Colorado Conference on the Estimation of Migration
Boulder and Aspen Lodge, September 24-26, 2004

Inferring Migration Flows From Birthplace-Specific Population Stocks, the Case of Mexico

Virgilio Partida
National Population Council, Mexico

Four indirect methods were applied:

1. Estimates based on infant migration propensities
2. Estimates based on age-specific net migration
3. Estimates based on two successive distributions of population stocks
4. Estimates based on iterative biproporcional adjustment

Main assumption

All the four methods suppose that the age patterns of conditional survivorship proportions of migrants, in two successive censuses, are similar

Four Regions of Mexico

Estimates based on infant migration propensities

The assumption is that proportionality between any two age groups in two successive censuses exists:

$$\frac{S_{ij}^{t-10}(x+h)}{S_{ij}^{t-10}(x)}=\frac{S_{ij}^{t}(x+h)}{S_{ij}^{t}(x)}$$

and only the level varies. We use the age-to-infant (ATI) migration ratio $r_{ij}(x)$, that is to say, the conditional survivorship proportions of migrants at any age divided by the child conditional survivorship proportion:

$$r_{ij}^{t}(x,-5)=\frac{S_{ij}^{t}(x)}{S_{ij}^{t}(-5)} \qquad x=0,5,10,\ldots$$

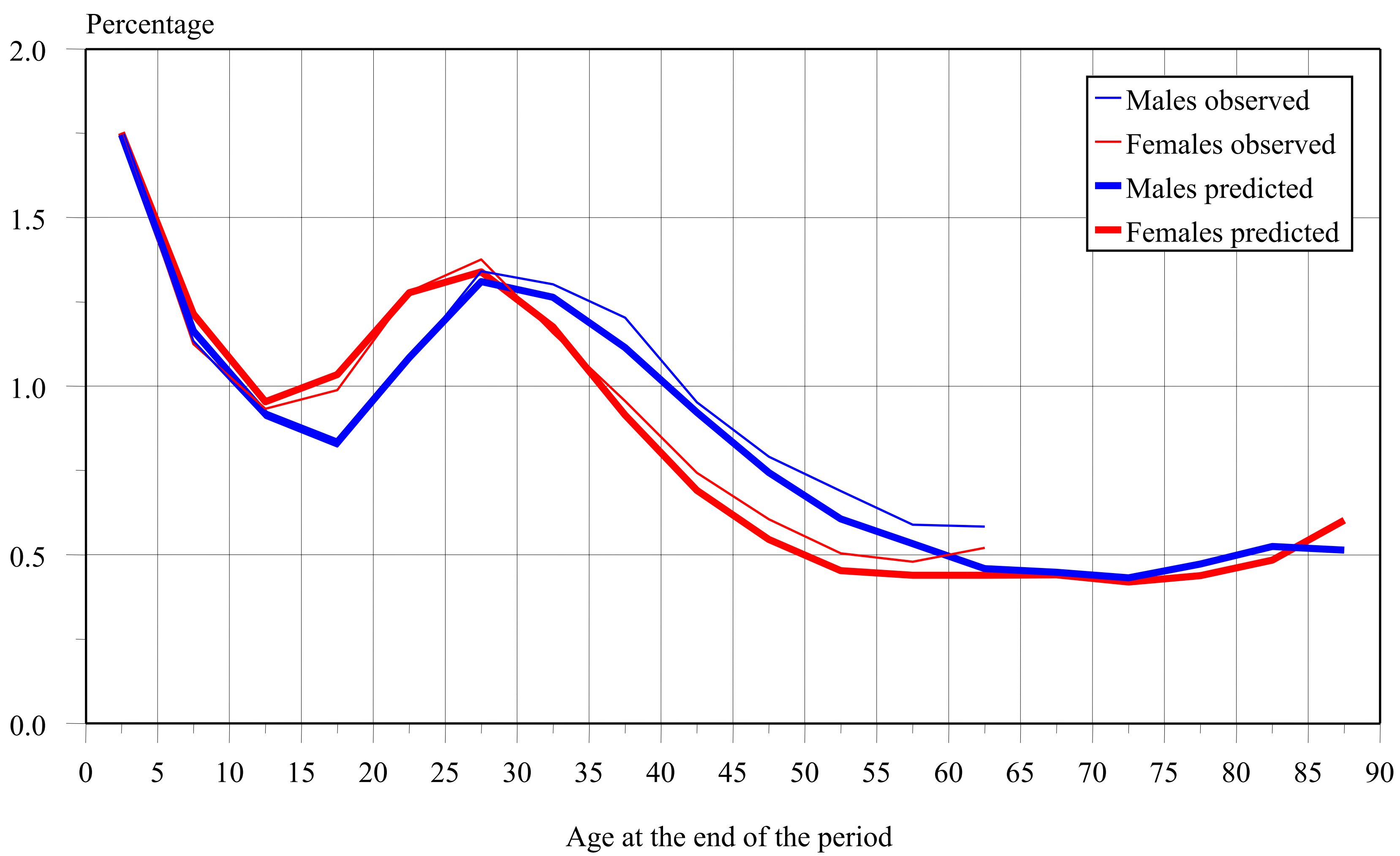
then, proportionality assumption can be written as:

$$\mathrm{S}_{\mathrm{ij}}^{\mathrm{t}}(x)=\mathrm{r}_{\mathrm{ij}}^{\mathrm{t\text{-}10}}(x,-5)\mathrm{S}_{\mathrm{ij}}^{\mathrm{t}}(-5) \quad \text{or} \quad \mathrm{S}_{\mathrm{ij}}^{\mathrm{t\text{-}10}}(x)=\mathrm{r}_{\mathrm{ij}}^{\mathrm{t}}(x,-5)\mathrm{S}_{\mathrm{ij}}^{\mathrm{t\text{-}10}}(-5)$$

