $n \to \infty$ where k is the number of groups.

USA	1	2	3	4	5	6	7	8	9	10
1	0.206	0.235	0.177	0.142	0.104	0.071	0.040	0.014	0.007	0.004
2	0.141	0.147	0.154	0.127	0.129	0.110	0.093	0.059	0.027	0.013
3	0.124	0.132	0.145	0.118	0.124	0.103	0.083	0.074	0.065	0.032
4	0.132	0.122	0.111	0.100	0.117	0.098	0.100	0.079	0.073	0.068
5	0.138	0.092	0.102	0.093	0.096	0.085	0.104	0.089	0.090	0.111
6	0.130	0.096	0.078	0.097	0.099	0.095	0.090	0.089	0.099	0.127
7	0.074	0.078	0.093	0.115	0.092	0.083	0.097	0.114	0.117	0.137
8	0.039	0.063	0.074	0.086	0.086	0.125	0.110	0.151	0.129	0.137
9	0.013	0.029	0.049	0.084	0.094	0.128	0.138	0.147	0.165	0.153
10	0.003	0.006	0.017	0.038	0.059	0.102	0.145	0.184	0.228	0.218

DIAGONALIZABLE:

- **Diagonalizable** if can be written in the form:

$$\boldsymbol{P = \Sigma D \Sigma^{-1}}$$

Σ = matrix of right eigenvectors of *P.*

D = diagonal matrix of eigenvalues of P.

Σ^{-1} = inverse of Σ and matrix of left eigenvectors of *P.*

CONSEQUENCE:

- The set of defective (non-diagonalizable) transition matrices has zero Lebesgue measure when considered as a subset of $\mathbb{C}^{k\times k}$.
- Almost every transition matrix we encounter in practice will be diagonalizable.

LIMITING BEHAVIOR

- What happens to $P_{i,j}^{(n)}$ as n becomes very large?
- If the transition matrix is **ergodic, doubly stochastic** and **diagonalizable** then $P_{i,j}^{(n)} \rightarrow 1/k$ as $n \rightarrow \infty$ where k is the number of groups

Matrices Raised to the Power of n = 1

		SWEDEN									
		1	2	3	4	5	6	7	8	9	10
P^1	1	0.1800	0.1230	0.1090	0.1060	0.0960	0.1160	0.0790	0.0880	0.0560	0.0470
	2	0.1360	0.1290	0.1240	0.0850	0.1200	0.0940	0.0910	0.0730	0.0730	0.0750
	3	0.1190	0.1200	0.1080	0.1120	0.0960	0.0960	0.1020	0.0990	0.0820	0.0660
	4	0.1020	0.1120	0.1190	0.0940	0.0900	0.1120	0.0950	0.1080	0.0830	0.0850
	5	0.0990	0.1100	0.1060	0.1000	0.0910	0.0990	0.1010	0.1000	0.0950	0.0990
	6	0.0870	0.0930	0.1020	0.1090	0.0950	0.0920	0.0980	0.1150	0.1060	0.1030
	7	0.0950	0.0890	0.0870	0.0930	0.1050	0.0970	0.1050	0.1070	0.1130	0.1090
	8	0.0820	0.0840	0.0820	0.1110	0.1000	0.0950	0.1090	0.1060	0.1190	0.1120
	9	0.0620	0.0680	0.0770	0.0910	0.1060	0.1100	0.1170	0.0940	0.1320	0.1430
	10	0.0380	0.0720	0.0860	0.0990	0.1010	0.0890	0.1030	0.1100	0.1410	0.1610

