# Estimating Bilateral Migration

Guy J. Abel

- A common problem with bilateral migration data is that it is unavailable or outdated.
  - Often collected in censuses
- In some cases there are other data sources available that provide information on the in- and out-migration totals
  - Population registers

A. Auxi	liary D	ata			B. Primary Data: Marginal Totals							
Origin		De	estina	tion			Origin		De	estina	tion	
	Α	В	С	D	Sum			A	В	С	D	Sum
A		100	30	70	200		Α					250
В	50		45	5	100		В					75
C	60	35		40	135		C					125
D	20	25	20		65		D					150
Sum	130	160	95	115	500		Sum	150	200	50	200	600

- This provides a data estimation challenge, where the marginal tables totals are known but the cell values are known.
- Similar data estimation challenges exist for more detailed migration flow tables, for example:
  - In- and out-migration totals by age in each region are known, but the origin-destination migration flow table for each age group is missing.
  - Required by multi-regional cohort-component models
  - Estimating international migration flows from stocks (see for example Abel (2013))

- A popular approach to estimate values in a contingency table based on known marginal tables and an initial contingency table is the Iterative Proportional Fitting Procedure (IPFP).
- First described by Deming and Stephan (1940), the IPFP has since been widely studied in a number of different disciplines and under a number of synonyms such as raking, matrix scaling or the RAS algorithm
  - Lovelace et al. (2015) gives a good overview of the application of IPFP in social sciences.
  - Lomax and Norman (2016) for another overview more specific to demography.
- Mathematical approach to iteratively adjust a *seed* contingency table  $\mu_{ij}^{(0)} = m_{ij}$  to known row and column totals  $(n_{i+} \text{ and } n_{+j})$

$$\mu_{ij}^{(t+1)} = \frac{\mu_{ij}^{(t)}}{\mu_{i+}^{(t)}} n_{i+} \qquad \mu_{ij}^{(t+2)} = \frac{\mu_{ij}^{(t+1)}}{\mu_{+i}^{(t+1)}} n_{+j}$$

Α.	Auxii	liary	Data

Origin	Destination								
_	Α	В	С	D	Sum				
A		100	30	70	200				
В	50		45	5	100				
C	60	35		40	135				
D	20	25	20		65				
Sum	130	160	95	115	500				

A. Auxi	liary Dat	a					
Origin		Des	tination	1			Total
_	Α	В	С	D	Sum		In
A		100	30	70	200	1	30
В	50		45	5	100	1	60
C	60	35		40	135		95
D	20	25	20		65	1	15
Sum	130	160	95	115	500	5	00

A. Auxil	iary Dat	:a			
Origin		Des	tination	1	
_	Α	В	С	D	Sum
A		100	30	70	200
В	50		45	5	100
C	60	35		40	135
D	20	25	20		65
Sum	130	160	95	115	500

A. Auxil	liary Dat	a			
Origin		Des	tinatio	1	
_	Α	В	С	D	Sum
A		100	30	70	200
	50		45	5	100
C	60	35		40	135
D	20	25	20		65
Sum	130	160	95	115	500

В

Α

C

D

# **IPFP**

A B C D

 $\mathsf{Sum}$ 

A. Auxil	liary Dat	а						
Origin		Des	tination	1			Totals	
_	Α	В	С	D	Sum	In	Out	Net
A		100	30	70	200	130	200	-70
В	50		45	5	100	160	100	60
C	60	35		40	135	95	135	-40
D	20	25	20		65	115	65	50
Sum	130	160	95	115	500	500	500	0

Sum

ln

Out

Totals In Out Net

A A!	!: D-4									
A. Auxil	iary Dat	a								
Origin		Des	tination	1					Totals	
_	Α	В	С	D	Sum			In	Out	Net
Α		100	30	70	200			130	200	-70
В	50		45	5	100			160	100	60
C	60	35		40	135			95	135	-40
D	20	25	20		65			115	65	50
Sum	130	160	95	115	500			500	500	0
B. Cons	trained L	Estimates	s							
Origin		Des	tination	1		Targ	et		Totals	
	Α	В	С	D	Sum	In	Out	In	Out	Net
A B C						150 200 50	250 75 125			
D						200	150			
Sum						600	600			

In

Totals

Out

-70 

-40

Net

-70

-40

B. Constrained Estimates

Α

D

Sum

Origin

Α

В

Sum

В

Destination

C

IPFP							
A. Auxiliary L	Data						
Origin	Dest	ination		_		Totals	
			 	_			

A. Auxi	liary Data							
Origin		Desi	tination				Totals	
_	А	В	С	D	Sum	In	Out	Net

Sum

Target

lη

Out

Prigin		Des	tination		
	Α	В	С	D	Sum
١		100	30	70	200
3	50		45	5	100

D

Out

250.00

75.00

125.00

150.00

600.00

In

139.21

215.10

117.40

128.29

600.00

Net

-110.79

140.10

-7.60

0.00

-21.71

B. Constrained Estimates

Α

37.50

55.56

46.15

139.21

Destination

37.50

33.75

46.15

117.40

В

125.00

32.41

57.69

215.10

C

D

87.50

37.04

128.29

3.75

Origin

Α

В

Sum

IPFP		
A. Auxiliary D	ata	
Origin	Destination	Totals
·		 

A. Auxilia	ary Dat	a			
Origin		Des	tination		
	Α	В	С	D	Sum
		100			

rigin		Des	stination					iotais
	Α	В	С	D	Sum		In	Out
		100	30	70	200		130	200
3	50		45	5	100		160	100
·	60	35		40	135		95	135

Α		100	30	70	200	130	200	-70
В	50		45	5	100	160	100	60
C	60	35		40	135	95	135	-40
D	20	25	20		65	115	65	50
Sum	130	160	05	115	500	500	500	Λ

	50		73	3	100		00 100	, ,
C	60	35		40	135		95 135	-40
D	20	25	20		65	1	15 65	5 50
Sum	130	160	95	115	500	5	00 500	0