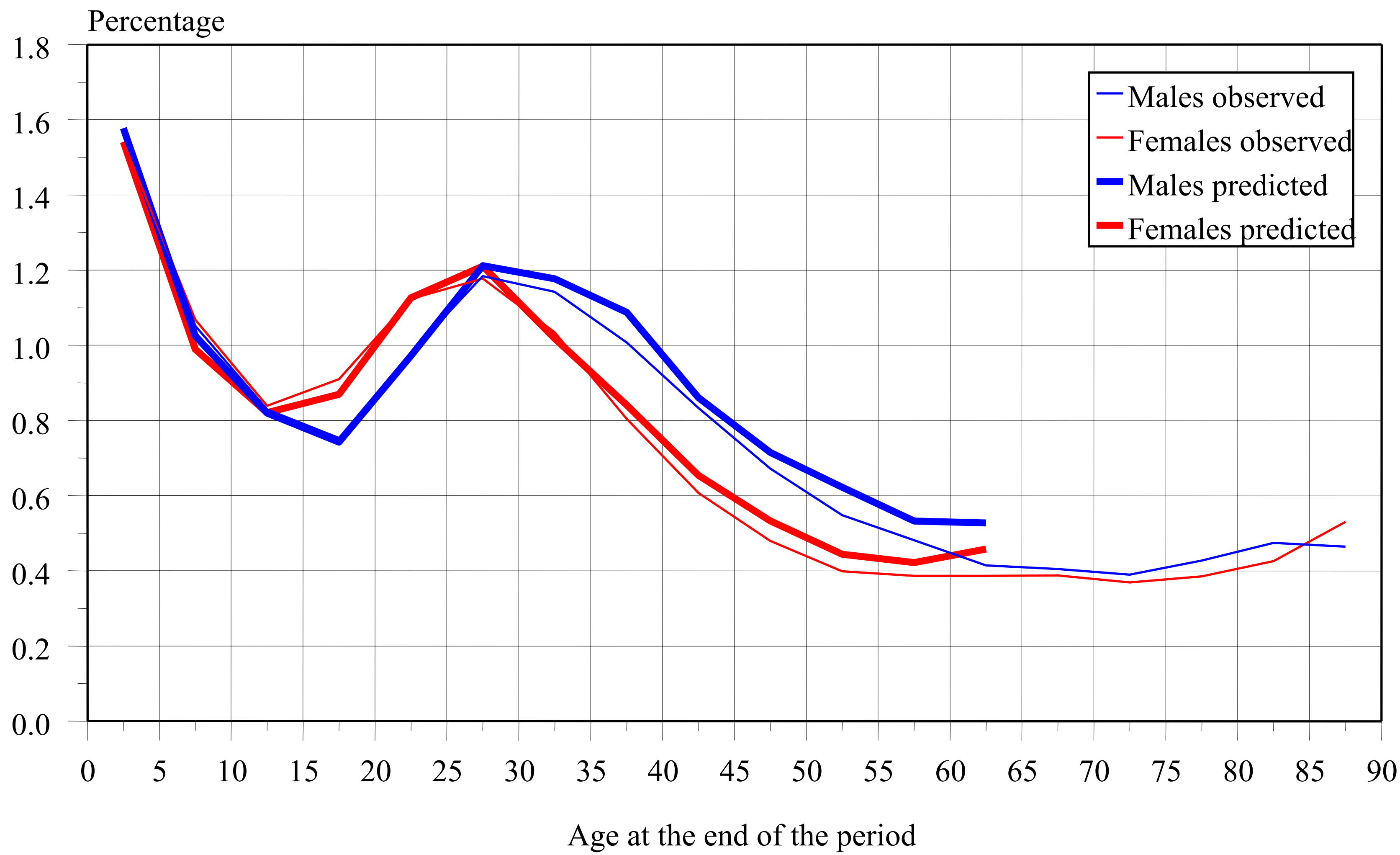
MAPE statistics as mesaure of goodness of fit for ATI procedure, 1985-2000

Region of destiny	Migration survivorship proportions, 1985-1990, using the 1995-2000 age-to-infant migration ratios Region of origin				Migration survivorship proportions, 1995-2000, using the 1985-1990 age-to-infant migration ratios Region of origin			
	Border	North Central	Central	South	Border	North Central	Central	South
				All				
Border		17.6	19.7	18.5		22.0	24.6	24.1
North Central	6.1		25.9	31.5	6.7		36.3	47.0
Central	10.6	12.0		20.5	12.5	14.6		26.4
South	10.9	15.8	9.2		11.8	19.9	10.3	
All	**16.5**				**21.4**			
				Males				
Border		17.2	19.2	16.9		21.3	23.2	21.7
North Central	5.9		23.7	30.5	6.7		32.6	44.9
Central	9.1	10.4		20.4	10.3	12.3		26.1
South	10.8	16.2	8.6		11.4	20.3	9.6	
All	**15.8**				**20.0**			
				Females				
Border		18.0	20.2	20.1		22.6	25.9	26.5
North Central	6.2		28.2	32.5	6.7		40.0	49.0
Central	12.2	13.6		20.6	14.8	16.9		26.7
South	11.0	15.4	9.8		12.3	19.5	11.0	
All	**17.3**				**22.7**			