|| P^n - A || = 0.20552

		USA									
		1	2	3	4	5	6	7	8	9	10
P^1	1	0.2060	0.2350	0.1770	0.1420	0.1040	0.0710	0.0400	0.0140	0.0070	0.0040
	2	0.1410	0.1470	0.1540	0.1270	0.1290	0.1100	0.0930	0.0590	0.0270	0.0130
	3	0.1240	0.1320	0.1450	0.1180	0.1240	0.1030	0.0830	0.0740	0.0650	0.0320
	4	0.1320	0.1220	0.1110	0.1000	0.1170	0.0980	0.1000	0.0790	0.0730	0.0680
	5	0.1380	0.0920	0.1020	0.0930	0.0960	0.0850	0.1040	0.0890	0.0900	0.1110
	6	0.1300	0.0960	0.0780	0.0970	0.0990	0.0950	0.0900	0.0890	0.0990	0.1270
	7	0.0740	0.0780	0.0930	0.1150	0.0920	0.0830	0.0970	0.1140	0.1170	0.1370
	8	0.0390	0.0630	0.0740	0.0860	0.0860	0.1250	0.1100	0.1510	0.1290	0.1370
	9	0.0130	0.0290	0.0490	0.0840	0.0940	0.1280	0.1380	0.1470	0.1650	0.1530
	10	0.0030	0.0060	0.0170	0.0380	0.0590	0.1020	0.1450	0.1840	0.2280	0.2180

|| P^n - A || = 0.47155

- The adjacent transition matrices are $k \times k = 10 \times 10$ ergodic, doubly stochastic and diagonalizable.
- When we raise the matrices to successively higher and higher powers, their elements tend to $P_{i,j}^{(n)} = 1/k = 1/10$

CONCLUSION:

Over a long enough time line, everyone has an equal chance of being in any income decile irrespective of the decile to which they originally belonged.

A MEASURE OF MOBILITY

The quicker a transition matrix converges to equilibrium
the greater the level of income mobility within the associated cohort.

The quicker a transition matrix converges to equilibrium:

- The sooner every individual has an equal chance of being any given group irrespective of the initial group to which they were originally assigned.
- The quicker the initial income advantages or disadvantages held by individuals dissipate.
- The less ‘sticky’ the relative position of individuals within the income distribution.
- The quicker the Markov chain defined by the observed transition matrix ‘forgets’ about its initial state.

WHAAT DETERMINES THE RATE OF CONVERGENCE TO EQUILIBRIUM?

- The **smaller** the magnitude of the second largest eigenvalue $|\lambda_2| \in [0,1)$, the **quicker** the transition matrix will converge to equilibrium.

Diagonalizability n times

$$P^n = (\Sigma D \Sigma^{-1})^n = (\Sigma D \Sigma^{-1}) \cdot (\Sigma D \Sigma^{-1}) \cdot \ldots \cdot (\Sigma D \Sigma^{-1}) = \Sigma D^n \Sigma^{-1}$$

Ergodicity

$$= \Sigma\, [■■■1@0 \;\&■0@\lambda_2^n \;\&\cdots\&■0@0 \;@■\vdots \;\&\vdots\;\&\ddots\&\vdots@■0 \;\&0\;\&\cdots\&\lambda_k^n]\, \Sigma^{-1} = \Sigma\, [a_{1,1}]\, \Sigma^{-1} + \Sigma$$

$[a_{j,j}]$ is the matrix with element $a_{j,j}$ equal to 1 and all other elements equal to 0.

The smaller $|\lambda_2|$ is the quicker these terms will tend to zero. Recall $1 > |\lambda_2| \geq |\lambda_3| \geq \ldots \geq |\lambda_k|$.

$\to 0 \;\; as \;\; n \to \infty$ since $|\lambda_j| < 1$

$$\underset{n \to \infty}{\to}_{as} [■\mathbf{e}, \&■\mathbf{r}_2, \&■\ldots, \&\mathbf{r}_k] \cdot [■■■1@0 \;\&■0@0 \;\&\cdots\&■0@0 \;@■\vdots\&\vdots \;\&\ddots\&\vdots@■0\&0 \;\&\cdots\&0] \cdot [■\mathbf{e}^T @■\mathbf{l}_2 @$$

Double Stochasticity

$$= [■■■k^{-1} @k^{-1} \;\&■k^{-1} @k^{-1} \;\&\cdots\&■k^{-1} @k^{-1} \;@■\vdots \;\&\vdots \;\&\ddots\&\vdots@■k^{-1} \&k^{-1} \&\cdots\&k^{-1}] \equiv$$

Equilibrium

Thus $|\lambda_2| \in [0,1)$ measures how quickly the transition matrix converges to equilibrium and therefore the degree of mobility described by the transition matrix.