Sum

250.00

125.00

150.00

600.00

75.00

**Target** 

lη

150

200

200

600

50

Out

250

125

150

600

268.61 -118.61

139.37

-97.74

76.97

0.00

60.63

147.74

123.03

600.00

### **IPFP**

Α

В

D

Sum

116.23

30.13

53.64

200.00

40.41

59.86

49.73

150.00

15.97

14.37

19.66

50.00

136.41

200.00

5.85

57.74 147.74

A. Auxilia	ary Data	1							
Origin		Des	tination					Totals	
_	Α	В	С	D	Sum		ln	Out	Net

A. Auxilia	ary Dat	а					
Origin	Destination						
	Α	В	С	D	Sum		
4		100	30	70	200		

	А	В	C	D	Sum
Α		100	30	70	200
В	50		45	5	100
C	60	35		40	135
D	20	25	20		65
Sum	120	160	0E	115	500

В	50		45	5	100	160	100	60
C	60	35		40	135	95	135	-40
D	20	25	20		65	115	65	50
Sum	130	160	95	115	500	500	500	0

C	60	35		40	135	95	135	-40
D	20	25	20		65	115	65	50
Sum	130	160	95	115	500	500	500	0

B. Const.	rained E	stimates	5							
Origin		Des	tination			Targ	et		Totals	
	A	В	С	D	Sum	In	Out	In	Out	Net

B. Const	trained E	stimates	5							
Origin		Des	tination			Targ	et		Totals	
_	Α	В	С	D	Sum	In	Out	In	Out	Net

268.61

60.63

123.03

600.00

150

200

200

600

50

250

125

150

600

75

150.00

200.00

200.00

600.00

Out

250.00

125.00

150.00

600.00

75.00

In

161.27

199.07

56.61

183.05

600.00

Net

-88.73

124.07

-68.39

33.05

0.00

B. Constrained Estimates

Α

49.99

50.65

60.63

161.27

Destination

14.86

17.78

23.96

56.61

В

108.17

25.50

65.40

199.07

C

D

126.96

7.23

48.86

183.05

### IDED

Origin

Α

В

Sum

IPFP								
A. Auxilia	ary Data							
Origin		Dest	tination				Totals	
	Α	В	С	D	Sum	In	Out	Net

A. Auxiii	ary Dat	a			
Origin		Des	tination		
_	Α	В	С	D	Sum
A		100	30	70	200
D	Ε0		4.5		100

	, ,	_	•	_	• • • • • • • • • • • • • • • • • • • •	•••	0 4.0
		100	30	70	200	130	200
	50		45	5	100	160	100
	60	35		40	135	95	135
	20	25	20		65	115	65
1	130	160	95	115	500	500	500

$\overline{}$		100	30	70	200	130	200	-10
В	50		45	5	100	160	100	60
C	60	35		40	135	95	135	-40
D	20	25	20		65	115	65	50
Sum	130	160	95	115	500	500	500	0

В	50		45	5	100	100	100	00
C	60	35		40	135	95	135	-40
D	20	25	20		65	115	65	50
Sum	130	160	95	115	500	500	500	0

Sum

250.00

125.00

150.00

600.00

75.00

**Target** 

lη

150

200

200

600

50

Out

250

125

150

600

Out

200

100

135

500

Totals

Out

70.10

126.10

143.27

600.00

260.53 -110.53

65

Net

-70

60

-40

50

0

Net

129.90

-76.10

56.73

0.00

In

130

160

95

115

500

In

150.00

200.00

50.00

200.00

600.00

### IPFP

Α

B C

D

Sum

Origin

Α

В

D

Sum

Ī				
•	A. Auxiliary Data			

A. Auxiliary Data	
Origin	Destination

C

30

45

20

95

Destination

13.13

15.70

21.17

50.00

C

D

70

5

40

115

D

138.72

7.90

53.38

200.00

Sum

200

100

135

500

Sum

260.53

70.10

126.10

143.27

600.00

65

**Target** 

lη

150

200

200

600

50

Out

250

125

150

600

75

В

100

35

25

160

В

108.68

25.61

65.71

200.00

Α

50

60

20

130

B. Constrained Estimates

Α

46.49

47.11

56.40

500

Totals

Out

250.00

125.00

150.00

600.00

75.00

65

115

500

In

155.49

198.47

51.56

194.48

600.00

Net -70

60

-40

50

0

Net

-94.51

123.47

-73.44

44.48

0.00

20

130

B. Constrained Estimates

Α

49.74

46.70

59.05

155.49

25

160

В

104.29

25.39

68.79

198.47

20

95

Destination

12.60

16.80

22.16

51.56

C

D

Sum

Origin

Α

В

D

Sum

IPFP		
A. Auxiliary D	)ata	
Origin	Destination	Totals

65

500

Sum

250.00

125.00

150.00

600.00

75.00

**Target** 

lη

150

200

200

600

50

Out

250

125

150

600

75

Origin		Des	tination		
_	А	В	С	D	Sum
A		100	30	70	200
В	50		45	5	100

				-		
_	Α	В	С	D	Sum	In
Α		100	30	70	200	130
В	50		45	5	100	160
C	60	35		40	135	95

115

D

133.11

8.45

52.91

In

150.00

200.00

50.00

200.00

600.00

500

Totals

Out

254.20

72.98

125.05

147.78

600.00

0

Net

-104.20

127.02

-75.05

52.22

0.00

### **IPFP**

Sum

Origin

Α

В

Sum

A. Auxili	ary Dat	:a			
Origin		Des	stination		
_	Α	В	С	D	Sum
Α		100	30	70	200
В	50		45	5	100
C	60	35		40	135
D	20	25	20		65

500

Sum

254.20

72.98

125.05

147.78

600.00

**Target** 

lη

150

200

200

600

50

Out

250

125

150

600

75

# B. Constrained Estimates

Α

47.99

45.05

56.96

150.00

130

В

105.09

25.59

69.32

200.00

95

Destination

12.22

16.29

21.49

50.00

C

115

D

136.89

8.69

54.41

200.00

115

500

In

152.17

199.30

50.57

197.96

600.00

**Target** 

lη

150

200

200

600

50

Out

250

125

150

600

75

135

500

Totals

Out

250.00

125.00

150.00

600.00

75.00

65

-40

50

0

Net

-97.83

124.30

-74.43

47.96

0.00

### **IPFP**

Origin

Α

В

D

Sum

A. Auxili	ary Dat	а						
Origin		Des	tination				Totals	
_	Α	В	С	D	Sum	In	Out	Net
A		100	30	70	200	130	200	-70
В	50		45	5	100	160	100	60