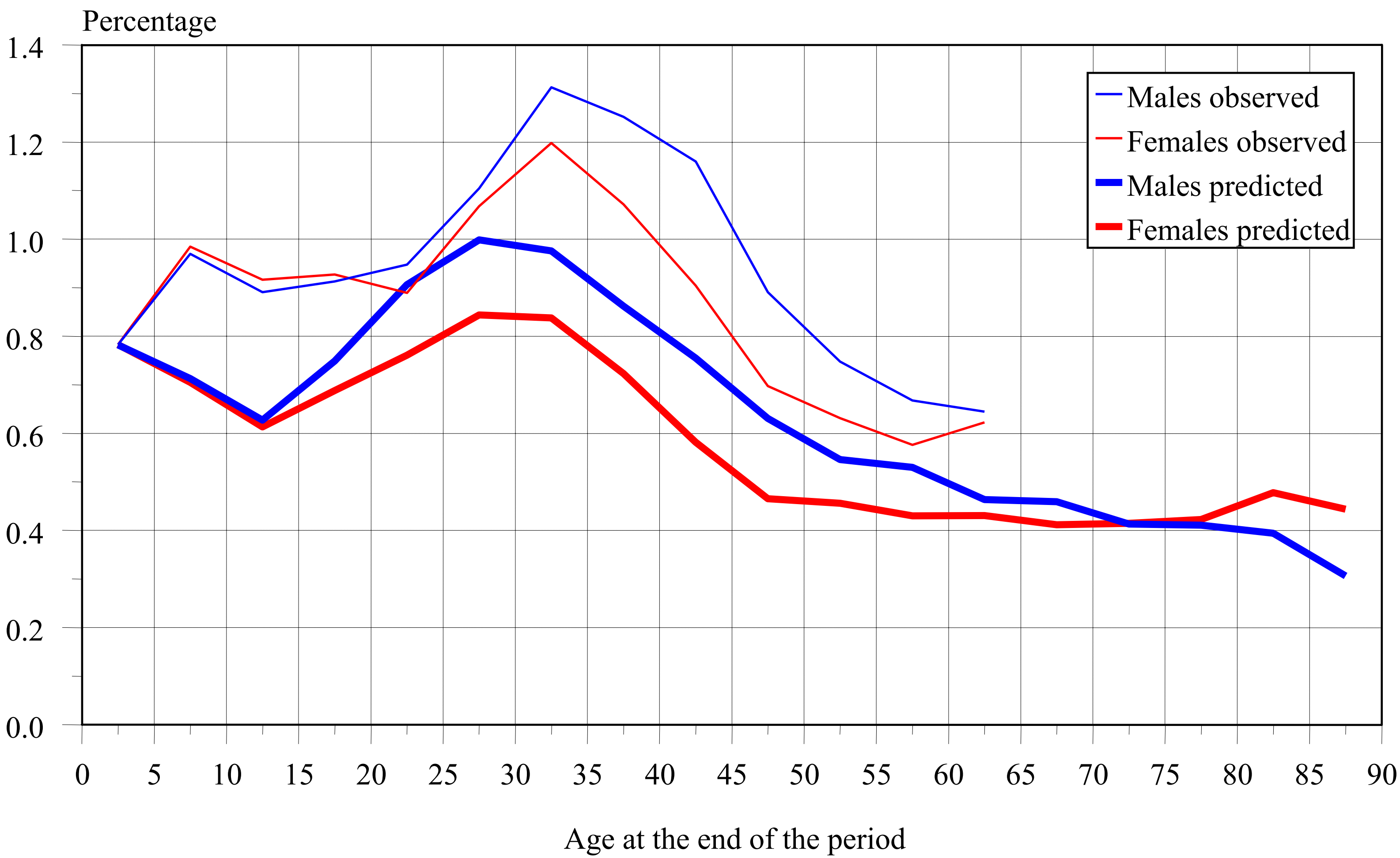
Observed and predicted conditional migration proportions from Border to North Central 1985-1990, using the 1995-2000 age-to-infant ratios (best fitting)



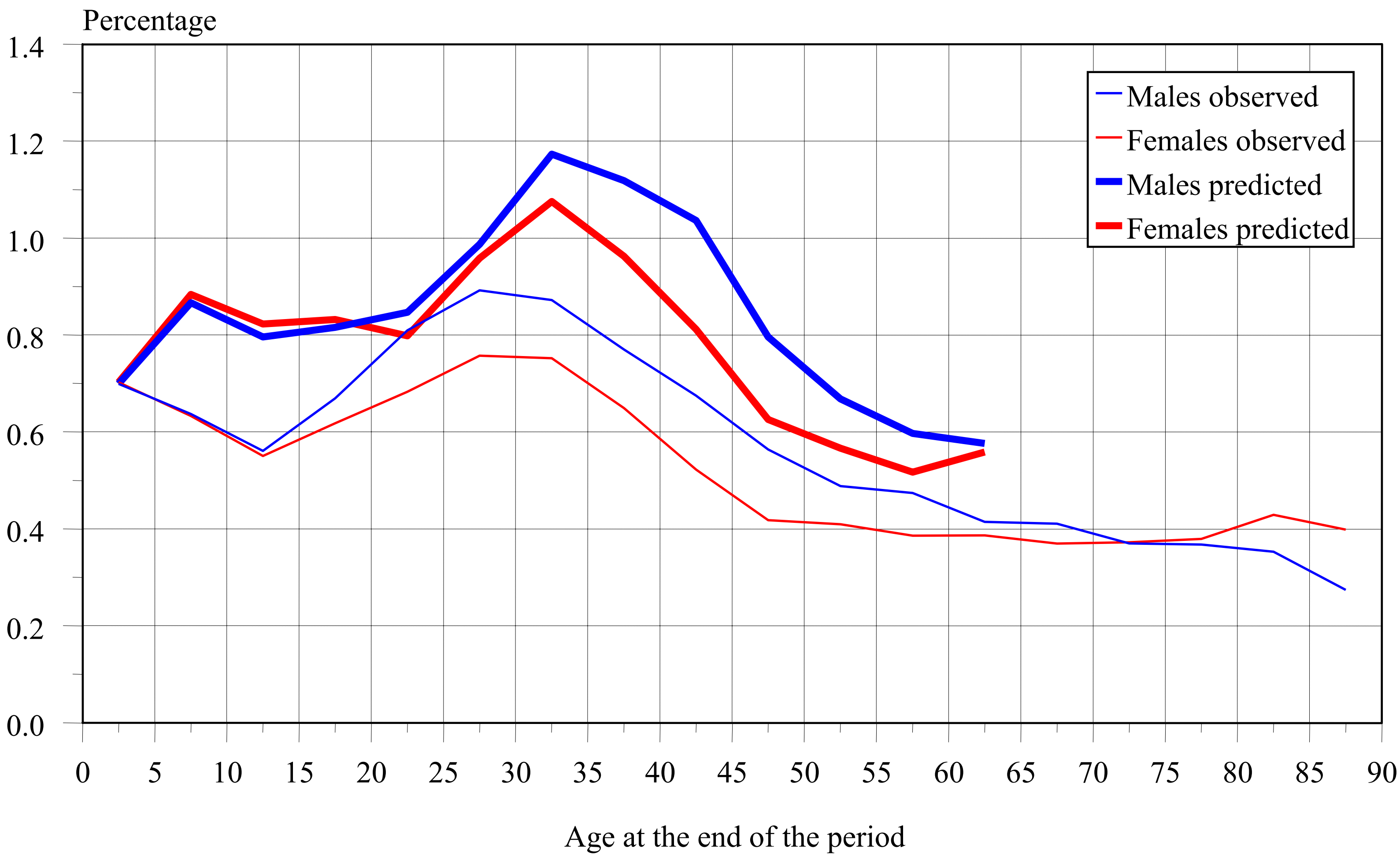
Observed and predicted conditional migration proportions from Border to North Central 1995-2000, using the 1985-1990 age-to-infant ratios (best fitting)



Observed and predicted conditional migration proportions from Central to North Central 1985-1990, using the 1995-2000 age-to-infant ratios (worst fitting)



Observed and predicted conditional migration proportions from Central to North Central 1995-1990, using the 1985-1990 age-to-infant ratios (worst fitting)



Estimates based on age-specific net migration

The problem is finding coefficients $\beta_{ji}(x)$ multiplying the conditional proportions that

$$N_i^t(x) = \sum_{j \neq i} \beta_{ji}(x) S_{ji}^{t-10}(x) \frac{K_j^t(x)}{K_i^t(x)} - \sum_{j \neq i} \beta_{ij}(x) S_{ij}^{t-10}(x) + \beta_{iI}(x) S_I^t(x)$$

be satisfied, where $S_{iI}(x)$ is the propotion for net international migration in region i.