COMPARISONS FOR SWEDEN & USA

- What happens to $P_{i,j}^{(n)}$ as n becomes very large?
- If the transition matrix is **ergodic, doubly stochastic** and **diagonalizable** then $P_{i,j}^{(n)} \to 1/k$ as $n \to \infty$ where k is the number of groups.

Matrices Raised to the Power of n = 1

		SWEDEN									
		1	2	3	4	5	6	7	8	9	10
P^1	1	0.1800	0.1230	0.1090	0.1060	0.0960	0.1160	0.0790	0.0880	0.0560	0.0470
	2	0.1360	0.1290	0.1240	0.0850	0.1200	0.0940	0.0910	0.0730	0.0730	0.0750
	3	0.1190	0.1200	0.1080	0.1120	0.0960	0.0960	0.1020	0.0990	0.0820	0.0660
	4	0.1020	0.1120	0.1190	0.0940	0.0900	0.1120	0.0950	0.1080	0.0830	0.0850
	5	0.0990	0.1100	0.1060	0.1000	0.0910	0.0990	0.1010	0.1000	0.0950	0.0990
	6	0.0870	0.0930	0.1020	0.1090	0.0950	0.0920	0.0980	0.1150	0.1060	0.1030
	7	0.0950	0.0890	0.0870	0.0930	0.1050	0.0970	0.1050	0.1070	0.1130	0.1090
	8	0.0820	0.0840	0.0820	0.1110	0.1000	0.0950	0.1090	0.1060	0.1190	0.1120
	9	0.0620	0.0680	0.0770	0.0910	0.1060	0.1100	0.1170	0.0940	0.1320	0.1430
	10	0.0380	0.0720	0.0860	0.0990	0.1010	0.0890	0.1030	0.1100	0.1410	0.1610

|| P^n - A || = 0.20552

		USA									
		1	2	3	4	5	6	7	8	9	10
P^1	1	0.2060	0.2350	0.1770	0.1420	0.1040	0.0710	0.0400	0.0140	0.0070	0.0040
	2	0.1410	0.1470	0.1540	0.1270	0.1290	0.1100	0.0930	0.0590	0.0270	0.0130
	3	0.1240	0.1320	0.1450	0.1180	0.1240	0.1030	0.0830	0.0740	0.0650	0.0320
	4	0.1320	0.1220	0.1110	0.1000	0.1170	0.0980	0.1000	0.0790	0.0730	0.0680
	5	0.1380	0.0920	0.1020	0.0930	0.0960	0.0850	0.1040	0.0890	0.0900	0.1110
	6	0.1300	0.0960	0.0780	0.0970	0.0990	0.0950	0.0900	0.0890	0.0990	0.1270
	7	0.0740	0.0780	0.0930	0.1150	0.0920	0.0830	0.0970	0.1140	0.1170	0.1370
	8	0.0390	0.0630	0.0740	0.0860	0.0860	0.1250	0.1100	0.1510	0.1290	0.1370
	9	0.0130	0.0290	0.0490	0.0840	0.0940	0.1280	0.1380	0.1470	0.1650	0.1530
	10	0.0030	0.0060	0.0170	0.0380	0.0590	0.1020	0.1450	0.1840	0.2280	0.2180

|| P^n - A || = 0.47155

SWEDEN:

$|\lambda_{2,SWE}| = 0.187$

USA:

$|\lambda_{2,USA}| = 0.435$

CONCLUSION:

$|\lambda_{2,SWE}| < |\lambda_{2,USA}|$

Income mobility is higher in Sweden than in the USA!

HIGHER DIMENSIONAL MATRICES & BOOTSTRAPPING

- Results on previous slide constructed using 10×10 matrices, i.e. *k = 10* groups. As stated earlier, better to use transition matrices with as many groups as possible. Table below shows the same measures of mobility calculated for *k=10, 20, 40 & 50*.