Α		100	30	70	20
В	50		45	5	10
C	60	35		40	13
D	20	25	20		6
Sum	130	160	95	115	50

В

103.35

25.58

70.37

199.30

Destination

12.01

16.75

21.81

50.57

C

B. Constrained Estimates

Α

49.32

45.03

57.82

152.17

45	5	100	
	40	135	
20		65	
95	115	500	

D

134.63

8.94

54.39

197.96

Sum

250.00

125.00

150.00

600.00

-40 50 0

Net

125.80

-75.01

50.82

0.00

Totals

Out

74.20

125.01

149.18

600.00

251.62 -101.62

In

150.00

200.00

50.00

200.00

600.00

B. Constrained Estimates

Α

48.62

44.39

56.99

150.00

### **IPFP**

Origin

Α

В

Sum

A. Auxili	iary Dat	а						
Origin		Des	tination				Totals	
	Α	В	С	D	Sum	In	Out	Net
A		100	30	70	200	130	200	-70
В	50		45	5	100	160	100	60

**Target** 

lη

150

200

200

600

50

Out

250

125

150

600

75

D	50		73	J	100	100	100
C	60	35		40	135	95	135
D	20	25	20		65	115	65
Sum	130	160	95	115	500	500	500

Sum

251.62

125.01

149.18

600.00

74.20

_	00	33		+0	133	93	Τ,
D	20	25	20		65	115	(
Sum	130	160	95	115	500	500	50

D

136.02

9.03

54.95

200.00

Destination

11.88

16.56

21.57

50.00

В

103.72

25.67

70.62

200.00

C

500

In

199.22 150.00

150.84

199.72

50.22

600.00

65

500

Totals

Out

250.00

125.00

600.00

75.00

-70 60 -40

50

Net

-99.16

124.72

-74.78

49.22

0.00

0

20

130

B. Constrained Estimates

Α

49.14

44.39

57.31

150.84

25

160

В

103.05

25.66

71.01

199.72

D

Sum

Origin

Α

В

D

Sum

IPFP		
A. Auxiliary D	ata	
Origin	Destination	Totals

A. Auxi	iliary Data								
Origin		Desi	tination					Totals	
_	Α	В	С	D	Sum	-	In	Out	Net

65

**Target** 

lη

150

200

200

600

50

Out

250

125

150

600

75

500

Sum

250.00

75.00

125.00

150.00

600.00

Origin		Destination							
_	Α	В	С	D	Sum				
A		100	30	70	200				
В	50		45	5	100				

115

D

135.15

9.13

54.95

199.22

20

95

Destination

11.80

16.73

21.69

50.22

C

_							
	Α	В	С	D	Sum	In	(
Α		100	30	70	200	130	
В	50		45	5	100	160	
C	60	35		40	135	95	

B. Constrained Estimates

Α

48.87

44.14

56.99

150.00

160

В

103.20

25.70

71.10

200.00

### **IPFP**

Sum

Origin

Α

В

D

Sum

A. Auxilia	ary Dat	а							
Origin		Des	tination			-		Totals	
_	Α	В	С	D	Sum	-	In	Out	Net
A		100	30	70	200		130	200	-70

500

Sum

250.62

125.00

149.69

600.00

74.69

	Α	ь	C	D	Juili
А		100	30	70	200
В	50		45	5	100
C	60	35		40	135
D	20	25	20		65

115

D

135.67

9.16

55.16

200.00

95

Destination

11.75

16.66

21.59

50.00

C

		_

**Target** 

lη

150

200

200

600

50

Out

250

125

150

600

75

160

95

115

500

In

150.00

200.00

50.00

200.00

600.00

100

135

500

Totals

Out

250.62

74.69

125.00

149.69

600.00

65

60

-40

50

0

Net

-100.62

125.31

-75.00

50.31

Out

250.00

125.00

150.00

600.00

75.00

In

150.32

199.89

199.70

600.00

50.09

Net

-99.68

124.89

-74.91

49.70

0.00

B. Constrained Estimates

Α

49.07

44.14

57.11

150.32

Destination

11.72

16.73

21.63

50.09

C

D

135.34

9.20

55.16

199.70

В

102.94

25.70

71.25

199.89

### IDED

Origin

Α

В

Sum

IPFP									
A. Auxilia	ary Data								
Origin		Des	tination					Totals	
	Λ	R		D	Sum		In	Out	Not

Origin		Des	tination			
	Α	В	С	D	Sum	In
		100	30	70	200	130
В	50		45	5	100	160

	А	Ь	C	D	Sum
Α		100	30	70	200
В	50		45	5	100
C	60	35		40	135
D	20	25	20		65

$\overline{}$		100	30	70	200	130	200	-10
В	50		45	5	100	160	100	60
C	60	35		40	135	95	135	-40
D	20	25	20		65	115	65	50
Sum	130	160	95	115	500	500	500	0

C	60	35		40	135	95	135	-40
D	20	25	20		65	115	65	50
Sum	130	160	95	115	500	500	500	0