An alternative procedure is to combine all age groups. Let $H_{ij}^{t}(x)$ be the age profile of conditional survivorship proportions of migrants:

$$H_{ij}^{t}(x)=\frac{S_{ij}^{t}(x)}{S_{ij}^{t}(\bullet)} \quad \text{with} \quad S_{ij}^{t}(\bullet)=\sum_{x} S_{ij}^{t}(x) \quad \text{and} \quad \sum_{x} H_{ij}^{t}(x)=1$$

or,

$$N_{i}^{t}(x)=\sum_{j\neq i}\beta_{ji}H_{ji}^{t-10}(x)\frac{K_{j}^{t}(x)}{K_{i}^{t}(x)}-\sum_{j\neq i}\beta_{ij}H_{ij}^{t-10}(x)+\beta_{iI}H_{I}^{t}(x)$$

where β_{ij} refers to the migration *level* from *i* to *j*, because β_{ij} is equal to $S_{ij}^{t}(\bullet)$. Coefficients β_{ij} were estimated by ordinary least squares regression.

Observed and predicted gross interregional migraproduction proportions (linear regression coefficients) by sex, using net migration residual estimation, 1995-2000

	Males				Females			
	Region of origin				Region of origin			
Region of destiny	Border	North Central	Central	South	Border	North Central	Central	South
				Observed				
Border		0.0885	0.0532	0.0162		0.0815	0.0474	0.0134
North Central	0.1549		0.0655	0.0152	0.1381		0.0659	0.0127
Central	0.0909	0.0719		0.0631	0.0718	0.0613		0.0558
South	0.0782	0.0714	0.1478		0.0574	0.0584	0.1541	
				Setting as emigration proportions				
Border		2.5602	1.8965	1.0931		-0.8745	2.7347	-1.5962
North Central	-0.9135		-0.6370	-0.0941	-0.9954		-1.5213	1.1605
Central	0.0709	-2.7486		-2.1256	0.3735	1.0126		0.5069
South	-0.0321	2.5280	0.7684		0.3927	2.5364	-0.3786	
				Setting as immigration proportions				
Border		-1.1472	-0.4610	2.4567		-0.5208	-0.1319	0.9284
North Central	2.9429		-0.0994	0.6138	4.6044		0.1106	-1.7801
Central	7.0542	-7.2256		10.7858	-1.9753	1.3419		1.1927
South	1.2060	-0.6101	-0.2067		-0.0426	-1.2229	1.0662	
				Average				
Border		0.7065	0.7178	1.7749		-0.6977	1.3014	-0.3339
North Central	1.0147		-0.3682	0.2599	1.8045		-0.7053	-0.3098
Central	3.5626	-4.9871		4.3301	-0.8009	1.1773		0.8498
South	0.5869	0.9590	0.2808		0.1750	0.6568	0.3438	
Net international migration	-0.3844	-0.9053	-1.2956	-1.5525	0.0312	-0.0934	0.3537	-0.9876

How can we avoid the negative values? Solving the system of lineal equations:

$$N_i^t(x) = \sum_{j \neq i} \beta_{ji}(x) S_{ji}^{t-10}(x) \frac{K_j^t(x)}{K_i^t(x)} - \sum_{j \neq i} \beta_{ij}(x) S_{ij}^{t-10}(x) + \beta_{iI}(x) S_I^t(x)$$

subject to the restriction of all coefficients $\beta_{ij}(x)$ be positive; coefficients $\beta_{iI}(x)$ for net international migration can be positive o negative.

A solution to the problem is an algorithm of quadratic programming:

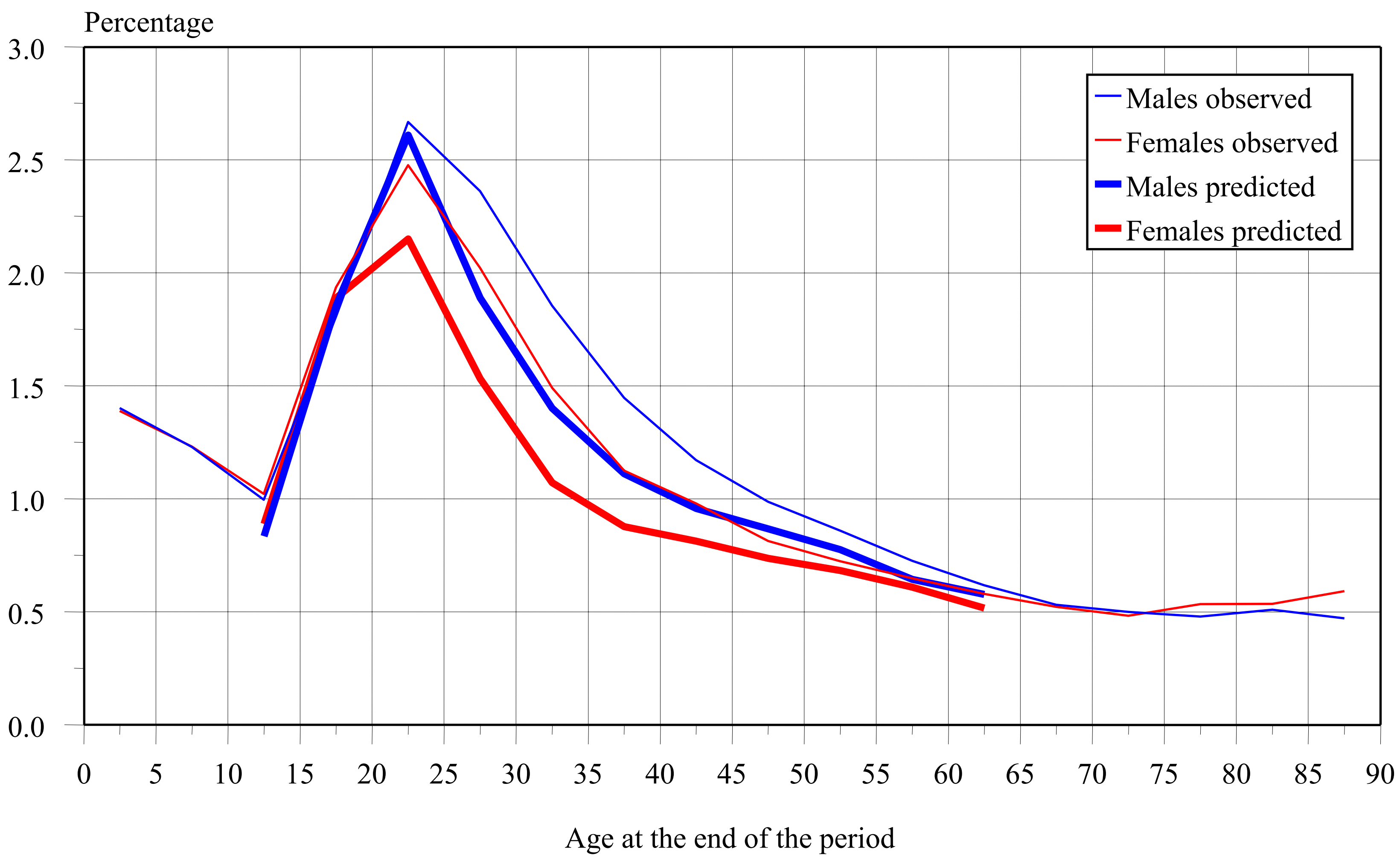
$$\text{maximize} \qquad \overline{g}^{\mathrm{T}}\,\overline{x} + \frac{1}{2}\overline{x}^{T}\boldsymbol{H}\,\overline{x} \qquad \text{subject to} \qquad \boldsymbol{A}\,\overline{x} = \overline{b}$$