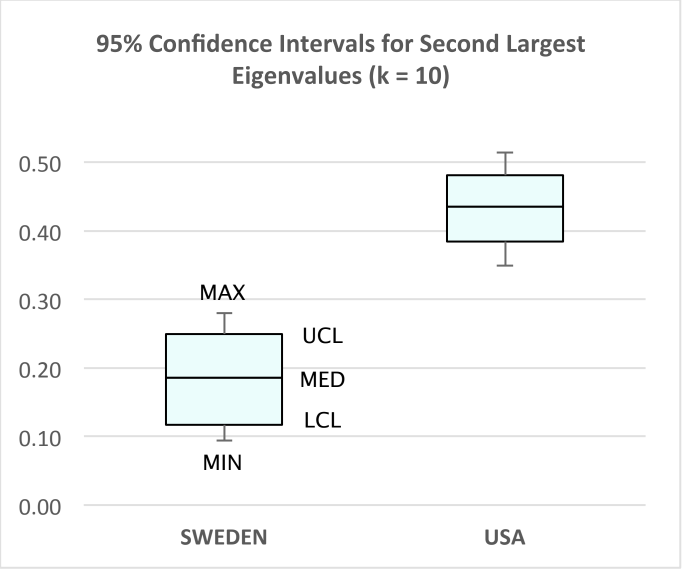
Cohort	$k \times k = 10 \times 10$			$k \times k = 20 \times 20$			$k \times k = 40 \times 40$			$k \times k = 50 \times 50$		
	$\|\lambda_2\|$	$n(\varepsilon)$	$T(\varepsilon)$	$\|\lambda_2\|$	$n(\varepsilon)$	$T(\varepsilon)$	$\|\lambda_2\|$	$n(\varepsilon)$	$T(\varepsilon)$	$\|\lambda_2\|$	$n(\varepsilon)$	$T(\varepsilon)$
SWEDEN	0.187	6	30	0.191	6	30	0.193	6	30	0.189	6	30
USA	0.435	12	60	0.436	12	60	0.435	12	60	0.435	12	60

Results for kxk transition matrices derived from cohorts in Sweden and USA.

- For all values of *k* investigated same result emerges, $|\lambda_{2,SWE}| < |\lambda_{2,USA}|$ such that income mobility is greater in Sweden than in USA.
- Note *k cannot* be taken arbitrarily large as resulting transition matrices may not be ergodic. One also runs into problems numerically calculating the eigenvalues of transition matrices when *k* is *very* large.

BOOTSTRAPPING:

- Also possible to create a 95% confidence interval for $|\lambda_2|$ by taking
  1. Taking 5,000 independent samples of 1,000 individuals from each cohort;
  2. Forming the resultant transition matrices;
  3. Calculating $|\lambda_{2,SWE}|$ and $|\lambda_{2,USA}|$;
  4. Collating results and removing the upper and lower 2.5th percentile.
- Produces a 95% confidence interval for our measure of mobility analogous to confidence interval for *β* produced in elasticity measures.



SECTION VI

CONCLUSION & BIBLIOGRAPHY

CONCLUSION

- There is a need to derive a summary measure of income mobility from transition matrices, particularly when elasticity methods are not appropriate.
- This measure is provided by calculating the magnitude of the second largest eigenvalue $|\lambda_2| \in [0,1)$ of the transition matrix.
- This measures how long it takes for the initial income advantages or disadvantages held by individuals dissipate.
- The closer $|\lambda_2|$ is to **0** the **greater** the level of income mobility. The closer $|\lambda_2|$ is to **1** the **lower** the level of income mobility.
- The measure is applicable to high dimensional matrices and confidence intervals for the measure can be produced using bootstrapping techniques.
- Explained technique in terms of income mobility, but applicable to any form of mobility.

Bibliography

1. *Gibbons M.* 2008. Intergenerational Economic Mobility in New Zealand. Labour, Employment and Work in New Zealand.

2. *Bhattacharya D. & Mazumder B.* 2011. A non-parametric analysis of black white differences in intergenerational mobility in the United States. Quantitative Economics, 2(3), 335-379.

3. *Seneta E.* 2006. Non-negative matrices and Markov chains. Springer Science & Business Media.