Sum

250.00

125.00

150.00

600.00

75.00

**Target** 

lη

150

200

200

600

50

Out

250

125

150

600

Sum

Α

В

D

Sum

A. Auxilia	ary Dat	a						
Origin	Destination						Totals	
	Α	В	С	D	Sum	In	Out	Net
A		100	30	70	200	130	200	-70

500

Sum

250.24

125.00

149.88

600.00

74.88

	Α	В	C	D	Sum
Α		100	30	70	200
В	50		45	5	100
C	60	35		40	135
D	20	25	20		65

115

D

135.54

9.21

55.25

200.00

95

C

11.70

16.70

21.60

50.00

**Target** 

lη

150

200

200

600

50

Out

250

125

150

600

75

160

95

115

500

In

150.00

200.00

50.00

200.00

600.00

100

135

65

500

Totals

Out

74.88

125.00

149.88

600.00

250.24 -100.24

60

-40

50

0

Net

125.12

-75.00

50.12

0.00

B. Constrained E	stimates
Origin	Destination

Α

48.97

44.04

56.99

150.00

130

160

В

103.00

25.71

71.29

95

115

500

In

150.12

199.96

50.03

199.88

600.00

100

135

500

Totals

Out

250.00

125.00

150.00

600.00

75.00

65

60

-40

50

0

Net

-99.88

124.96

-74.97

49.88

0.00

50

60

20

130

B. Constrained Estimates

Α

49.04

44.04

57.04

150.12

35

25

160

В

102.90

25.71

71.35

199.96

45

20

95

Destination

11.69

16.73

21.62

50.03

C

В

D

Sum

Origin

Α

В

D

Sum

IPFP	
A. Auxiliary Data	

A. Auxilia	ary Data								
Origin	Destination							Totals	
_	Α	В	С	D	Sum		In	Out	N

A. Auxiliary Data								
Origin		Des	tination					
_	Α	В	С	D	Sum			
		100	30	70	200			

100

135

500

Sum

250.00

125.00

150.00

600.00

75.00

65

**Target** 

lη

150

200

200

600

50

Out

250

125

150

600

75

5

40

115

D

135.41

9.23

55.24

Out

250.09

125.00

149.95

600.00

74.95

In

150.00

200.00

50.00

200.00

600.00

Net

-100.09

125.05

-75.00

50.05

0.00

B. Constrained Estimates

Α

49.00

44.01

56.99

150.00

### **IPFP**

Origin

Α

В

Sum

A. Auxilia	ary Data	)							
Origin		Des	tination					Totals	
	Α	В	С	D	Sum		In	Out	Net

	•				
Origin		Des	tination		
	Α	В	С	D	Sum
		100	30	70	200
	50		45	5	100

Α		100	30	70	200	130
В	50		45	5	100	160
C	60	35		40	135	9!
D	20	25	20		65	11!
Sum	130	160	95	115	500	500

D

135.49

9.23

55.28

200.00

Destination

11.68

16.72

21.60

50.00

В

102.92

25.72

71.36

200.00

C

		100	50	70	200	130	200	-10
В	50		45	5	100	160	100	60
C	60	35		40	135	95	135	-40
D	20	25	20		65	115	65	50
Sum	130	160	95	115	500	500	500	(

C	60	35		40	135	95	135	-40
D	20	25	20		65	115	65	50
Sum	130	160	95	115	500	500	500	0

Sum

250.09

125.00

149.95

600.00

74.95

**Target** 

lη

150

200

200

600

50

Out

250

125

150

600

В	50		45	5	100	160	100	60
C	60	35		40	135	95	135	-40
D	20	25	20		65	115	65	50
Sum	130	160	95	115	500	500	500	0

Out

250.00

125.00

150.00

600.00

75.00

In

150.05

199.98

50.01

199.96

600.00

Net

-99.95

124.98

-74.99

49.96

0.00

B. Constrained Estimates

Α

49.03

44.01

57.01

150.05

Destination

11.68

16.73

21.61

50.01

C

D

135.44

9.24

55.28

199.96

В

102.88

25.72

71.39

199.98

Origin

Α

В

Sum

IPFP		
A. Auxiliary D	)ata	
Origin	Destination	 Totals

A. Aux	iliary Data								
Origin		Des	tination					Totals	
-	Α	В	С	D	Sum		In	Out	Net

Irigin		Des	tination			
_	Α	В	С	D	Sum	In
4		100	30	70	200	130
2	FΩ		45	F	100	160

	А	В	C	D	Sum	ın
		100	30	70	200	130
	50		45			160
	60	35		40	135	95
-	20	25	20		65	115

Α		100	30	70	200	130	200	-70
В	50		45	5	100	160	100	60
C	60	35		40	135	95	135	-40
D	20	25	20		65	115	65	50
Sum	130	160	95	115	500	500	500	0

C	60	35		40	135	95	135	-40
D	20	25	20		65	115	65	50
Sum	130	160	95	115	500	500	500	0