The algorithm was applied by age and sex. g is a vector of 16 elements all equal to –1/16, $\boldsymbol{H}$ is a diagonal matrix of order 16 with all cells equal to 2/16, x is a vector containing the 16 coefficients (12 for interregional flows and 4 for net international migration) and $\boldsymbol{A}$ is a matrix of 4 rows by 16 columns that contains the variables meetly suitable.

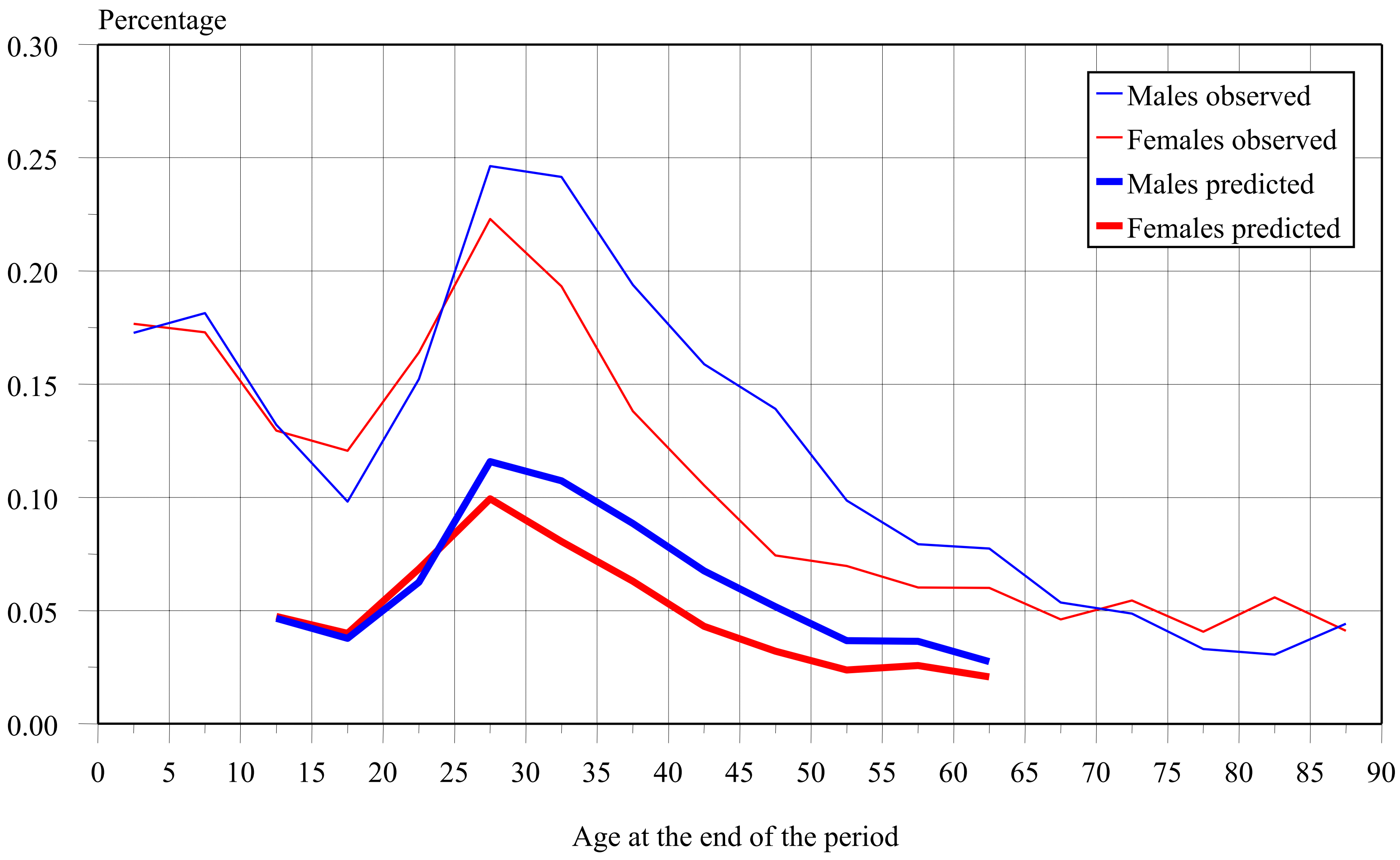
MAPE statistics as mesaure of goodness of fit for netmigration method, 1995-2000