Sum

250.00

125.00

150.00

600.00

75.00

**Target** 

lη

150

200

200

600

50

Out

250

125

150

600

500

150.00

200.00

50.00

200.00

600.00

65

500

250.04 -100.04

74.98

125.00

149.98

600.00

50

Net

125.02

-75.00

50.02

0.00

0

### **IPFP**

D

Α

В

C

D

Sum

Sum

A. Auxil	liary Data	a			
Origin		Des	stination		
_	А	В	С	D	Sum
Α		100	30	70	200
В	50		45	5	100
C	60	35		40	135

65

500

R	Constrained	Ectimat

49.02

43.99

56.99

150.00

95 115

135.47

9.24

55.29

200.00

20

11.67

16.72

21.60

50.00

20

130

25

160

102.89

25.72

71.39

200.00

B. Con:	strained E	stimates	5						
Origin		Des	tination			Targ	get		Totals
	Α	В	С	D	Sum	In	Out	In	Out

250.04

125.00

149.98

600.00

74.98

150

200

200

600

50

250

125

150

600

95

115

500

In

150.02

199.99

50.00

199.98

600.00

100

135

500

Totals

Out

250.00

125.00

150.00

600.00

75.00

65

Net -70

60

-40

50

0

Net

-99.98

124.99

-75.00

49.98

0.00

20

130

B. Constrained Estimates

Α

49.03

43.99

56.99

150.02

25

160

В

102.87

25.72

71.40

199.99

### IDED

D

Sum

Origin

Α

В

D

Sum

IPFP		
A. Auxiliary Da	ata	
Origin	Destination	Totals

65

**Target** 

lη

150

200

200

600

50

Out

250

125

150

600

75

500

Sum

250.00

125.00

150.00

600.00

75.00

Origin		Des	tination				Totals	
	Α	В	С	D	Sum	In	Out	
A		100	30	70	200	130	200	

08		200			
	А	В	С	D	Sum
A		100	30	70	200
В	50		45	5	100
C	60	35		40	135

115

D

135.45

9.24

55.29

199.98

20

95

Destination

11.67

16.73

21.60

50.00

C

160

115

500

In

150.00

200.00

50.00

200.00

600.00

95

200

100

135

500

Totals

Out

250.01

125.00

149.99

600.00

74.99

65

-70

60

-40

50

0

Net

-100.01

125.01

-75.00

50.01

0.00

### **IPFP**

Sum

Origin

Α

В

D

Sum

130

B. Constrained Estimates

Α

49.02

43.99

56.99

150.00

160

В

102.88

25.72

71.40

200.00

A. Auxilia	ary Data							
Origin		Desi	tination				Totals	
	Α	В	С	D	Sum	In	Out	Net

500

Sum

250.01

125.00

149.99

600.00

74.99

**Target** 

lη

150

200

200

600

50

Out

250

125

150

600

75

Origin		Des	LIIIaLIOII		
_	А	В	С	D	Sum
А		100	30	70	200
В	50		45	5	100

	Α	В	С	D	Sum
Α		100	30	70	200
В	50		45	5	100
C	60	35		40	135
D	20	25	20		65

115

D

135.46

9.24

55.29

200.00

95

Destination

11.67

16.73

21.60

50.00

C

Out

200

100

135

65

500

Totals

Out

250.00

125.00

150.00

600.00

75.00

Net

-70

60

-40

50

0

Net

-99.99

125.00

-75.00

49.99

0.00

In

130

160

95

115

500

In

150.01

200.00

199.99

600.00

50.00

### **IPFP**

Α

В

D

Sum

Origin

Α

В

D

Sum

Α	Auxiliary	Da

Α.	Auxiliary	Dat

A. Auxiliary	Data
Origin	

Α

50

60

20

130

B. Constrained Estimates

Α

49.03

43.99

56.99

150.01

Destination

C

30

45

20

95

Destination

11.67

16.73

21.60

50.00

C

D

70

5

40

115

D

135.46

9.24

55.29

199.99

Sum

200

100

135

500

Sum

250.00

125.00

150.00

600.00

75.00

65

**Target** 

In

150

200

200

600

50

Out

250

125

150

600

75

В

100

35

25

160

В

102.87

25.72

71.41

135

500

250.01 -100.01

125.00

-75.00

50.00

0.00

75.00

125.00

150.00

600.00

65

60

-40

50

0

B. Constrained Estimates

49.03

43.99

56.99

150.00

102.87

25.72

71.41

200.00

11.67

16.73

21.60

50.00

### **IPFP**

Α

В

Sum

A. Auxilia	ary Dat	а							
Origin		Des	tination			-		Totals	
	Α	В	С	D	Sum	-	In	Out	Net
A		100	30	70	200	-	130	200	-70

	А	Ь	C	D	Sum
Α		100	30	70	200
В	50		45	5	100
C	60	35		40	135
D	20	25	20		65

135.46

200.00

9.24

55.29 125.00

Α		100	30	70	200	130
В	50		45	5	100	160
C	60	35		40	135	95
D	20	25	20		65	115
Sum	130	160	95	115	500	500

Origin		Des	tination			Targ	et		Totals	
	Α	В	С	D	Sum	ln	Out	ln	Out	Net

150

200

200

600

50

250

75

125

150

600

150.00

200.00

50.00

200.00

600.00

250.01

75.00

150.00

Out

250.00

75.00

125.00

150.00

600.00

In

150.00

200.00

50.00

200.00

600.00

Net

-100.00

125.00

-75.00

50.00

0.00

B. Constrained Estimates

Α

49.03

43.99

56.99

150.00

### **IPFP**

Origin

Α

В

Sum

A. Auxilia	ary Dat	а							
Origin		Des	stination			•		Totals	
	Α	В	С	D	Sum		 In	Out	Net
Λ		100	20	70	200	•	 120	200	70

Α		100	30	70	200
В	50		45	5	100
C	60	35		40	135
D	20	25	20		65
Sum	130	160	95	115	500

Destination

11.67

16.73

21.60

50.00

C

D

135.46

9.24

55.29

200.00

Sum

250.00

125.00

150.00

600.00

75.00

В

102.87

25.72

71.41

200.00

		Iotals	
Sum	In	Out	Net
200	130	200	-70
100	160	100	60
135	95	135	-40
C =		C =	

73	5	100	100	100	00
	40	135	95	135	-40
20		65	115	65	50
95	115	500	500	500	0

**Target** 

lη

150

200

200

600

50

Out

250

125

150

600

95

115

500

In

150.00

200.00

50.00

200.00

600.00

100

135

500

Totals

Out

250.00

75.00

125.00

150.00

600.00

65

60

-40

50

0

Net

-100.00

125.00

-75.00

50.00

0.00

20

130

B. Constrained Estimates

Α

49.03

43.98

56.99

150.00

25

160

В

102.87

25.72

71.41

200.00

D

Sum

Origin

Α

В

D

Sum

IPFP		
A. Auxiliary D	ata	
Origin	Destination	Totals

A. Auxiii	ary Dat	a			
Origin		Des	tination		
	Α	В	С	D	Sum
A		100	30	70	200

65

**Target** 

lη

150

200

200

600

50

Out

250

125

150

600

75

500

Sum

250.00

125.00

150.00

600.00

75.00

_					
	Α	В	С	D	Sum
Α		100	30	70	200
В	50		45	5	100
C	60	35		40	135

115

D

135.46

9.25

55.29

200.00

20

95

Destination

11.67

16.73

21.60

50.00

C

95

115

500

In

150.00

200.00

50.00

200.00

600.00

100

135

500

Totals

Out

250.00

75.00

125.00

150.00

600.00

65

60

-40

50

0

Net

-100.00

125.00

-75.00

50.00

0.00

20

130

B. Constrained Estimates

Α

49.03

43.98

56.99

150.00

25

160

В

102.87

25.72

71.41

200.00

D

Sum

Origin

Α

В

D

Sum

IPFP		
A. Auxiliary D	ata	
Origin	Destination	Totals

A. Auxiii	ary Dat	a			
Origin		Des	tination		
	Α	В	С	D	Sum
A		100	30	70	200

65

**Target** 

lη

150

200

200

600

50

Out

250

125

150

600

75

500

Sum

250.00

125.00

150.00

600.00

75.00

_					
	Α	В	С	D	Sum
Α		100	30	70	200
В	50		45	5	100
C	60	35		40	135

115

D

135.46

9.25

55.29

200.00

20

95

Destination

11.67

16.73

21.60

50.00

C

Out

250.00

75.00

125.00

150.00

600.00

In

150.00

200.00

50.00

200.00

600.00

Net

-100.00

125.00

-75.00

50.00

0.00

### **IPFP**

Origin

Α

В

Sum

A. Auxilia	ary Dat	·a							
Origin		Des	tination			-		Totals	
	Α	В	С	D	Sum	-	In	Out	Net
Δ		100	30	70	200	•	130	200	-70

Origin		Des	tination	1			Totals		
_	Α	В	С	D	Sum	-	ln	Out	Net
A		100	30	70	200	-	130	200	-70
В	50		45	5	100		160	100	60
C	60	35		40	135		95	135	-40
D	20	25	20		65		115	65	50
Sum	130	160	95	115	500		500	500	0

Sum

250.00

75.00

125.00

150.00

600.00

**Target** 

lη

150

200

200

600

50

Out

250

125

150

600

75

Sum	130	100
B. Consti	rained	Estimates

Α

49.03

43.98

56.99

150.00

Destination

11.67

16.73

21.60

50.00

C

D

135.46

9.25

55.30

200.00

В

102.87

25.72

71.41

- Willekens (1999) calls the seed data an *auxiliary* table and notes that it should be information on a variables related to migration.
  - Typically past migration flow data
  - Distances or travel costs between the origin-destination pairs have been used where no past data exists
  - Limited testing to see which seeds work best for estimating migration
- The marginal data is then known as primary data.
  - Partial observations on the number of migrations

## nipfp

- The mipfp package by Barthélemy and Suesse (2018) implements IPFP in R using the Ipfp() function
- Can be used for multi-dimensional marginal constraint problems.
- Three inputs
  - seed a matrix of auxiliary data to aid estimation
  - target.list a list of dimensions that are being targeted (see next point)
  - target.data a list of targets related to target.list
- R numbers dimension of arrays with
  - 1 row
  - 2 column
  - 3 table
  - •
- The target.list might involve
  - a single target, e.g. column totals target.list = list(2)
  - multiple targets, e.g. row and column totals target.list = list(1, 2)
  - sums over cells rather than margins of array, e.g. cells summed over tables target.list = list(c(1, 2))

```
> r <- LETTERS[1:4]
> m0 \leftarrow matrix(data = c(0, 100, 30, 70,
                       50, 0, 45, 5,
                       60, 35, 0, 40,
                       20, 25, 20, 0),
              nrow = 4, ncol = 4, byrow = TRUE,
              dimnames = list(orig = r, dest = r))
> addmargins(m0)
    dest
       A B C D Sum
orig
  A 0 100 30 70 200
  В
      50 0 45 5 100
  C
      60 35 0 40 135
  D
      20 25 20 0 65
  Sum 130 160 95 115 500
```

```
> orig_tot <- c(250, 75, 125, 150)
> dest_tot <- c(150, 200, 50, 200)
> names(orig_tot ) <- names(dest_tot) <- r</pre>
>
> orig_tot
 A B C D
250 75 125 150
> dest_tot
 A B C D
150 200 50 200
>
> # check sums are equal
> sum(orig_tot)
[1] 600
> sum(dest_tot)
[1] 600
```

```
> library(mipfp)
> Ipfp(seed = m0, target.list = list(1, 2),
      target.data = list(orig_tot, dest_tot))
Call:
Ipfp(seed = m0, target.list = list(1, 2), target.data = list(orig_tot,
   dest tot))
Method: ipfp - convergence: TRUE
Estimates:
   dest.
orig A B
   A 0.00000 102.87046 11.67024 135.459297
  B 49.02778 0.00000 16.72686 9.245364
  C 43.98433 25.72033 0.00000 55.295339
  D 56.98789 71.40921 21.60290 0.000000
```

```
> # save the result
 y0 <- Ipfp(seed = m0, target.list = list(1, 2),
            target.data = list(orig_tot, dest_tot))
>
 # view with totals
> addmargins(y0$x.hat)
    dest
                                           D Sum
orig
  Α
      0.00000 102.87046 11.67024 135.459297 250
  В
      49.02778 0.00000 16.72686 9.245364 75
  С
      43.98433 25.72033 0.00000 55.295339 125
  D
       56.98789 71.40921 21.60290 0.000000 150
  Sum 150.00000 200.00000 50.00000 200.000000 600
```

## Three dimensions

Origin         Destination         Origin         Destination           A         B         C         D         Sum         A         B         C         D           A         80         10         55         145         A         A         B         C         D           B         30         20         0         50         B         B         C         C         B         C         D         T         C         D         D         T         D         T         D         T         D         T         T         D         T         D         T         T         D         T         D         T         D         T         D         T         D         T         D         T         D         T         D         T         D         T         D         T         D         T         D         T         D         T         D         T         D         D         T         D         D         D         T         D         D         D         D         D         D         D         D         D         D         D         D         D         D         D	Auxiliary Data - Low Skill					Primary	/ Data					
A 80 10 55 145 A B 30 20 0 50 B C 50 15 10 75 C	Origin		De	estina	tion		Origin		De	estination		
B 30 20 0 50 B C 50 15 10 75 C		Α	В	С	D	Sum		Α	В	С	D	
C 50 15 10 75 C	A		80	10	55	145	Α					
	В	30		20	0	50	В					
D 5 20 10 35 D	C	50	15		10	75	C					
	D	5	20	10		35	D					
Sum 85 115 40 65 305 Sum 150 200 50 200	Sum	85	115	40	65	305	Sum	150	200	50	200	