Region of destiny	Region of origin			
	Border	North Central	Central	South
		All		
Border		13.6	47.3	67.6
North Central	48.1		17.8	46.6
Central	51.0	38.1		17.2
South	59.6	52.2	53.3	
All	**42.7**			
		Males		
Border		13.4	47.7	67.7
North Central	49.2		18.1	45.7
Central	52.5	40.4		18.2
South	59.1	53.3	53.4	
All	**43.2**			
		Females		
Border		13.9	46.9	67.5
North Central	47.0		17.5	47.5
Central	49.5	35.7		16.1
South	60.2	51.2	53.1	
All	**42.2**			

Observed and predicted conditional migration proportions from North Central to Border1995-2000, using the 1985-1990 proportions and net migration ratios (best fitting)



Observed and predicted conditional migration proportions from Border to South, 1995-2000, using the 1985-1990 proportions and net migration ratios (worst fitting)



Estimates based on two successive distributions of population stocks

This procedure uses more complex data: migrants by place of residence at census time and place of residence five years before, classified by place of birth. If we consider our four regions, the problem consist in solving four systems of simultaneous equations. If ${}_k r_{ij}$ denote the secondary to primary (STP) migration ratio

$$ {}_k r_{ij} = \frac{{}_k S_{ij}}{{}_i S_{ij}} $$

for Border region of residence at end of period, for example,

$$
\begin{aligned}
{}_B K_B^{90} &= {}_B S_{BB} \; ({}_B r_{BB} \, {}_B K_B^{85}) + {}_N S_{NB} \; ({}_B r_{NB} \, {}_B K_N^{85}) + {}_C S_{CB} \; ({}_B r_{CB} \, {}_B K_C^{85}) + {}_S S_{SB} \; ({}_B r_{SB} \, {}_B K_S^{85}) \\
{}_N K_B^{90} &= {}_B S_{BB} \; ({}_N r_{BB} \, {}_N K_B^{85}) + {}_N S_{NB} \; ({}_N r_{NB} \, {}_N K_N^{85}) + {}_C S_{CB} \; ({}_N r_{CB} \, {}_N K_C^{85}) + {}_S S_{SB} \; ({}_N r_{SB} \, {}_N K_S^{85}) \\
{}_C K_B^{90} &= {}_B S_{BB} \; ({}_C r_{BB} \, {}_C K_B^{85}) + {}_N S_{NB} \; ({}_C r_{NB} \, {}_C K_N^{85}) + {}_C S_{CB} \; ({}_C r_{CB} \, {}_C K_C^{85}) + {}_S S_{SB} \; ({}_C r_{SB} \, {}_C K_S^{85}) \\
{}_S K_B^{90} &= {}_B S_{BB} \; ({}_S r_{BB} \, {}_S K_B^{85}) + {}_N S_{NB} \; ({}_S r_{NB} \, {}_S K_N^{85}) + {}_C S_{CB} \; ({}_S r_{CB} \, {}_S K_C^{85}) + {}_S S_{SB} \; ({}_S r_{SB} \, {}_S K_S^{85})
\end{aligned}
$$

So we have four simultaneous equations with four unknowns: conditional survivorship proportions of primary migrants ${}_i S_{ij}$.

Most of results are acceptable, but 17 are unreal (negative) for males and 18 for females, because the solution is unrestricted. To avoid the negative values, we use again the algorithm of quadratic programming. The lineal equations systems in were organized by birthplace. The lineal equations system for those born in Border region, for example, is:

$$
\begin{aligned}
{}_{B}K_{B}^{90} &= {}_{B}S_{BB} \quad {}_{B}K_{B}^{85} + {}_{B}S_{NB} \quad {}_{B}K_{N}^{85} + {}_{B}S_{CB} \quad {}_{B}K_{C}^{85} + {}_{B}S_{SB} \quad {}_{B}K_{S}^{85} \\
{}_{B}K_{N}^{90} &= {}_{B}S_{BN} \quad {}_{B}K_{B}^{85} + {}_{B}S_{NN} \quad {}_{B}K_{N}^{85} + {}_{B}S_{CN} \quad {}_{B}K_{C}^{85} + {}_{B}S_{SN} \quad {}_{B}K_{S}^{85} \\
{}_{B}K_{C}^{90} &= {}_{B}S_{BC} \quad {}_{B}K_{B}^{85} + {}_{B}S_{NC} \quad {}_{B}K_{N}^{85} + {}_{B}S_{CC} \quad {}_{B}K_{C}^{85} + {}_{B}S_{SC} \quad {}_{B}K_{S}^{85} \\
{}_{B}K_{S}^{90} &= {}_{B}S_{BS} \quad {}_{B}K_{B}^{85} + {}_{B}S_{NS} \quad {}_{B}K_{N}^{85} + {}_{B}S_{CS} \quad {}_{B}K_{C}^{85} + {}_{B}S_{SS} \quad {}_{B}K_{S}^{85}
\end{aligned}
$$

and in a similar way for the other three regions. We adopt this approach because, seen down, the conditional proportions of secondary migration sum up one.

The idea is to estimate “correcting factors” for the proportions of 2000 census to satisfy the population by birthplace and residence of 1990 census:

$$
\begin{aligned}
{}_B K_B^{90} &= {}_B\beta_{BB}\ {}_B S_{BB}\quad {}_B K_B^{85} + {}_B\beta_{NB}\ {}_B S_{NB}\quad {}_B K_N^{85} + {}_B\beta_{CB}\ {}_B S_{CB}\quad {}_B K_C^{85} + {}_B\beta_{SB}\ {}_B S_{SB}\quad {}_B K_S^{8}\\
{}_B K_N^{90} &= {}_B\beta_{BN}\ {}_B S_{BN}\quad {}_B K_B^{85} + {}_B\beta_{NN}\ {}_B S_{NN}\quad {}_B K_N^{85} + {}_B\beta_{CN}\ {}_B S_{CN}\quad {}_B K_C^{85} + {}_B\beta_{SN}\ {}_B S_{SN}\quad {}_B K\\
{}_B K_C^{90} &= {}_B\beta_{BC}\ {}_B S_{BC}\quad {}_B K_B^{85} + {}_B\beta_{NC}\ {}_B S_{NC}\quad {}_B K_N^{85} + {}_B\beta_{CC}\ {}_B S_{CC}\quad {}_B K_C^{85} + {}_B\beta_{SC}\ {}_B S_{SC}\quad {}_B K\\
{}_B K_S^{90} &= {}_B\beta_{Bs}\ {}_B S_{BS}\quad {}_B K_B^{85} + {}_B\beta_{NS}\ {}_B S_{NS}\quad {}_B K_N^{85} + {}_B\beta_{CS}\ {}_B S_{CS}\quad {}_B K_C^{85} + {}_B\beta_{SS}\ {}_B S_{SS}\quad {}_B K_S^{85}
\end{aligned}
$$

Vector x contains the 16 correcting factors and $\boldsymbol{A}$ matrix —again of 4 rows by 16 columns— contains conditional survivorship proportions multiplied per migrants by place of birth.

MAPE statistics as mesaure of goodness of fit for 1985-1990 conditional survivorship proportions of migrants, using 1995-2000 secondary migration proportions

Region of destiny	Region of origin			
	Border	North Central	Central	South
		All		
Border		49.4	24.6	23.8
North Central	53.2		52.1	23.7
Central	52.4	61.4		58.6
South	41.2	47.1	40.6	
All	**44.0**			
		Males		
Border		49.2	24.2	18.4
North Central	52.9		50.5	24.3
Central	51.9	61.3		58.9
South	43.8	47.2	40.4	
All	**43.6**			
		Females		
Border		49.5	25.1	29.3
North Central	53.5		53.7	23.1
Central	52.8	61.6		58.3
South	38.6	47.1	40.7	
All	**44.4**			

Estimates based on iterative biproporcional adjustment

Between three methods proposed by Rogers, Raymer and Jordan, the first one (age-to-infant ratios) gives the best results for Mexico (the smallest MAPE statistic); however, proportions' values do not strictly satisfy the lineal equations system:

$$K_B^t = S_{BB}\,K_B^{t-5} + S_{NB}\,K_N^{t-5} + S_{CB}\,K_C^{t-5} + S_{SB}\,K_S^{t-5}$$
$$K_N^t = S_{BN}\,K_B^{t-5} + S_{NN}\,K_N^{t-5} + S_{CN}\,K_C^{t-5} + S_{SN}\,K_S^{t-5}$$
$$K_C^t = S_{BC}\,K_B^{t-5} + S_{NC}\,K_N^{t-5} + S_{CC}\,K_C^{t-5} + S_{SC}\,K_S^{t-5}$$
$$K_S^t = S_{BS}\,K_B^{t-5} + S_{NS}\,K_N^{t-5} + S_{CS}\,K_C^{t-5} + S_{SS}\,K_S^{t-5}$$

We can write migration flows for Mexico's four regions system as a bivariate array:

Region of destiny	Region of origin				
	Border	North Central	Central	South	Total
Border	K_{BB}	K_{NB}	K_{CB}	K_{SB}	$K_B(t)$
North Central	K_{BN}	K_{NN}	K_{CN}	K_{SN}	$K_N(t)$
Central	K_{BC}	K_{NC}	K_{CC}	K_{SC}	$K_C(t)$
South	K_{BS}	K_{NS}	K_{CS}	K_{SS}	$K_S(t)$
Total	$K_B(t-5)$	$K_N(t-5)$	$K_C(t-5)$	$K_S(t-5)$	

Assuming again that the age patterns of migration flows are similar in two successive censuses, the values for initial array can be estimates as:

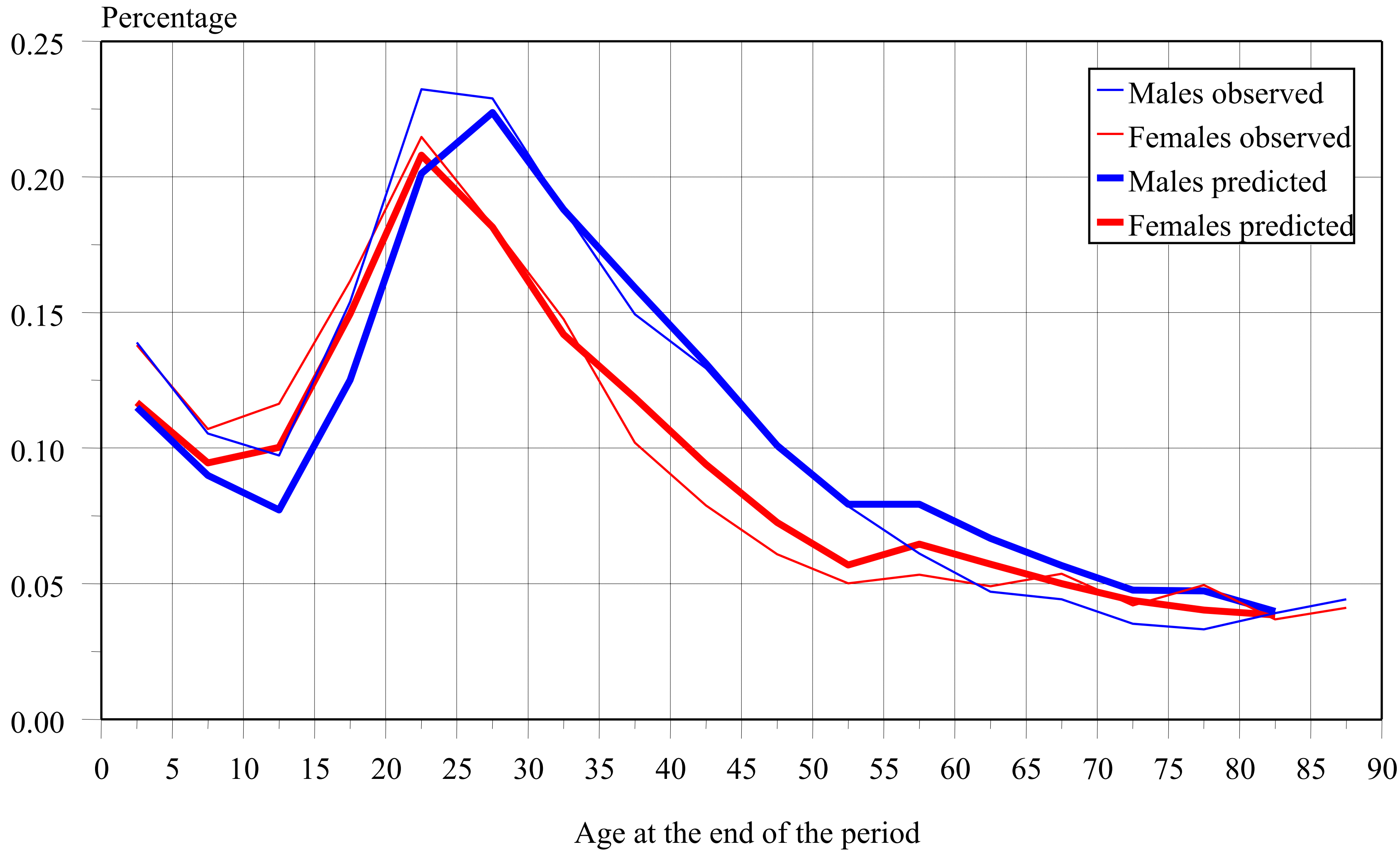
$$K_{ij}^{(0)} = S_{ij}^{t-10} K_i^{t-5}$$

where "(0)" indicates the iteration zero or initial array.