#### Auxiliary Data - High Skill

Origin	Destination									
	Α	В	С	D	Sum					
Α		20	20	15	55					
В	20		25	5	50					
C	10	20		30	60					
D	15	5	10		30					
Sum	45	45	55	50	195					

10 B

Α

# ... with 22 more rows

High

20

## IPFP More Complicated Data Situations

- The IPFP can be used for more complex data situations with more than two dimensions.
- Key to using the mipfp() function is setting the inputs for target.data.

```
> library(tidyverse)
 d <- expand_grid(orig = r, dest = r, skill = c("Low", "High")) %>%
    mutate(flow = c(0, 0, 80, 20, 10, 20, 55, 15, 30, 20, 0, 0, 20, 25, 0, 5, 50, 1)
>
> d
# A tibble: 32 x 4
   orig dest skill
                       flow
   <chr> <chr> <chr> <chr> <dbl>
 1 A
         Α
               Low
                          0
 2 A
         Α
               High
         В
               Low
                         80
         В
               High
                         20
 5 A
         С
               Low
                        10
         С
               High
                         20
         D
               Low
                         55
         D
                        15
               High
9 B
               Low
                         30
         Α
```

#### Estimating Detailed Bilateral Migration

```
> m1 <- xtabs(formula = flow ~ orig + dest + skill, data = d)
> m1
, , skill = High
   dest
orig A B C D
   A 0 20 20 15
  B 20 0 25 5
  C 10 20 0 30
  D 15 5 10 0
, , skill = Low
   dest
orig A B C D
  A 0 80 10 55
  B 30 0 20 0
  C 50 15 0 10
  D 5 20 10 0
```

```
> addmargins(m1)
, , skill = High
    dest
       A
          В
              C
                   D Sum
orig
  Α
       0
          20
              20
                  15
                      55
  В
      20
          0
              25
                  5
                      50
  С
      10
          20
              0
                  30
                      60
                  0
 D
       15
           5
              10
                      30
      45
          45
  Sum
              55
                  50 195
   skill = Low
    dest
orig
       Α
          В
              C
                   D Sum
                  55 145
  Α
       0
          80
              10
  В
       30
          0
              20
                  0
                      50
  С
          15
       50
              0
                  10
                      75
 D
          20
       5
              10
                  0
                      35
  Sum
      85 115
              40
                 65 305
, , skill = Sum
    dest
       Α
            В
               С
                    D Sum
orig
```

```
> addmargins(m1)[,,skill = "Sum"]
    dest
orig
       A B
            C
                D Sum
      0 100 30 70 200
 Α
 В
      50 0 45 5 100
 C
      60 35 0 40 135
 D
      20
         25 20
               0 65
 Sum 130 160 95 115 500
```

```
> y1 <- Ipfp(seed = m1, target.list = list(1, 2),
             target.data = list(orig_tot, dest_tot))
> addmargins(y1$x.hat)
, , skill = High
     dest.
                          В
                                                D
                                                          Sum
orig
               Α
        0.000000
                  20.574092
                              7.780161
  Α
                                        29.026992
                                                   57.381245
  В
       19.611112
                   0.000000
                              9.292698
                                         9.245364
                                                    38.149174
  C
       7.330722 14.697330
                              0.000000
                                        41.471504 63.499556
  D
       42.740914
                  14.281842
                             10.801451
                                         0.000000
                                                   67.824208
  Sum
       69.682749
                  49.553264
                             27.874310
                                        79.743860 226.854183
   skill = Low
     dest.
orig
               Α
                          В
                                                D
                                                          Sum
  Α
        0.000000
                  82,296369
                              3.890080 106.432305 192.618755
  В
       29.416668
                   0.000000
                              7.434158
                                         0.000000
                                                    36.850826
  C
       36.653611
                  11,022997
                              0.000000
                                        13.823835 61.500444
  D
       14.246971
                  57,127369
                             10.801451
                                         0.000000
                                                   82,175792
       80.317251 150.446736
                             22.125690 120.256140 373.145817
  Sum
   skill = Sum
```

## nipfp

```
> addmargins(y1$x.hat)[,,skill = "Sum"]
    dest
                         В
                                              D
                                                       Sum
orig
              Α
                            11.670241 135.459297 250.000000
  Α
       0.000000 102.870462
  В
      49.027781 0.000000
                            16.726856
                                       9.245364 75.000000
  C
      43.984334 25.720327
                             0.000000 55.295339 125.000000
  D
      56.987886 71.409211
                            21.602903
                                       0.000000 150.000000
  Sum 150.000000 200.000000 50.000000 200.000000 600.000000
```

```
> y1$x.hat %>%
    as.data.frame.table(responseName = "est") %>%
   as_tibble()
# A tibble: 32 x 4
   orig dest skill
                      est
   <fct> <fct> <fct> <dbl>
 1 A
        Α
              High
 2 B
         Α
              High 19.6
 3 C
        Α
              High 7.33
 4 D
        Α
              High 42.7
 5 A
        В
              High 20.6
 6 B
        В
              High
                    0
 7 C
        В
              High 14.7
 8 D
        В
              High 14.3
 9 A
         С
              High 7.78
10 B
         С
               High
                     9.29
# ... with 22 more rows
```

- Plane (1981) developed a proportional adjustment algorithm for estimating bilateral migration flows to match both
  - Constraints on the net migration of each region
  - Total sum of the bilateral migration flows
- Requires knowledge of
  - Past bilateral migration flows
  - Current (target) total migration flows (over whole system)
  - Current (target) net migration flows
  - Distance matrix to correspond
- No application of this method in R, although in migest package the cm\_net\_tot() function provides a similar set of estimates
  - Unable to incorporate distance matrix

```
> addmargins(m0)
     dest
orig
             В
                C
                    D Sum
  Α
          100 30
                   70 200
  В
       50
            0 45
                    5 100
  C
       60
           35
                0
                   40 135
  D
       20
           25 20
                      65
  Sum 130 160 95 115 500
>
  # observed net
  library(migest)
  sum turnover(m0)
# A tibble: 4 x 5
  region in_mig out_mig turn
                                  net
  <chr>>
          <dbl>
                   <dbl> <dbl> <dbl>
             130
                     200
                            330
                                  -70
 Α
2 B
             160
                     100
                            260
                                   60
3 C
              95
                     135
                            230
                                  -40
4 D
             115
                      65
                            180
                                    50
```

Estimate migration flows to match new net migration and grand total.

```
> y1 <- cm net tot(net tot = c(-100, 125, -75, 50), tot = 600,
                 m = m0, verbose = FALSE)
> addmargins(y1$n)
    dest.
orig
                                                 Sum
      0.00000 136.22513 32.93756 79.068944 248.23163
  Α
  В
      49.88761 0.00000 42.28296 4.833488
                                            97.00406
  C
      74.27815 50.62851
                         0.00000 47.977516 172.88418
      24.06590 35.15032 22.66377 0.000000 81.87999
  Sum 148.23166 222.00396 97.88429 131.879947 599.99986
> sum turnover(v1$n)
# A tibble: 4 x 5
  region in_mig out_mig turn
                               net
  <chr> <dbl> <dbl> <dbl> <dbl> <dbl>
      148. 248. 396. -100.
2 B
       222. 97.0 319.
                             125.
3 C
        97.9 173. 271. -75.0
      132.
4 D
             81.9 214. 50.0
```

- The requirement on the total sum of the bilateral flow for the algorithm is not realistic. (1981) method not widely adpoted
  - In many countries the overall number of migrant flows, that is demographically consistent with natural population change, is typically not known.
  - If the overall number of migrant flows is known, it is typically obtained from a comprehensive population register, and thus bilateral migration or total in- and out-migration flows already exist. If it is the later, can use IPFP approaches.
- In recent years I have been working on a method that constrains only to the net migration totals.
  - Unpublished, work in progress, use at own risk
  - Method is available in the cm\_net() function in the migest package
- Potential uses
  - Update bilateral migration flows from surveys or administrative data to match known demographic consistent net migration totals
  - Estimate bilateral migration flows from known net migration totals using non-migration data as a seed (if no migration flow data available)

```
> y2 <- cm_net(net_tot = c(-100, 125, -75, 50), m = m0, verbose = FALSE)
> addmargins(y2$n)
    dest.
orig
                                                Sum
 Α
    0.00000 124.97056 27.96585 71.121910 224.05832
 В
      40.00942 0.00000 33.56693 4.065067
                                           77.64142
 C
      64.36422 46.92119 0.00000 43.597199 154.88260
      19.68451 30.74980 18.34980 0.000000 68.78412
 Sum 124.05815 202.64155 79.88258 118.784175 525.36645
> sum turnover(y2$n)
# A tibble: 4 x 5
 region in_mig out_mig turn
                              net
 <chr> <dbl>
                <dbl> <dbl> <dbl>
      124.
                224. 348. -100.
1 A
2 B
   203. 77.6 280. 125.
3 C
       79.9 155.
                       235. -75.0
        119. 68.8 188. 50.0
4 D
```

# Exercise (ex7.R)

# 0. a) Load the KOSTAT2021. Rproj file.

```
Run the qetwd() below. It should print the directory where the
      KOSTAT2021. Rproj file is located.
getwd()
      b) Load the packages used in this exercise
library(tidyverse)
library(mipfp)
##
##
##
# 1. Run the code below to read in the bilateral data in uk census 2011 tidy.csv
    from the ONS 2011 British Census
cen11 <- read csv("./data/uk census 2011 tidy.csv")</pre>
cen11
# 2. Run the code below to read in the bilateral data in
     uk nhs hesa 2018.csv from the British administrative data (National
     Health Service patient records and Higher Education Statistics Authority)
nhs18 <- read csv("./data/uk nhs hesa 2018 tidy.csv")
nhs18
# 3. Run the code below to create data with abbreviated region names - to make
     it easier to view the matrices for each time period
     Note: the census data is more detailed (orig - dest - age - sex) than the
           administrative data (orig - dest)
cen11 <- cen11 %>%
  mutate(orig full = orig.
```

#### References I

- Abel, Guy J. 2013. "Estimating global migration flow tables using place of birth data." *Demographic Research* 28 (March): 505–46. https://doi.org/10.4054/DemRes.2013.28.18.
- Barthélemy, Johan, and Thomas Suesse. 2018. "<br/>b>mipfp</b>: An <i>R</i> Package for Multidimensional Array Fitting and Simulating Multivariate Bernoulli Distributions." Journal of Statistical Software 86 (Code Snippet 2). https://doi.org/10.18637/jss.v086.c02.
- Deming, W. Edwards, and Frederick F Stephan. 1940. "On a Least Squares Adjustment of a Sampled Frequency Table When the Expected Marginal Totals are Known." *The Annals of Mathematical Statistics* 11 (4): 427–44. https://doi.org/10.1214/aoms/1177731829.
- Lomax, Nik, and Paul Norman. 2016. "Estimating population attribute values in a table: 'Get me started in' iterative proportional fitting." Professional Geographer 68 (3): 451–61. https://doi.org/10.1080/00330124.2015.1099449.
- Lovelace, Robin, Mark Birkin, Dimitris Ballas, and Eveline van Leeuwen. 2015. "Evaluating the Performance of Iterative Proportional Fitting for Spatial Microsimulation: New Tests for an Established Technique." Journal of Artificial Societies and Social Simulation 18 (2): 1–15. https://doi.org/10.18564/jasss.2768.
- Plane, David A. 1981. "Estimation of Place-to-Place Migration Flows from Net Migration Totals: A Minimum Information Approach." *International Regional Science Review* 6 (1): 33–51. https://doi.org/10.1177/016001768100600103.
- Willekens, Frans. 1999. "Modeling approaches to the indirect estimation of migration flows: from entropy to EM." Mathematical Population Studies 7 (3): 239–78, 308. https://doi.org/10.1080/08898489909525459.