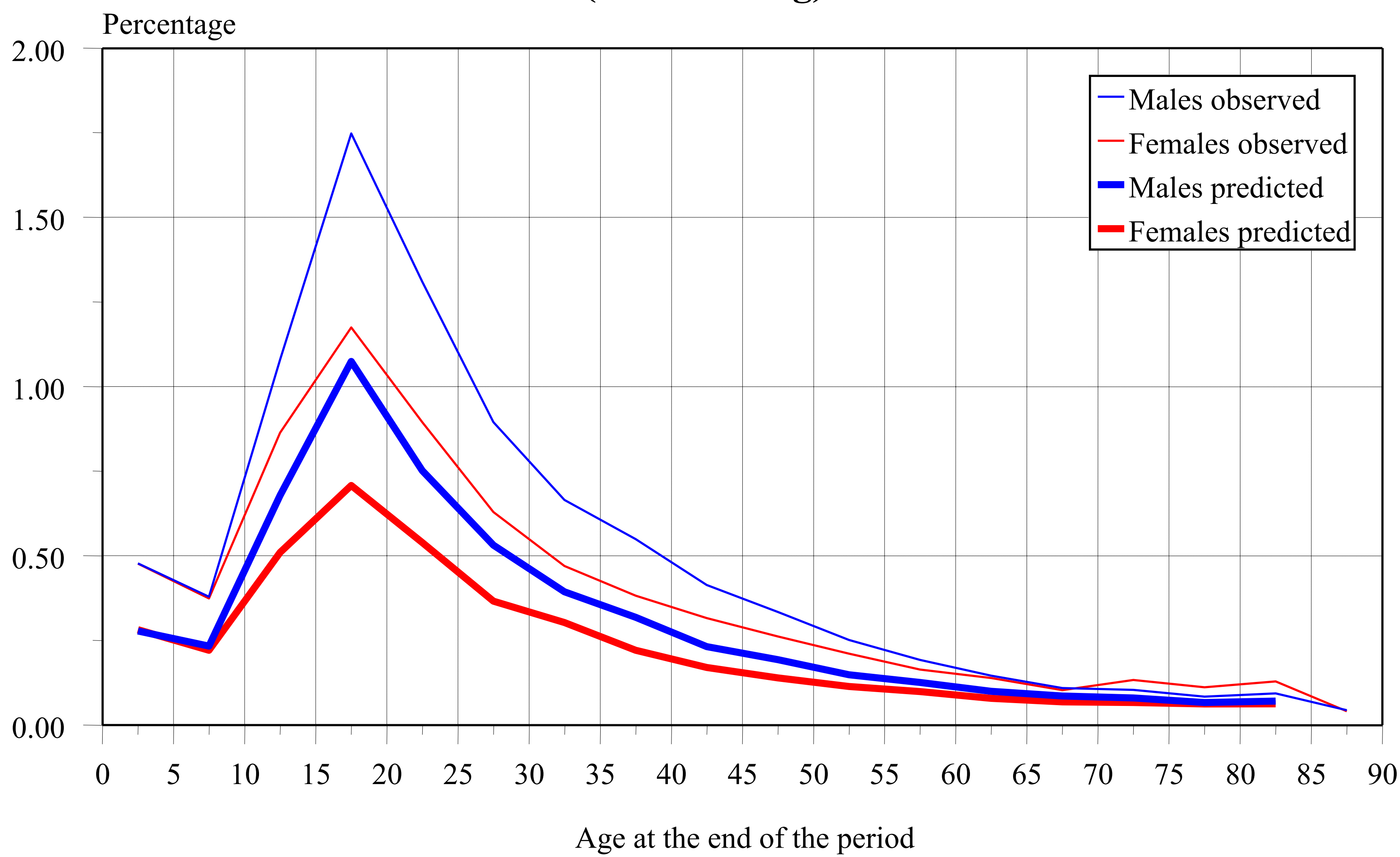
MAPE statistics as mesaure of goodness of fit for 1995-2000 conditional survivorship proportions of migrants, using 1985-1990 migration proportions and iterative biproportional adjustment method

Region of destiny	Region of origin			
	Border	North Central	Central	South
		All		
Border		31.6	14.7	41.1
North Central	15.8		49.7	8.7
Central	12.1	29.2		35.5
South	17.6	11.3	14.6	
All	**23.5**			
		Males		
Border		32.3	15.5	40.4
North Central	13.7		46.7	8.1
Central	11.0	25.0		35.4
South	19.2	10.6	13.3	
All	**22.6**			
		Females		
Border		30.9	13.9	41.7
North Central	18.0		52.7	9.4
Central	13.3	33.4		35.7
South	15.9	12.1	15.9	
All	**24.4**			

Observed and predicted conditional migration proportions from North Central to South, 1995-2000, using the 1985-1990 migration proportions and iterative biproportional adjustment method (best fitting)



Observed and predicted conditional migration proportions from South to Border, 1995-2000, using the 1985-1990 migration proportions and iterative biproportional adjustment method (worst fitting)



Conclusions

Results of first method suggested by Rogers, Raymer and Jordan (ATI) were generally good, but not for the other two, because they turned unrealistic values, mainly negative, for the proportions. Looking for avoiding inadequate values, a possibility of guaranteeing positive values was explored using a quadratic programming algorithm. Nevertheless, the procedure generally underestimates migration level.

Another method, based on the iterative biproporcional adjustment algorithm, was intended. This procedure uses the same data than the other three, and its results were good and similar to those of Method 1 (ATI); but the fourth method presents the advantage that satisfies the number of migrants by region of origin and region of destiny from census.